

**UNIVERSIDAD CATÓLICA SANTO TORIBIO DE MOGROVEJO**  
**FACULTAD DE INGENIERÍA**  
**ESCUELA DE INGENIERÍA DE SISTEMAS Y COMPUTACIÓN**



**APLICACIÓN WEB PARA LA ELABORACIÓN DE PERFILES DE  
CONSUMIDOR BASADA EN MINERÍA DE DATOS Y ARQUITECTURA  
CLOUD PARA EL APOYO AL PROCESO DE CONVERSIÓN DE LEADS  
EN LA ASOCIACIÓN AIESEC EN PERÚ**

**TESIS PARA OPTAR EL TÍTULO DE  
INGENIERO DE SISTEMAS Y COMPUTACIÓN**

**AUTOR**

**WALTER MIGUEL RIOJA CURO**

**ASESOR**

**JUAN ANTONIO TORRES BENAVIDES**

<https://orcid.org/0000-0002-0133-119X>

**Chiclayo, 2020**

**APLICACIÓN WEB PARA LA ELABORACIÓN DE PERFILES  
DE CONSUMIDOR BASADO EN MINERÍA DE DATOS Y  
ARQUITECTURA CLOUD PARA EL APOYO AL PROCESO  
DE CONVERSIÓN DE LEADS EN LA ASOCIACIÓN AIESEC  
EN PERÚ**

PRESENTADA POR:  
**WALTER MIGUEL RIOJA CURO**

A la Facultad de Ingeniería de la  
Universidad Católica Santo Toribio de Mogrovejo  
para optar el título de

**INGENIERO DE SISTEMAS Y COMPUTACIÓN**

APROBADA POR:

Marlon Eugenio Vilchez Rivas  
PRESIDENTE

Ernesto Ludwig Nicho Córdova  
SECRETARIO

Juan Antonio Torres Benavides  
ASESOR

## **DEDICATORIA**

A mis padres, por enseñarme lo que sé, formarme en quien soy, y me dieron tantos cafés para culminar este trabajo. ¡Lo logramos!

A los interesados en el mundo del data science, esperando este trabajo avive su interés por la inteligencia de negocios y la orientación de la tecnología a este rubro.

## **AGRADECIMIENTOS**

A Dios, por guiarme en el camino de la verdad, la paz,  
el bien y brindarme a las mejores personas para  
compartir la vida.

A mi familia: mi Chani, Walter, Michelle, Lucia,  
Arturo, Ángel y mi madrina Bertha, por acompañarme  
durante mi recorrido por la universidad, apoyándome  
y alentándome a perseguir mis sueños y viajes.

A Daniela, José y Guadalupe, por apoyarme durante la  
etapa de desarrollo de mi aplicación.

A mi asesor de tesis, el Ing. Juan Torres, por su  
paciencia al aguantar mi terquedad y resolver mis  
dudas, y su dedicación para llevar adelante esta  
investigación, puliéndola y apoyándome en todo  
momento. Gracias totales. Arrivederci. Tanti auguri.

## RESUMEN

La conversión de leads resulta vital en el cumplimiento de metas de organizaciones dedicadas a la difusión de productos a través del marketing digital. AIESEC en Perú es un ejemplo de las dificultades que puede afrontar una asociación para obtener consumidores finales. Debido a esta problemática se inicia una investigación tecnológico aplicada a nivel preexperimental con el objetivo de apoyar dicho proceso a través de una aplicación basada en la segmentación de perfiles de consumidor mediante técnicas de minería de datos. De este modo se facilita el lead nurturing y la toma de decisiones al mostrar reportes detallados para proponer productos a la medida. El software planteado como producto acreditable ha sido desarrollado siguiendo una hibridación entre las metodologías CRISP-DM y RUP y una arquitectura cloud computing. Para evaluar el impacto de la herramienta se ha considerado la validación de expertos, usuarios y consumidores respecto a los beneficios obtenidos por la aplicación, concluyendo en una reducción al mínimo del tiempo de respuesta a leads, incremento de la calidad de información presentada a los miembros asociados de la organización y clientes en un 81% y 85%, respectivamente; así como la verificación de la calidad de software, siendo esta un 94%. Esta investigación representa un ejemplo del impacto generado por herramientas de inteligencia de negocios en empresas con recursos financieros limitados, otorgándoles una ventaja competitiva y facilitando el logro de los objetivos organizacionales

**PALABRAS CLAVE:** minería de datos, marketing digital, clusterización, perfiles de consumidor, computación en la nube.

## **ABSTRACT**

Lead conversion is vital in meeting the goals of organizations dedicated to the diffusion of products through digital marketing channels. AIESEC in Peru is an example of the difficulties that an association can face to obtain final consumers. Due to this problem, a technological applied investigation at a pre-experimental level is initiated with the objective of supporting this process through an application based on the segmentation of consumer profiles through data mining techniques. This facilitates lead nurturing and decision making by displaying detailed reports to propose tailored products. The software proposed as credible product has been developed following a hybridization between the CRISP-DM and RUP methodologies and a cloud computing architecture. To assess the impact of the tool, the validation of experts, users and consumers regarding the benefits obtained by the application has been considered, concluding in a minimization of lead response time, increasing the quality of information presented to members partners of the organization and clients in 81% and 85%, respectively; as well as the verification of software quality, this being 94%. This research represents an example of the impact generated by business intelligence tools in companies with limited financial resources, giving them a competitive advantage and facilitating the achievement of organizational objectives.

**KEYWORDS:** Data mining, digital marketing, clustering, customer profile, cloud computing.

## ÍNDICE

<b>I.</b>	<b>INTRODUCCIÓN .....</b>	<b>15</b>
<b>II.</b>	<b>MARCO TEÓRICO .....</b>	<b>18</b>
2.1.	ANTECEDENTES .....	18
2.1.1.	ANTECEDENTES INTERNACIONALES .....	18
2.1.2.	ANTECEDENTES NACIONALES.....	19
2.1.3.	ANTECEDENTES LOCALES.....	21
2.2.	BASES TEÓRICO CIENTÍFICAS .....	24
2.2.1.	ARQUITECTURA DE SOFTWARE.....	24
2.2.1.1.	Arquitectura orientada a servicios.....	25
2.2.1.2.	Cloud computing .....	26
2.2.2.	MARKETING DIGITAL .....	27
2.2.2.1.	Conversión de leads.....	28
2.2.2.2.	Lead nurturing .....	29
2.2.3.	MINERÍA DE DATOS .....	29
2.2.3.1.	Técnicas dirigidas .....	30
2.2.3.2.	Técnicas no dirigidas.....	31
2.2.3.3.	Herramientas de minería de datos.....	32
2.2.3.4.	Metodologías de minería de datos.....	33
2.2.4.	METODOLOGÍAS DE DESARROLLO DE SOFTWARE .....	37
2.2.4.1.	CommonKads .....	37
2.2.4.2.	RUP.....	38
<b>III.</b>	<b>METODOLOGÍA .....</b>	<b>39</b>
3.1.	TIPO Y NIVEL DE INVESTIGACIÓN .....	39
3.1.1.	TIPO DE INVESTIGACIÓN.....	39
3.1.2.	NIVEL DE INVESTIGACIÓN .....	39
3.2.	DISEÑO DE INVESTIGACIÓN .....	39
3.3.	POBLACIÓN, MUESTRA Y MUESTREO .....	40
3.3.1.	POBLACIÓN .....	40
3.3.2.	MUESTRA .....	40
3.3.3.	MUESTREO.....	41
3.4.	CRITERIOS DE SELECCIÓN.....	41

<b>3.5.</b>	<b>OPERACIONALIZACIÓN DE VARIABLES .....</b>	<b>41</b>
<b>3.5.1.</b>	<b>VARIABLES.....</b>	<b>42</b>
<b>3.5.1.1.</b>	<b>Variable independiente .....</b>	<b>42</b>
<b>3.5.1.2.</b>	<b>Variable dependiente .....</b>	<b>42</b>
<b>3.5.1.3.</b>	<b>INDICADORES.....</b>	<b>43</b>
<b>3.6.</b>	<b>TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS .....</b>	<b>44</b>
<b>3.7.</b>	<b>PROCEDIMIENTOS .....</b>	<b>44</b>
<b>3.7.1.</b>	<b>METODOLOGÍA DE DESARROLLO .....</b>	<b>44</b>
<b>3.7.2.</b>	<b>ANÁLISIS DE RIESGOS.....</b>	<b>46</b>
<b>3.7.3.</b>	<b>PRODUCTO ACREDITABLE .....</b>	<b>47</b>
<b>3.7.4.</b>	<b>MANUAL DE USUARIO.....</b>	<b>47</b>
<b>3.8.</b>	<b>PLAN DE PROCESAMIENTO Y ANÁLISIS DE DATOS.....</b>	<b>47</b>
<b>3.9.</b>	<b>MATRIZ DE CONSISTENCIA .....</b>	<b>48</b>
<b>3.10.</b>	<b>CONSIDERACIONES ÉTICAS.....</b>	<b>49</b>
<b>IV.</b>	<b>RESULTADOS .....</b>	<b>50</b>
<b>4.1.</b>	<b>EN BASE A LA METODOLOGÍA UTILIZADA .....</b>	<b>50</b>
<b>4.1.1.</b>	<b>ITERACIÓN #1: COMPRENSIÓN DEL NEGOCIO (CRISP-DM) .....</b>	<b>50</b>
<b>4.1.1.1.</b>	<b>Determinación de los objetivos del negocio.....</b>	<b>50</b>
<b>4.1.1.2.</b>	<b>Evaluación de la situación.....</b>	<b>51</b>
<b>4.1.1.3.</b>	<b>Determinación de los objetivos de DM .....</b>	<b>52</b>
<b>4.1.1.4.</b>	<b>Producción de un plan de proyecto.....</b>	<b>52</b>
<b>4.1.2.</b>	<b>ITERACIÓN #2: COMPRENSIÓN DE LOS DATOS (CRISP-DM).....</b>	<b>54</b>
<b>4.1.2.1.</b>	<b>Recolección de datos iniciales .....</b>	<b>54</b>
<b>4.1.2.2.</b>	<b>Descripción de los datos .....</b>	<b>55</b>
<b>4.1.2.3.</b>	<b>Exploración de los datos .....</b>	<b>57</b>
<b>4.1.2.4.</b>	<b>Verificación de la calidad de los datos .....</b>	<b>64</b>
<b>4.1.3.</b>	<b>ITERACIÓN #3: PREPARACIÓN DE LOS DATOS (CRISP-DM).....</b>	<b>65</b>
<b>4.1.3.1.</b>	<b>Selección de datos .....</b>	<b>65</b>
<b>4.1.3.2.</b>	<b>Limpieza de los datos .....</b>	<b>66</b>
<b>4.1.3.3.</b>	<b>Estructuración de los datos.....</b>	<b>68</b>
<b>4.1.3.4.</b>	<b>Integración de los datos.....</b>	<b>68</b>
<b>4.1.3.5.</b>	<b>Formateo de los datos.....</b>	<b>73</b>
<b>4.1.4.</b>	<b>ITERACIÓN #4: MODELADO (CRISP-DM).....</b>	<b>75</b>
<b>4.1.4.1.</b>	<b>Selección de la técnica de modelado.....</b>	<b>75</b>
<b>4.1.4.2.</b>	<b>Generación del plan de prueba .....</b>	<b>75</b>

4.1.4.3.	Construcción del modelo.....	76
4.1.4.4.	Evaluación del modelo .....	77
4.1.5.	ITERACIÓN #5: DISEÑO (RUP).....	78
4.1.5.1.	Diagrama de contexto de diseño.....	78
4.1.5.2.	Diagrama de clases general.....	78
4.1.5.3.	Diseño de base de datos.....	78
4.1.5.4.	Diagramas de estados .....	79
4.1.5.5.	Diseño de interfaces .....	79
4.1.5.6.	Diseño de la arquitectura.....	87
4.1.6.	ITERACIÓN #6: EVALUACIÓN (CRISP-DM) .....	87
4.1.6.1.	Evaluación de los resultados.....	87
4.1.6.2.	Proceso de revisión .....	87
4.1.6.3.	Determinación de futuras fases .....	88
4.1.7.	ITERACIÓN #7: IMPLEMENTACIÓN Y PRUEBA (CRISP-DM Y RUP) ....	88
4.1.7.1.	Diagrama de componentes (RUP).....	88
4.1.7.2.	Plan de implantación (CRISP-DM) .....	89
4.1.7.3.	Monitorización y mantenimiento (CRISP-DM).....	89
4.1.7.4.	Informe final (CRISP-DM).....	89
4.1.7.5.	Pruebas unitarias y de integración (RUP).....	89
4.1.7.6.	Revisión del proyecto (CRISP-DM).....	92
4.2.	EN BASE A LOS OBJETIVOS DE LA INVESTIGACIÓN .....	<b>93</b>
4.2.1.	REDUCIR EL TIEMPO DE RESPUESTA A LEADS TRAS SU REGISTRO EN LA PLATAFORMA .....	93
4.2.2.	SEGMENTAR LEADS A TRAVÉS DE TÉCNICAS DE MINERÍA DE DATOS....	94
4.2.3.	INCREMENTAR LA CALIDAD DE INFORMACIÓN BRINDADA A LOS DIRECTIVOS DE LA ORGANIZACIÓN PARA LA TOMA DE DECISIONES ESTRATÉGICAS.....	96
4.2.4.	INCREMENTAR LA CALIDAD DE LA INFORMACIÓN PRESENTADA A LEADS EN EL ACOMPAÑAMIENTO DIGITAL DEL PROCESO DE CONVERSIÓN.....	97
4.2.5.	VERIFICAR LA CALIDAD DE LA APLICACIÓN .....	99
4.3.	IMPACTOS ESPERADOS .....	<b>101</b>
4.3.1.	IMPACTOS ECONÓMICOS.....	101
4.3.2.	IMPACTOS SOCIALES .....	101
4.3.3.	IMPACTOS EN TECNOLOGÍA .....	101
4.3.4.	IMPACTOS EN LA FORMACIÓN DE CADENAS PRODUCTIVAS .....	102
V.	DISCUSIÓN .....	<b>103</b>

<b>VI. CONCLUSIONES.....</b>	<b>105</b>
<b>VII. RECOMENDACIONES.....</b>	<b>106</b>
<b>VIII. LISTA DE REFERENCIAS .....</b>	<b>107</b>
<b>IX. ANEXOS.....</b>	<b>113</b>
<b>ANEXO N° 01. CONSTANCIA DE APROBACIÓN DEL PRODUCTO ACREDITABLE DE LA ENTIDAD DONDE SE EJECUTÓ LA TESIS .....</b>	<b>113</b>
<b>ANEXO N° 02. ANÁLISIS DE RIESGOS .....</b>	<b>114</b>
<b>ANEXO N° 03. LISTA DE COTEJOS PARA USABILIDAD DEL SOFTWARE ..</b>	<b>117</b>
<b>ANEXO N° 04. MANUAL DE USUARIO.....</b>	<b>119</b>
<b>ANEXO N° 05. REPORTES DE GESTIÓN 2017-2018.....</b>	<b>129</b>

## ÍNDICE DE TABLAS

TABLA I: CLASIFICACIÓN DE TÉCNICAS DE MINERÍA DE DATOS [40] .....	30
TABLA II: COMPARACIÓN ENTRE LAS HERRAMIENTAS DE MINERÍA DE DATOS RAPIDMINER, WEKA Y ORANGE [54].....	33
TABLA III: RESUMEN DE LAS CORRESPONDENCIAS ENTRE KDD, SEMMA Y CRISP-DM [57] .....	36
TABLA IV: DISTRIBUCIÓN DE LA MUESTRA 2 .....	41
TABLA V: INDICADORES .....	43
TABLA VI: TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS .....	44
TABLA VII: MATRIZ DE CONSISTENCIA .....	48
TABLA VIII: OBJETIVOS DE DM.....	52
TABLA IX: PLAN DEL PROYECTO .....	53
TABLA X: DESCRIPCIÓN DE LOS DATOS .....	55
TABLA XI: DISTRIBUCIÓN DE FRECUENCIAS PARA EL ATRIBUTO EDAD EN EL CONJUNTO INICIAL DE LEADS .....	57
TABLA XII: DISTRIBUCIÓN DE FRECUENCIAS PARA EL ATRIBUTO DEPARTAMENTO EN EL CONJUNTO INICIAL DE LEADS .....	59
TABLA XIII: DISTRIBUCIÓN DE FRECUENCIAS PARA EL ATRIBUTO MES PROYECTADO PARA VIAJAR EN EL CONJUNTO INICIAL DE LEADS .....	60
TABLA XIV: DISTRIBUCIÓN DE FRECUENCIAS PARA EL ATRIBUTO CICLO DE ESTUDIOS EN EL CONJUNTO INICIAL DE LEADS .....	61
TABLA XV: DISTRIBUCIÓN DE FRECUENCIAS PARA EL ATRIBUTO CARRERA PROFESIONAL EN EL CONJUNTO INICIAL DE LEADS .....	62
TABLA XVI: DISTRIBUCIÓN DE FRECUENCIAS PARA EL ATRIBUTO PROGRAMA DE INTERÉS EN EL CONJUNTO INICIAL DE LEADS .....	64
TABLA XVII: SELECCIÓN DE DATOS .....	66
TABLA XVIII: ATRIBUTOS GENERADOS PARA LA TABLA LEAD_REGISTRO .....	68
TABLA XIX: DETALLE DE LA TABLA DESTINO_REGION .....	69
TABLA XX: DETALLE DE LA TABLA DESTINO_PAIS .....	69
TABLA XXI: DETALLE DE LA TABLA CAT_CARRERA .....	69
TABLA XXII: DETALLE DE LA TABLA CARRERA .....	70
TABLA XXIII: DETALLE DE LA TABLA PRODUCTO.....	70
TABLA XXIV: DETALLE DE LA TABLA CAT_INVERSION .....	71
TABLA XXV: DETALLE DE LA TABLA CAT_CICLO .....	71
TABLA XXVI: DETALLE DE LA TABLA DEPARTAMENTO.....	72
TABLA XXVII: DETALLE DE LA TABLA PERFIL_DESEADO.....	72
TABLA XXVIII: DETALLE DE LA TABLA PRODUCTO_PERFIL .....	73

TABLA XXIX: FORMATEO DE LA TABLA LEAD_REGISTRO .....	74
TABLA XXX: COMPARACIÓN DE TIEMPOS DE RESPUESTA A LEADS TRAS SU REGISTRO ...	93
TABLA XXXI: COMPARACIÓN DEL GRADO DE ACEPTACIÓN DE LA INFORMACIÓN POR LOS USUARIOS .....	97
TABLA XXXII: COMPARACIÓN DEL GRADO DE ACEPTACIÓN DE LA INFORMACIÓN BRINDADA AL CONSUMIDOR .....	99
TABLA XXXIII: INTERESADOS INTERNOS.....	114
TABLA XXXIV: INTERESADOS EXTERNOS .....	114
TABLA XXXV: MATRIZ DE RIESGOS DEL PROYECTO.....	115
TABLA XXXVI: MATRIZ DE CONTROLES DE RIESGOS .....	116

## ÍNDICE DE FIGURAS

FIG. 1. METODOLOGÍAS MÁS UTILIZADAS PARA MINERÍA DE DATOS, ANALYTICS Y OTROS PROYECTOS DE DATA SCIENCE. [59].....	36
FIG. 2. EXPORTACIÓN DE DATOS EN LA HERRAMIENTA PODIO .....	54
FIG. 3. GRÁFICO DE LÍNEAS DE LA DISTRIBUCIÓN DE EDADES .....	58
FIG. 4. GRÁFICO DE ANILLO DE LA DISTRIBUCIÓN DE DEPARTAMENTOS .....	60
FIG. 5. GRÁFICO DE COLUMNAS DE LA DISTRIBUCIÓN DE MESES DESEADOS PARA VIAJAR	61
FIG. 6. GRÁFICO DE COLUMNAS DE LA DISTRIBUCIÓN DE CICLO DE ESTUDIOS .....	62
FIG. 7. GRÁFICO DE COLUMNAS DE LA DISTRIBUCIÓN DE CARRERAS .....	64
FIG. 8. GRÁFICO CIRCULAR DE LA DISTRIBUCIÓN DE PROGRAMAS DE INTERÉS .....	64
FIG. 9. CAPTURA DE DATOS INCOMPLETOS EN LA TABLA DE LEADS INTERESADOS EN EL PROGRAMA VOLUNTARIO GLOBAL.....	67
FIG. 10. CAPTURA DE DATOS INCOMPLETOS EN LA TABLA DE LEADS INTERESADOS EN EL PROGRAMA TALENTO GLOBAL .....	67
FIG. 11. FUNCIÓN BUSCARV PARA ELIMINACIÓN DE DUPLICADOS TRAS LA UNIFICACIÓN DE DATASETS .....	67
FIG. 12. MODELO DE PROCESO EN RAPIDMINER.....	77
FIG. 13. DIAGRAMA DE CONTEXTO DE DISEÑO.....	78
FIG. 14. DIAGRAMA DE CLASES GENERAL .....	78
FIG. 15. MODELO RELACIONAL DE BASE DE DATOS .....	78
FIG. 16. DIAGRAMA DE ESTADOS DE LA CLASE LEAD_REGISTRO .....	79
FIG. 17. REGISTRO DE LEADS EN EL ASISTENTE (MÓDULO CLIENTE) .....	79
FIG. 18. INICIO DE SESIÓN PARA ACCESO AL SISTEMA .....	80
FIG. 19. MENÚ PRINCIPAL .....	80
FIG. 20. LISTADO DE PERFILES .....	81
FIG. 21. DETALLE DE UN PERFIL DE CONSUMIDOR .....	82
FIG. 22. DETALLE DE SUBGRUPO (CLÚSTER) DE PERFIL.....	82
FIG. 23. LISTADO DE LEADS .....	83
FIG. 24. DETALLE DE LEAD .....	83
FIG. 25. REPORTE DE TASAS DE CONVERSIÓN .....	84
FIG. 26. INTERFAZ DE REGISTRO EN EL ASISTENTE (MÓDULO CLIENTE) .....	85
FIG. 27. INTERFAZ DE REGISTRO EN EL ASISTENTE (MÓDULO CLIENTE) .....	85
FIG. 28. REPORTE DE APPROVALS (CONSUMIDORES).....	86
FIG. 29. ARQUITECTURA DE LA APLICACIÓN.....	87
FIG. 30. DIAGRAMA DE COMPONENTES .....	88

FIG. 31. ANÁLISIS DE LA VELOCIDAD DE CARGA DEL FORMULARIO DE REGISTRO DE LEADS .....	90
FIG. 32. ANÁLISIS DE LA VELOCIDAD DE CARGA DEL MENÚ PRINCIPAL DEL SISTEMA .....	90
FIG. 33. RESULTADOS DEL ANÁLISIS DEL SOFTWARE DR. WATSON .....	91
FIG. 34. RESULTADOS DE LA PRUEBA AUTOMATIZADA DE CAJA BLANCA MEDIANTE W3C .....	91
FIG. 35. RESULTADOS DE LA PRUEBA DE CONTROLADORES MEDIANTE POSTMAN .....	92
FIG. 36. CORREO ENVIADO AUTOMÁTICAMENTE AL LEAD CON LOS PROGRAMAS A LOS QUE CALIFICA.....	93
FIG. 37. MENSAJE DE TEXTO ENVIADO AL LEAD TRAS SU REGISTRO .....	94
FIG. 38. ÁRBOL BINARIO PARA REGLAS DE CLUSTERIZACIÓN DEL PERFIL #11 EN RAPIDMINER.....	95
FIG. 39. SEGMENTACIÓN DE CLÚSTERES DEL PERFIL #11 EN MYSQL. ....	95
FIG. 40. CAPTURA DE PANTALLA DE LA INTERFAZ DE REPORTE DE TIEMPO DE PROCESAMIENTO POR PROGRAMAS. ....	96
FIG. 41. CAPTURA DE PANTALLA DE LA INTERFAZ DE REPORTE DE TASA DE CONVERSIÓN POR PERFILES. ....	96
FIG. 42. CAPTURA DE PANTALLA DE LA INTERFAZ DE CLÚSTERES DEL PERFIL #07 Y ENVÍO DE CORREOS. ....	98
FIG. 43. CAPTURA DE PANTALLA DEL CORREO ENTREGADO EXITOSAMENTE. ....	98
FIG. 44. MATRIZ DE DATOS DEL JUICIO DE EXPERTOS. ....	100
FIG. 45. CÁLCULO ESTADÍSTICO DEL COEFICIENTE ALFA DE CRONBACH. ....	100
FIG. 46. PROMEDIO DE CALIFICACIONES DE EXPERTOS. ....	100
FIG. 47. CARACTERÍSTICAS DEL PÚBLICO OBJETIVO DE CADA PROGRAMA DE INTERCAMBIO .....	103
FIG. 48. CARTA DE APROBACIÓN DEL PRODUCTO ACREDITABLE.....	113
FIG. 49. CAPTURA DE PANTALLA DEL CORREO CONTENIENDO LOS REPORTE DE GESTIÓN 2017-2018.....	129

## I. INTRODUCCIÓN

Gracias a nuestro mundo más tecnificado, el flujo de información en las organizaciones es incomparable respecto a una década atrás; hoy en día la capacidad de recolección de datos y el procesamiento de información son pilares para la toma de decisiones. Para ello, las empresas hacen uso de herramientas que faciliten estas tareas; según Microsoft en 2014 [1] el 97% de las PYMEs resalta la relevancia de la aplicación de tecnologías de la información y la comunicación (TIC) para su negocio, no obstante, estas se limitan al uso de las funcionalidades más básicas de Microsoft Office, desaprovechando otras alternativas que permitan el procesamiento de datos para impulsar la productividad. Para cambiar esto, [2] reafirma la necesidad de las ONGs, empresas sociales y asociaciones civiles por utilizar herramientas de inteligencia de negocios (BI, del inglés *Business Intelligence*) para el procesamiento de información, haciendo un balance entre el alto costo de inversión para su implementación y los beneficios a obtener tras su aplicación.

En el mismo 2014, EY [3] resalta el impacto de *big data* en la manera en que los negocios compiten y operan. Desde ese año señalan la emergencia de tecnologías y canales digitales accesibles que permitan un análisis de datos más rápido y sencillo para el usuario; posteriormente, [4] complementa en 2019 esta afirmación al presentar en su portafolio soluciones que integran tecnologías de *machine learning* y *cloud computing* para el análisis en tiempo real de datos en las empresas. Asimismo, Deloitte en 2019 [5] señala al análisis de datos complementado con soluciones de inteligencia artificial como catalizadores que permiten a las empresas trascender hoy en día, observando que la cultura organizacional corporativa suele ser el obstáculo para que este tipo de tecnologías sean implementadas (escenario que posteriormente fue visto dentro de la situación problemática de la asociación).

Queda claro que las organizaciones deben hacer uso de tecnologías de procesamiento de datos para la toma oportuna de decisiones, no obstante, el mercado ofrece estos servicios a costos elevados que sobrepasan el presupuesto de las PYMEs, resultando en inaccesibilidad por motivos económicos. Asimismo, la selección de una herramienta resulta en una tarea complicada debido a que cada empresa cuenta con distintas necesidades y tipos de datos, por lo que es necesario

un software a medida, que además de costoso exige también inversiones considerables de tiempo.

De este modo llegamos a AIESEC en Perú, una asociación civil presente en trece regiones del país, la cual gestiona intercambios al extranjero de voluntariado y pasantías profesionales como herramientas para desarrollar en jóvenes las capacidades de liderazgo activo y sostenible a través de una plataforma mediante *e-commerce*. A nivel internacional la organización analiza las tasas de conversión entre *lead*, *sign-up* y cliente (indicadores relevantes según [6]) para determinar el grado de productividad de la asociación y así tomar decisiones, siendo estos de 16% de conversión *lead/sign-up* y 14% de *sign-up/consumidor* según el reporte de gestión alcanzado por AIESEC (ver *Anexo N° 05*). Sin embargo, en Perú dicho análisis no se aprovecha en su totalidad debido al tiempo mal gestionado de los miembros asociados (se dedican a tareas operativas en lugar de actividades propias del nivel táctico/estratégico según corresponda). Asimismo, el ritmo de los resultados operacionales de AIESEC se ha desacelerado (de 37% de crecimiento relativo en 2015 a un 5.99% en 2018), trayendo una baja productividad que afecta la sostenibilidad de la asociación en distintas sedes del país. La redundancia de información es preocupante, pero sobre todo el problema viene en el abordaje de su público objetivo, el cual no está tipificado (no se identifican sus características comunes), y la generación personalizada de paquetes de intercambio, puesto que tarda más de lo necesario debido a la inexperiencia de los asociados y el desaprovechamiento de los datos que obtienen de ellos, los cuales son almacenados en una aplicación web a modo de un CRM. Frente al escenario descrito surgió la interrogante: ¿cómo apoyar el proceso de conversión de *leads* en la asociación? La presente tesis denominada “APLICACIÓN WEB PARA LA ELABORACIÓN DE PERFILES DE CONSUMIDOR BASADO EN MINERÍA DE DATOS Y ARQUITECTURA CLOUD PARA EL APOYO AL PROCESO DE CONVERSIÓN DE LEADS EN LA ASOCIACIÓN AIESEC EN PERÚ” se inicia con el propósito de desarrollar una herramienta que permita almacenar y procesar los datos de *leads* para ser clasificados por perfiles de consumidor y subgrupos por características comunes, de modo que la asociación analice información procesada para establecer “productos a medida”, facilitando la toma de decisiones especializadas para mejorar los ratios de conversión, y así hacer frente al decrecimiento que genera

insostenibilidad financiera en AIESEC en Perú. Esta investigación se justifica en la elaboración de algoritmos basados en la conceptualización del funcionamiento de tecnologías que siguen las tendencias de *big data*, *data analytics* y *machine learning*; así como en el aporte de un software a la medida para una organización sin fines de lucro con recursos financieros limitados, permitiendo así maximizar su impacto y otorgándole una ventaja competitiva.

Para profundizar en este tema se realizó una investigación tecnológico aplicada donde la población se distribuyó en dos grupos: por un lado, jóvenes interesados en adquirir alguno de los programas de la asociación; y por el otro, los asociados miembros de áreas de marketing e intercambios salientes. Se planteó desarrollar un sistema web que a través de diversos algoritmos clasifique a sus *leads* y procese sus datos para apoyar el proceso de conversión en AIESEC en Perú; para lo que se tuvo que ponderar los factores determinantes en la adquisición de un programa de intercambio, clasificar *leads* en base a perfiles de consumo acorde a las reglas de la asociación para después segmentarlos y desarrollar un software que unifique estas funcionalidades. Como objetivos específicos, se planteó reducir el tiempo de respuesta a un *lead* tras su registro, segmentar los perfiles generados a través de técnicas de minería de datos, mejorar la calidad de información determinante para la toma de decisiones, incrementar la satisfacción del *lead* respecto al soporte digital en el proceso de conversión y validar la aplicación desarrollada.

## II. MARCO TEÓRICO

### 2.1. Antecedentes

Para la presente investigación se consideraron los siguientes antecedentes:

#### 2.1.1. Antecedentes internacionales

Deligianni [7] realiza un estudio extenso sobre la innovación basada en datos a la que toda empresa debería estar orientada tras la tendencia que vemos hoy en día respecto a las grandes cantidades de datos, también llamada *big data*. Específicamente esta tesis se enfoca en las Organizaciones no gubernamentales, las cuales al no contar con suficientes recursos económicos suele tomar decisiones basadas solo en la experiencia de aquellas personas que las gestionan. El resultado final de esta investigación es un conjunto de elementos que pueden guiar a las ONGs para basar sus decisiones en los datos que están a su alcance. La guía propuesta de elementos y sugerencias para la utilización de datos que proporcionen información útil para la toma de decisiones en ONGs sirvió para guiar la presente investigación, puesto que se aplicó en una asociación civil sin fines lucrativos. Asimismo, las tendencias actuales y protección de datos que están moldeando la gestión de datos hoy en día, plasmadas en este antecedente, sirvieron de apoyo para generar un producto a la vanguardia que apoye a AIESEC en Perú.

Naber [8] presenta una investigación que surge con la intención de analizar la lealtad de los clientes de la empresa VIATA para determinar los motivos que influyen en su permanencia con la compañía y diferenciar las características con aquellos que optan por la competencia. Para ello aplican un modelo que evalúa factores recientes, frecuentes y monetarios de los clientes (RFM, del inglés *Recency, Frequency, Monetary*) para aplicar minería de datos a través de las técnicas de tablas de decisión, árboles de decisión, bosque aleatorio, redes bayesianas y k vecinos más cercanos (kNN, del inglés *k-nearest neighbors*), con el fin de analizar la lealtad de los clientes para proponer estrategias de marketing. Tras obtener

todos los análisis estadísticos de la exactitud de los algoritmos planteados, se concluyó que la técnica kNN era la más indicada para segmentar clientes en base a su comportamiento y márgenes de error; así se pueden construir estrategias de marketing considerando varios subgrupos de consumidores, dependiendo de su grado de lealtad. El aporte para la presente investigación fue la aplicación y comparación de técnicas de aprendizaje supervisado para clasificar clientes, sirviendo como base para determinar su aplicación en la presente situación problemática, similar a la presentada por Naber.

Manero, Rimiru y Otieno [9] describen la problemática en una empresa proveedora de servicios móviles en Kenia, la cual enfrenta una pérdida de clientes que conlleva a un impacto financiero negativo para la compañía. Frente a esto, aplicaron la metodología KDD con aplicación de diseño completamente aleatorio (CRD, del inglés *Completely Randomized Design*) para obtener una segmentación de clientes a través de técnicas de clasificación y clusterización mediante minería de datos. Se obtuvo clústers tras análisis de componentes principales (PCA, del inglés *Principal Component Analysis*), determinando dos como los óptimos para concentrarse y fidelizar como clientes tras aplicar estrategias de marketing. El aporte principal de este antecedente fue la diferenciación entre técnicas de clasificación y clusterización (aprendizaje dirigido y aprendizaje no dirigido) para la segmentación de clientes, trayendo como resultado similar a uno de los resultados esperados en la presente investigación.

### **2.1.2. Antecedentes nacionales**

Carpio [10] analiza los niveles de riesgo de impago y morosidad en los créditos financieros otorgados por la CRAC Los Andes, los cuales representaban amenazas de consecuencias graves para la entidad financiera. En esta investigación se desarrolló un modelo de predicción de la morosidad en el otorgamiento de crédito aplicando técnicas de minería de datos, utilizando para ello información

histórica. Con el objetivo de calibrar los modelos planteados para *credit scoring*, se aplicaron técnicas como las matrices de errores, validación cruzada, y curvas ROC (del inglés *Receiver Operating Characteristic*). Tras la aplicación de la metodología CRISP-DM se determinó el algoritmo de bosques aleatorios como el idóneo para el modelo predictivo. Así, se redujo la tasa de riesgo, minimizó el índice de morosidad y agilizó la evaluación de crédito hecha por la CRAC. Los aportes para la presente investigación se sustentan en la creación de modelos inteligentes basados en un estudio extensivo de datos no estructurados dentro de una empresa para desarrollar un sistema experto, así como la aplicación de la metodología CRISP-DM para la aplicación de minería de datos.

Grández [11] narra el desaprovechamiento de datos y características de los clientes que determinaban la compra de suplementos nutricionales en una empresa distribuidora, situación que conllevaba a un desconocimiento de las campañas de marketing más adecuadas para aplicar a su cartera de clientes. Se aplicaron algoritmos inteligentes para el análisis de datos tales como el modelo de asociación, el modelo de clústeres y de redes neuronales (los tres de Microsoft); para basar la aplicación de minería de datos basándose en la metodología CRISP-DM a través del software Visual Studio 2015 y como gestor de base de datos SQL Server 2014. Se obtuvo porcentajes para cada paquete que combinaba determinados productos basándose en los patrones de consumo establecidos para cada cliente, favoreciendo al comercio al incrementar sus posibilidades de ingreso económico tras la ejecución y perfeccionamiento de las estrategias planteadas por el autor. Su aporte a la presente investigación fueron las reglas y técnicas de minería de datos para determinar patrones del consumo de los productos de la empresa, presentando algoritmos ya estructurados con resultados que fueron analizados para determinar su aplicabilidad en la situación problemática y su potencial impacto.

Leon [12] analiza la deserción de pacientes en un proceso psicoterapéutico durante un tiempo menor a los cuatro meses de tratamiento. Tras aplicar la metodología SEMMA, se desarrolló un modelo predictivo que determinaba el éxito o fracaso del proceso y tratamiento psicoterapéutico a través de los datos brindados por el paciente, así como la predicción del número de citas efectivas en caso este desertase. Se logró diseñar los prototipos propuestos en la investigación, no obstante, se concluyó que, si bien los objetivos se cumplieron, era necesario un mayor volumen de datos para evaluar e incrementar los niveles de precisión de los algoritmos aplicados. El aporte a la presente investigación fue la aplicación de otras técnicas y herramientas de minería de datos, así como algoritmos especializados (árboles de decisión) para identificar patrones de comportamiento de sujetos basados en características comunes, así como la consideración del número de datos requeridos para un proceso de minería de datos de esta naturaleza.

### **2.1.3. Antecedentes locales**

Quepuy [13] investiga en una institución educativa la falta de disponibilidad de información útil para la toma de decisiones, la cual desencadenaba fallos intuitivos y empíricos que resultaban en disconformidad por parte de los directivos debido a sus consecuencias negativas sobre los resultados del negocio. Debido a esto se implementó un sistema para gestionar y dar soporte a la toma de decisiones estratégicas en la empresa a través de la aplicación de algoritmos de bayes y clúster basándose en la metodología Ralph Kimball. Se concluyó que con un 95% de confianza el sistema propuesto mejoraba la toma de decisiones, así como se facilitó el acceso oportuno a la información, resultando en la satisfacción de la gerencia respecto al sistema y el tiempo de procesamiento. La relación con la presente investigación estuvo en la aplicación de teorías de clúster en un sistema que mejoró el acceso a la información y la disposición de datos mostrados en un *dashboard* que satisfizo a la alta dirección de una empresa.

Gastulo [14] describe el absentismo del personal del departamento de enfermería en una entidad prestadora de servicios de salud debido a solicitudes recurrentes de descansos médicos, rotación de personal y falta de atención en la salud de los colaboradores en tratamientos preventivos de enfermedades. Para esta situación problemática, presentó como solución un sistema web para gestionar los descansos médicos de los colaboradores de las diferentes áreas de las cuales está a cargo dicho departamento, basándose en patrones de comportamiento definidos tras un proceso de minería de datos; aplicando una hibridación entre las metodologías KDD y XP. Finalmente, el autor concluye que se alcanzó el objetivo de la investigación, proveyendo de información confiable y oportuna al departamento de enfermería y mejorando así la gestión de descansos médicos. Se tomó en consideración esta tesis como antecedente debido a la hibridación de metodologías: una de desarrollo de software con otra de minería de datos; asimismo las técnicas aplicadas para la generación de patrones de comportamiento en base a características comunes de un grupo de personas.

Fernández [15] investiga el cálculo impreciso de la producción de espárrago para una temporada determinada que generaba exceso o falta de inventario para dicha época, resultando en una disminución de los niveles de servicio para los compradores. Se aplicaron técnicas de series temporales y regresión para posteriormente compararlas y mostrar cada gráfico junto a los valores de los modelos predictivos en interfaces de un sistema web, así como las simulaciones por meses. Esta tesis también presenta una hibridación entre metodologías de minería de datos y desarrollo de software, siendo esta vez CRISP-DM y XP. Se determinó a ARIMA como la técnica más idónea para la predicción de la producción de espárragos con más de 86% de confiabilidad; asimismo se recomendó realizar estudios posteriores por la empresa para mejorar los modelos predictivos y otra clase de aplicación de herramientas para procesar sus datos. El aporte de este antecedente se centra en la aplicación y

posterior comparación de la efectividad de las técnicas aplicadas, siendo solamente de series temporales y regresión. Este trabajo de fin de grado consideró solo las técnicas de Holt-Winters y redes neuronal autorregresivas: ARMA (del inglés *Autorregressive Moving Average*) y ARIMA (del inglés *Autorregressive Integrated Moving Average*) debido a la naturaleza de los datos elegidos; estos modelos fueron considerados dentro de la selección de la técnica de minería de datos a elegir para la presente investigación.

## 2.2. Bases teórico científicas

### 2.2.1. Arquitectura de Software

El concepto de arquitectura ha sido revisado y propuesto por diferentes autores a lo largo de los años, resultando complicado llegar a un consenso sobre su definición y el límite existente entre diseño de arquitectura y diseño de aplicaciones. [16] realiza una revisión exhaustiva de literatura alrededor del término arquitectura de software, determinando tres tipos de definiciones:

- En el primero, los autores sostienen que la arquitectura de software corresponde a un nivel alto de abstracción de software. No obstante, estas no describen qué representa ese “alto nivel” ni establece si hace referencia a requerimientos funcionales o no funcionales; tampoco delimita las funcionalidades de la aplicación de la arquitectura de esta.
- En el segundo, la arquitectura es definida como una estructura y propiedades externamente visibles de un sistema. Si bien los autores que sostienen esta perspectiva del concepto hacen énfasis en los componentes del sistema y sus relaciones entre ellos y con el ambiente, no se incluye la funcionalidad del software en su definición ni resuelve el dilema del acceso exclusivo para aplicaciones de alto nivel.
- En el tercero, se determinan conceptos fundamentales y restricciones bajo las cuales los sistemas deben ser diseñados y desarrollados. Esta definición da pie a la admisión de nuevas propiedades y principios para la arquitectura de software y el concepto de estilos arquitectónicos. Asimismo, se dejan abiertos a interpretación los límites entre funcionalidades y arquitectura, así como el nivel requerido por la aplicación.

De acuerdo con la norma ISO/IEC/IEEE 42010:2011 [17], la arquitectura de software es un concepto o percepción abstracta de aquello que es fundamental para un sistema considerando su ambiente. Este está conformado por los elementos requeridos para

su construcción, relaciones tanto internas como externas al sistema y los principios de su diseño y evolución.

Sommerville [18] establece a la arquitectura como el puente entre las etapas de gestión de requerimientos y diseño. Asimismo, describe dos objetivos claros para determinar la arquitectura de software en un proyecto: para direccionar debates respecto al diseño del sistema y para documentar una arquitectura previa en uno ya existente. Coincidiendo con [17], este autor describe a los estilos arquitectónicos como un patrón representativo que surge para reaprovechar conocimiento generado por otros sistemas exitosamente, considerando sus beneficios y repercusiones en cada una de las etapas del desarrollo de software.

Existen diversos estilos, patrones o modelos de arquitectura de software, entre los más utilizados de acuerdo a Segura [19] se encuentran: cliente/servidor, arquitectura basada en componentes, arquitectura en capas, orientada a objetos, arquitectura orientada a servicios, modelo vista controlador y fachada de aplicación.

#### **2.2.1.1. Arquitectura orientada a servicios**

SOA (del inglés *Software Oriented Architecture*) es un estilo arquitectónico basado en el modelo de cliente-servidor, en el que se divide un sistema en servicios que no requieren estar ubicados en el mismo espacio físico [18], de este modo pueden coexistir independientemente de las tecnologías utilizadas [20].

Este tipo de aplicaciones dinámicas permite, además de la mejora y automatización de las operaciones de la compañía, la creación y reutilización de servicios y aplicaciones, preparándola para el cambio y evolución. Así, SOA consiste en una herramienta idónea para las organizaciones ágiles que requieren de cumplir con regulaciones y políticas internas [20].

En [19] se explica que SOA puede ser utilizado tanto para servicios existentes dentro de una organización como para el consumo de servicios de terceros.

[21] detallan algunas características y principios de SOA, los cuales son: reconocible y vinculado dinámicamente, autocontenido y modular, interoperabilidad, bajo acoplamiento, transparencia de ubicación, componibilidad y autonomía para la recuperación.

Entre los beneficios de SOA a nivel corporativo se consideran la mejora en la toma de decisiones, mejora de la productividad de los empleados y mejora de la relación con clientes y proveedores; a nivel de TI se consideran el desarrollo rápido y económico de aplicaciones más seguras, manejables, productivas y flexibles, [20].

#### **2.2.1.2. Cloud computing**

Una de las definiciones con mayor aceptación a nivel internacional es la del NIST [22], que define a la arquitectura *cloud* como un modelo que permite el acceso a la red de un modo conveniente, ubicuo y bajo demanda a un grupo de recursos computacionales configurables que puede ser fácilmente abastecido y publicado con un esfuerzo mínimo de gestión o interacción con el proveedor de servicios.

Erl [23] simplifica la definición de *cloud computing* como una forma especializada de computación distribuida que presenta modelos de utilización remota de recursos medibles y escalables.

[24], citados por [25], explican que *cloud computing* puede ser visto como servicios habilitados para la red que proveen otros de calidad garantizada y escalables según la demanda y que pueden ser accedidos a través de internet. Para ser compartidos a través de la red se utiliza la infraestructura proporcionada por el proveedor de servicio en la nube

(CSP, del inglés *Cloud Service Provider*); de este modo, el consumidor del servicio paga según la demanda, es decir el costo es progresivo según el uso que le dé. Así, *cloud computing* permite un nivel de abstracción entre los recursos computacionales requeridos y la arquitectura fundamental del sistema [24].

Según [22], las características primordiales de *cloud computing* son:

- Autoservicio bajo demanda: no es necesaria la interacción humana para obtener información de ciertos proveedores de servicios.
- Amplio acceso a la red: distintos dispositivos pueden acceder a la red utilizada y sus recursos.
- Agrupamiento de recursos: en base a sus requerimientos, diversos inquilinos de servicios pueden personalizar los servicios.
- Elasticidad rápida: la respuesta de los CSP debe ser rápida y escalable según las solicitudes de los consumidores.
- Servicio medido: los CSP deben contar con una capacidad de medición como parte del sistema para garantizar el control del uso de recursos.

### **2.2.2. Marketing digital**

El marketing digital representa la evolución de las técnicas de mercadotecnia, en la que los clientes y vendedores interactúan a través de tecnologías que van más allá del contacto físico, aprovechando artefactos y técnicas digitales con el fin de relacionarse y lograr una venta, interés o posicionar un producto o servicio [26].

De acuerdo con Internet World Stats [27], citado por [28], el desarrollo del ancho de banda en 2015 había incrementado la velocidad y uso de internet, así como las expectativas del usuario en casi 40% del mundo y casi 90% en varios países. Para el 2019 los

porcentajes se han acrecentado hasta un 58.8% de alcance a nivel global [29]. Por este motivo, [28] hace énfasis en el uso de canales digitales para la aplicación de técnicas de marketing.

[30] cita a Kotler, Kartajaya y Setiawan [31] y explica que el marketing digital debe ser adaptativo a la naturaleza dinámica del comportamiento del consumidor, por lo que se debe guiar al cliente desde la etapa inicial de captación hasta la recomendación de la marca a otros compradores.

#### **2.2.2.1. Conversión de *leads***

El concepto de *lead* difiere de empresa en empresa según la naturaleza o rubro de esta. Sin embargo, [32] sostiene que todas las definiciones concluyen en que un *lead* representa el arquetipo de cliente, dicho de otro modo, un prospecto del consumidor potencial. Por este motivo, los encargados de ventas en una compañía están a cargo de convertir el mayor número posible de *leads* para obtener la mejor tasa de conversión posible.

Eisenberg [33] describe las tasas de conversión como una medida de la habilidad para persuadir a los visitantes de una página a tomar las acciones que la compañía espera que tomen. Las metas del consumidor deben ser alcanzadas antes que las de la empresa; solo así se refleja la efectividad y satisfacción del cliente.

[34] enfatiza en el seguimiento de la conversión de *leads* a través de herramientas como *Google analytics* y *AdWords*, mapas de calor, NPS (del inglés *Net Promoter Score*) o encuestas de satisfacción al cliente, entre otros. Sin embargo, antes de estas actividades debe definirse qué es lo que representa conversión para la compañía, teniendo claro las metas del negocio y de las métricas de la aplicación.

#### **2.2.2.2. Lead nurturing**

Con el objetivo de convertir la mayor cantidad posible de *leads* en clientes finales se aplican diversas técnicas elegidas por la empresa de modo que las relaciones entre *lead* y organización se estrechen y fortalezcan. A este grupo de actividades se les conoce como *lead nurturing* [35].

Marketo [36], una compañía de Adobe, profundiza en la valorización de *leads* (*lead scoring*) como actividad previa al *lead nurturing*, considerando aspectos tales como las características demográficas, firmográficas, económicas y otras; el interés del *lead*, comportamiento y tiempo de conversión (tiempo de procesamiento) esperado. También se menciona a la automatización de marketing como tecnología útil para reducir el esfuerzo y especializarlo.

En [28], [31] y [34] se menciona que el *lead nurturing* personalizado es una de las herramientas actualmente más efectivas para la conversión de *leads*, precisando de segmentación a través de estudios de mercado o *data analytics*. En este sentido, el acercamiento particular para cada *lead* resulta altamente relevante a través de mensajes de texto, correos personalizados o promociones de productos ‘a la medida’.

#### **2.2.3. Minería de datos**

*Data mining* o minería de datos es un proceso de generación de conocimiento a través del descubrimiento de patrones en grandes volúmenes de datos mediante modelos probabilísticos, estadísticas, computación gráfica y otras técnicas [37]. Su objetivo es la extracción de información anteriormente desconocida para la organización, de modo que se brinde soporte a la toma de decisiones del negocio [38]. Las fuentes de datos incluyen, pero no se limitan a bases de datos, *data warehouses*, web y otros repositorios de información.

[39], referenciados por [40], clasifican los algoritmos de minería de datos en dos grupos:

- Algoritmos supervisados o dirigidos: predicen el valor de un atributo o etiqueta a través de otros ya conocidos previamente. Este proceso cuenta con las etapas de entrenamiento y prueba.
- Algoritmos no supervisados o no dirigidos: en lugar de usar datos históricos analizan actuales, identificando patrones y tendencias; esto debido a un grado insuficiente de madurez de la aplicación.

En [40] se presentan algunos algoritmos clasificados por el tipo de aprendizaje generado, como se muestra en la Tabla I.

TABLA I: CLASIFICACIÓN DE TÉCNICAS DE MINERÍA DE DATOS [40]

<b>Supervisados</b>	<b>No supervisados</b>
Árboles de decisión	Detección de desviaciones
Inducción neuronal	Segmentación
Regresión	Agrupamiento (“clustering”)
Series Temporales	Reglas de asociación
	Patrones Secuenciales

### 2.2.3.1. Técnicas dirigidas

- A. Árboles de Decisión: son herramientas de análisis para estructurar y evaluar situaciones inciertas. Esta técnica considera alternativas para diversas situaciones, asociando eventos probabilísticos en base a un proceso de decisión, así como los nodos finales o resultados para cada ruta [41].
- B. Redes Neuronales: son grupos no lineales de unidades interconectados y organizados, usualmente desarrollados bajo modelos matemáticos que son llevados a cabo bajo algoritmos digitales para reconocimiento de patrones, aprendizaje asociativo, predicción de series de tiempo, segmentación y más [42].

- C. Regresión: es una técnica estadística que realiza predicciones sobre una variable dependiente a partir de otras independientes, estando interrelacionadas entre sí [43].
- D. Series Temporales: son cadenas de valores analizados durante un periodo de tiempo determinado, cronológicamente ordenados. Para esta técnica es necesario analizar tendencias, estacionalidad y otras fluctuaciones irregulares, todo esto a través de modelos estadísticos [44].

#### **2.2.3.2. Técnicas no dirigidas**

- A. Detección de Desviaciones: detecta las diferencias – (desviación) entre un conjunto de datos respecto a un patrón normal de comportamiento de datos de entrenamiento, filtrando grandes cantidades de datos hasta determinar qué tan significativa es la desviación considerada [45].
- B. Segmentación: separa la totalidad de datos en subclases (exhaustivas o jerárquicas), pudiendo aplicar la técnica de clusterización. Esta técnica permite al usuario determinar los subconjuntos a través de tableros u otras herramientas visuales [46].
- C. Agrupamiento o clustering: divide la totalidad de datos en un número de grupos (algoritmo de “k-means”), detectando automáticamente clústeres mutuamente excluyentes para así agrupar los datos por características similares [43].
- D. Reglas de Asociación: establece un conjunto de reglas entre los objetos de una base de datos, identificando patrones de comportamiento a través de asociaciones entre dichos registros [47].
- E. Patrones Secuenciales: basándose en técnicas que evalúan el grado de similitud entre características de

datos, reconoce patrones en conjuntos de datos representados secuencialmente y sus atributos [48].

### 2.2.3.3. Herramientas de minería de datos

- A. Orange: es un software de minería de datos para expertos y principiantes que aplica algoritmos de machine learning para generar conocimiento a través de tableros que permiten la visualización interactiva de los datos. [49]
- B. WEKA: El entorno para el análisis del conocimiento de la universidad de Waikato (WEKA, del inglés *Waikato Environment for Knowledge Analysis*) es un conjunto de algoritmos de machine learning para desarrollar tareas de minería de datos. Utiliza herramientas para la preparación de datos, clasificación, regresión, clustering y reglas de asociación, así como la visualización de estos [50]. WEKA también brinda soporte para *deep learning*.
- C. Rapidminer: es un software utilizado para la aplicación de técnicas de minería de datos, machine learning, modelado de operaciones y preparación de datos. La especialidad de este programa radica en su análisis predictivo [51]. Forrester [52] y Gartner [53] describen a RapidMiner como una de las herramientas líderes para soluciones de machine learning y data science a través del análisis predictivo.

Ionos [54] realiza una comparación entre las tres aplicaciones descritas para minería de datos, concluyendo con lo mostrado en la Tabla II.

TABLA II: COMPARACIÓN ENTRE LAS HERRAMIENTAS DE MINERÍA DE DATOS  
RAPIDMINER, WEKA Y ORANGE [54]

	Características	Lenguaje de programación	Sistema operativo	Precio/Licencia
Rapidminer	Apto para todos los procesos, destaca en el análisis predictivo	Java	Windows, macOS, Linux	Freeware, diferentes versiones de pago
WEKA	Muchos métodos de clasificación	Java	Windows, macOS, Linux	Software libre (GPL)
Orange	Crea una visualización de datos atractiva sin que se requieran muchos conocimientos previos para ello.	Núcleo del software: C++, ampliación y lenguaje de entrada: Python	Windows, macOS, Linux	Software libre (GPL)

#### 2.2.3.4. Metodologías de minería de datos

A. KDD: El Modelo de Descubrimiento de Conocimiento en Bases de Datos (KDD, del inglés *Knowledge Discovery in Databases*) propuesto por Fayyad, Piatetsky-Shapiro y Smyth [55] es un proceso interactivo e iterativo que comprende cinco etapas:

1. Selección de datos: previa identificación y entendimiento de la realidad del proyecto así como sus objetivos y metas, se extraen los datos elegidos para el proceso.

2. Preparación de datos: se limpian y procesan los datos a través de técnicas de corrección.
3. Transformación de datos: se reducen y agrupan los datos según el interés del responsable del proyecto para posteriormente almacenarlos en una estructura arquitectónica adecuada.
4. Data mining: se aplican los algoritmos de minería de datos para descubrir patrones.
5. Interpretación y evaluación: tras la generación de conocimiento, este es validado y presentado visualmente para los dueños del negocio.

B. SEMMA: Es un proceso propuesto inicialmente por el instituto SAS, el cual corresponde a las siglas del inglés Sample, Explore, Modify, Model y Assess (muestra, exploración, modificación, modelado y evaluación en español). [56] citan a Olson y Delen [43] y explican que SEMMA facilita la exploración estadística y visualización de técnicas a través de la selección y transformación de variables predictivas relevantes, de modo que se obtenga un modelo y resultados mucho más precisos. [56] y [57] detallan las siguientes fases:

1. *Sample* (Muestra): se elige una fracción de la totalidad de datos que contenga información significativa que facilite su manipulación.
2. *Explore* (Exploración): se busca tendencias y anomalías para entender la naturaleza de los datos.
3. *Modify* (Modificación): se crean, seleccionan y transforman variables para el modelo que se elegirá en la próxima etapa.
4. *Model* (Modelado): se aplican los algoritmos elegidos de minería de datos.
5. *Assess* (Evaluación): se evalúa el grado de confiabilidad de los resultados del proceso de

minería de datos y se estima el grado de performance de este.

C. CRISP-DM El Modelo de Proceso Estándar para Minería de Datos (CRISP-DM, del inglés *Cross Industry Standards Process for Data Mining*), provee a un proyecto de seis etapas no estrictamente lineales que interactúan durante todas sus fases de desarrollo. Según Chapman et al. [58], dichas fases son:

1. Comprensión del negocio: el foco es la comprensión de objetivos y requisitos de proyecto.
2. Comprensión de los datos: tras una recopilación inicial de datos, identifica la calidad de datos y establece relaciones claras para definir una hipótesis de relación respecto al atributo oculto.
3. Preparación de los datos: consiste en la preparación de datos (siguiendo modelos establecidos) para dejarlos listos a la implementación de técnicas de minería de datos.
4. Modelado: se eligen y aplican las técnicas de modelado, asegurando que los valores sean óptimos.
5. Evaluación: se evalúa el modelo elegido y los pasos realizados para la construcción del modelo, asegurando que los objetivos del negocio fueron cumplidos.
6. Implantación: Se documenta y presenta el conocimiento adquirido para el uso del cliente, elaborando un plan de implantación, monitoreo y mantenimiento.

KD Nuggets realiza encuestas desde el 2002 respecto a las metodologías utilizadas para proyectos de minería de datos, analytics o data science. Piatetsky-Shapiro en la

última, aplicada en 2014 [59], concluye con CRISP-DM como la más utilizada, tal como se muestra en la Figura 1.

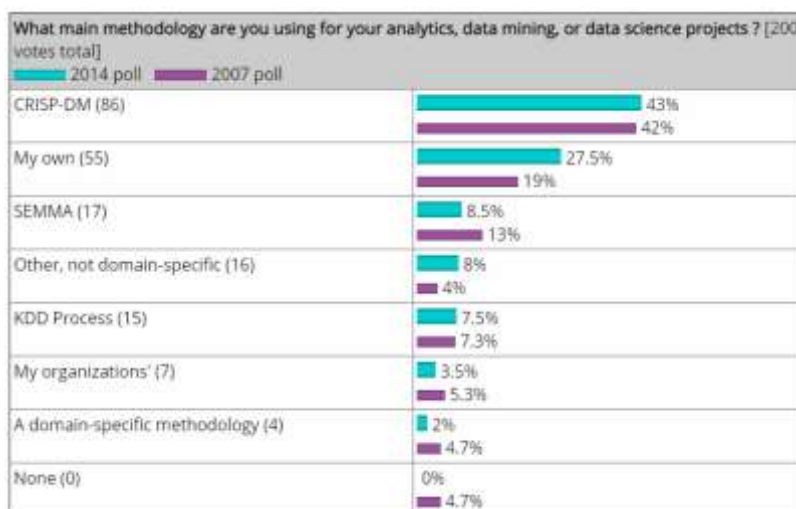


Fig. 1. Metodologías más utilizadas para minería de datos, *analytics* y otros proyectos de *data science*. [59]

En complemento, [57] realizan un trabajo comparativo y revisión de literaturas para establecer correspondencias entre los procesos KDD, SEMMA y CRISP-DM, obteniendo lo observado en la Tabla III. Se concluye que CRISP-DM es la metodología más completa, pero que el investigador puede integrar en SEMMA algunos pasos de KDD no incluidos en su modelo de proceso.

TABLA III: RESUMEN DE LAS CORRESPONDENCIAS ENTRE KDD, SEMMA Y CRISP-DM [57]

KDD	SEMMA	CRISP-DM
Pre KDD	-----	Comprensión del negocio
Selección	Muestra	Comprensión de los datos
Pre procesamiento	Exploración	
Transformación	Modificación	Preparación de los datos
Minería de datos	Modelado	Modelado
Interpretación/ Evaluación	Evaluación	Evaluación
Post KDD	-----	Implantación

Las metodologías para minería de datos tienen un enfoque ágil en su desarrollo y presentan un flujo claro para la generación de conocimiento; sin embargo, estas no son

suficientes para desarrollar software, puesto que no consideran componentes, arquitectura, despliegue y otros. Por este motivo es importante considerar otra metodología de desarrollo de software como complemento.

## **2.2.4. Metodologías de desarrollo de software**

### **2.2.4.1. CommonKads**

Bajo definición de Lagos [60] esta metodología ágil es aplicada para el desarrollo de sistemas basados en conocimiento (SBC), mediante modelos, cada uno con una serie de formularios que facilitan su implementación, interrelacionados entre sí para obtener rasgos del software y del entorno de este.

Los modelos de CommonKads se explican a continuación:

- A. Modelo de Organización (OM): con el propósito de adquirir conocimiento general de la situación y entorno de la organización.
- B. Modelo de Tareas (TM): describe las tareas realizadas en el entorno de la organización sobre la cual se implementará el sistema, proporcionando un marco para distribución de tareas.
- C. Modelo de Agente (AM): describe las capacidades y características de los ejecutores de tareas o agentes.
- D. Modelo de Comunicación (CM): describe el intercambio de datos e información entre los agentes involucrados durante la ejecución de tareas en el modelo de tareas.
- E. Modelo del Conocimiento de la Experiencia (EM): modela el conocimiento de la resolución de problemas determinados, por un agente que realiza una tarea.
- F. Modelo de Diseño (DM): describe la arquitectura y diseño técnico del sistema hasta su implementación.

#### **2.2.4.2. RUP**

Proceso Unificado Racional (RUP, del inglés *Rational Unified Process*), es una metodología de desarrollo de software basado en UML. En términos de [61], este proceso provee disciplina en la asignación de tareas y responsabilidades en una organización de desarrollo, asegurando la alta calidad de un producto de software que satisfaga las necesidades del usuario final. Cubre cinco etapas principalmente: modelado de negocio, captura de requisitos, análisis y diseño, implementación, pruebas y distribución. Asimismo, extiende tres flujos de soporte: administración de configuración y cambios, administración de proyecto y administración de entorno.

### III. METODOLOGÍA

#### 3.1. Tipo y nivel de investigación

##### 3.1.1. Tipo de investigación

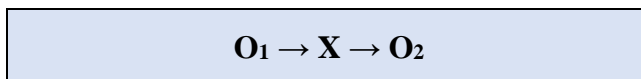
La presente corresponde a una investigación tecnológica aplicada puesto que sus resultados se orientan al desarrollo de una aplicación web que soporte el proceso de conversión de *leads* dentro de la problemática descrita previamente.

##### 3.1.2. Nivel de investigación

Cuantitativa preexperimental.

#### 3.2. Diseño de investigación

En base a los diseños establecidos por Hernández [62], para la presente investigación se utilizó el diseño de contrastación de tipo preprueba/posprueba con un solo grupo. Este diseño se diagrama a continuación:



Donde:

**O<sub>1</sub>** = Medición diagnóstica de la realidad problemática a través de encuestas y entrevistas para encontrar dificultades en la toma de decisiones dentro del proceso de conversión de *leads* por parte de los asociados, y necesidades actuales de los estudiantes.

**X** = Aplicación del software

**O<sub>2</sub>** = Medición del impacto generado por la propuesta sobre la realidad problemática. Debido al tiempo corto para la medición del efecto del software desarrollado, se procedió a aplicar la validación y aceptación de la propuesta mediante juicio de expertos, tanto de TI como directivos de la organización.

Para la preprueba y posprueba se evaluó a un grupo de personas que por su desempeño y relación con la organización fueron seleccionados. Dichos conjuntos son descritos en el inciso correspondiente a la población.

### 3.3. Población, muestra y muestreo

#### 3.3.1. Población

Considerando los aspectos mencionados en las observaciones de preprueba y posprueba, se tomó en cuenta dos grupos para la población, siendo estos:

- ✓ **Población 1** = 9245 *leads* no convertidos. Este número representa la totalidad de personas que no lograron adquirir un programa de intercambio tras su registro en la plataforma.
- ✓ **Población 2** = 409 asociados de AIESEC en Perú. La integran miembros de las áreas de marketing al consumidor (B2C, del inglés *Business to Customer*), intercambios salientes (de los programas Voluntario, Emprendedor y Talento Global) y finanzas. Esta población está constituida por personas en todos los rangos jerárquicos de la asociación.

#### 3.3.2. Muestra

La muestra se ha obtenido haciendo uso de la fórmula del *procedimiento para estimar el tamaño de la muestra representativa para una población finita* definida por Bernal [63]:

$$n = \frac{Z^2 PQN}{E^2(N - 1) + Z^2 PQ}$$

Donde:

**n** = Tamaño de la muestra por estimar.

**Z** = Nivel de confianza o margen de confiabilidad (95%, es decir,  $Z = 1,96$ ).

**N** = Tamaño de la población (Número).

**P** = Proporción (intervalo, calculado a partir de los datos de la muestra, en el cual nosotros “confiamos” se encuentra la proporción de la población. En este caso  $P = 0.5$ ).

**Q** =  $1 - P = 0.5$ .

**E** = Error de estimación (diferencia máxima entre la proporción muestral y la proporción proporcional que el investigador está dispuesto aceptar en función del nivel de confianza definido para el estudio. En este caso  $E = 0.05 * 5\%$ ).

✓ **Muestra 1**

La muestra de la población de estudiantes estuvo constituida por:

$$n = \frac{1.96^2 * 0.5 * 0.5 * 9245}{0.05^2(9245 - 1) + 1.96^2 * 0.5 * 0.5}$$

n = 369 individuos

✓ **Muestra 2**

La muestra de la población 2 estuvo constituida por:

TABLA IV: DISTRIBUCIÓN DE LA MUESTRA 2

<b>Rol</b>	<b># personas</b>
Directores nacionales	6
Gerentes locales	16
Directores locales	60
<b>Total</b>	<b>82</b>

### 3.3.3. Muestreo

Se aplicaron dos técnicas distintas de muestreo para determinar las cantidades de las muestras. Para el caso de estudiantes fue probabilístico de tipo aleatorio simple, mientras que para los asociados se eligió un no probabilístico de tipo discrecional mediante juicio de expertos.

### 3.4. Criterios de selección

La selección de la muestra de estudiantes fue aleatoria a través de la generación aleatoria de códigos mediante una herramienta informática. Acorde al muestreo discrecional, la selección de la muestra de asociados estuvo determinada por el juicio de los expertos del negocio, los cuales consideraron a aquellos miembros asociados dentro de los sectores tácticos y estratégicos (directores locales y gerentes nacionales, respectivamente), puesto que se requería que los individuos evaluados se encontraran en la capacidad de aportar en gran medida a los resultados operativos de la asociación a través de la toma de decisiones.

### 3.5. Operacionalización de variables

Las variables que se han utilizado como elementos básicos en el desarrollo de la hipótesis están identificadas de la siguiente manera:

### **3.5.1. Variables**

#### **3.5.1.1. Variable independiente**

Aplicación web

#### **3.5.1.2. Variable dependiente**

Proceso de conversión de leds

### 3.5.1.3.Indicadores

TABLA V: INDICADORES

Objetivo específico	Indicador(es)	Definición conceptual	Unidad de medida	Instrumento	Definición operacional
Reducir el tiempo de respuesta a <i>leads</i> tras su registro en la plataforma.	Tiempo de respuesta	Tiempo en que demora un miembro asociado en contactar a un <i>lead</i> después de haberse registrado en la plataforma online de la asociación.	Número de horas	Encuesta	Promedio de tiempo procesamiento en preprueba – Promedio de tiempo de procesamiento en posprueba
Segmentar <i>leads</i> a través de técnicas de minería de datos.	Número óptimo de grupos	Cantidad de grupos o segmentos de <i>leads</i> clasificados por perfiles.	Índice de Davies–Bouldin	Reporte del software utilizado	$\frac{1}{n_c} \sum_{i=1}^n R_i$ Donde $R_i = \max_{1 \leq n_c, i \neq j} R_{ij}$ , $R_{ij} = (S_i + S_j)/d_{ij}$ $S_i$ = distancia máxima entre los centros de clúster $i$ y $j$
Incrementar la calidad de información brindada a los directivos de la organización para la toma de decisiones estratégicas.	Nivel de aceptación de información	Nivel de calidad de la información provista por el sistema para los directivos y miembros de la asociación.	Grado de aceptación	Encuesta	Promedio de grado de aceptación en preprueba – Promedio de grado de aceptación en posprueba
Incrementar la calidad de la información presentada a <i>leads</i> en el acompañamiento digital del proceso de conversión.	Nivel de aceptación de información	Nivel de calidad de la información provista por el sistema para los consumidores.	Grado de aceptación	Encuesta	Promedio de grado de aceptación en preprueba – Promedio de grado de aceptación en posprueba
Verificar la calidad de la aplicación	Disponibilidad, usabilidad, portabilidad, testeabilidad	Idoneidad de la aplicación desarrollada respecto a su diseño, mantenibilidad, experiencia del usuario y otros requisitos no funcionales.	Juicio de expertos	Lista de cotejos	Validación de la aplicación mediante el juicio de expertos y la confiabilidad de este a través del alfa de Cronbach

### 3.6. Técnicas e instrumentos de recolección de datos

A continuación, se detallan las técnicas e instrumentos empleados para la recolección de datos.

TABLA VI: TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS

Técnicas	Instrumentos	Elementos de la población	Propósito
Entrevista	Cuestionario	Directores nacionales de operaciones, marketing y gerente general	Identificar los problemas existentes en la población durante el proceso de conversión de <i>leads</i> .
Juicio de expertos	Lista de cotejos	Expertos de TI	Verificar la calidad del software al culminar su desarrollo
Encuesta	Ficha de encuesta	Directores nacionales y locales	Realizar la preprueba y posprueba, obteniendo datos válidos y confiables para posterior análisis y procesamiento.

### 3.7. Procedimientos

#### 3.7.1. Metodología de desarrollo

Se decidió realizar un híbrido entre las metodologías CRISP-DM para la minería de datos, selección y aplicación de algoritmos inteligentes y RUP para el desarrollo del sistema y sus funcionalidades.

Para la primera metodología se ejecutaron todas sus fases, y para cubrir los modelos no cubiertos por esta se aplicó RUP, considerando la quinta iteración de diseño y unificando la sexta de implantación y prueba con las dos últimas de CRISP-DM: evaluación e implantación.

A continuación, se enumeran las actividades realizadas por cada iteración de la hibridación entre las metodologías descritas:

#### 1. Iteración #1: Comprensión del negocio (CRISP-DM)

En esta iteración se desarrollaron las siguientes actividades:

- ✓ Determinación de los objetivos del negocio.
- ✓ Evaluación de la situación.
- ✓ Determinación de los objetivos de DM.
- ✓ Producción de un plan de proyecto.

## **2. Iteración #2: Comprensión de los datos (CRISP-DM)**

En esta iteración se desarrollaron las siguientes actividades:

- ✓ Recolección de datos iniciales.
- ✓ Descripción de los datos.
- ✓ Exploración de los datos.
- ✓ Verificación de la calidad de los datos.

## **3. Iteración #3: Preparación de los datos (CRISP-DM)**

En esta iteración se desarrollaron las siguientes actividades:

- ✓ Selección de datos.
- ✓ Limpieza de los datos.
- ✓ Estructuración de los datos.
- ✓ Integración de los datos.
- ✓ Formateo de los datos.

## **4. Iteración #4: Modelado (CRISP-DM)**

En esta iteración se desarrollaron las siguientes actividades:

- ✓ Selección de la técnica de modelado.
- ✓ Generación del plan de prueba.
- ✓ Evaluación del modelo.

## **5. Iteración #5: Diseño (RUP)**

En esta iteración se desarrollaron las siguientes actividades:

- ✓ Diagrama de contexto de diseño.
- ✓ Diagrama de clases general.
- ✓ Diseño de base de datos.
- ✓ Diagramas de estados.
- ✓ Diseño de interfaces.
- ✓ Diseño de la arquitectura.

## **6. Iteración #6: Evaluación (CRISP-DM)**

En esta iteración se desarrollaron las siguientes actividades:

- ✓ Evaluación de los datos.
- ✓ Proceso de revisión.
- ✓ Determinación de futuras fases.

## **7. Iteración #7: Implementación y Prueba (CRISP-DM y RUP)**

En esta iteración se desarrollaron las siguientes actividades:

- ✓ Diagrama de componentes (RUP).

- ✓ Plan de implantación (CRISP-DM).
- ✓ Monitorización y mantenimiento (CRISP-DM).
- ✓ Informe final (CRISP-DM).
- ✓ Pruebas unitarias y de integración (RUP).
- ✓ Revisión del proyecto (CRISP-DM).

### **3.7.2. Análisis de riesgos**

El análisis de riesgos en el desarrollo de la presente tesis se efectuó con la finalidad de identificar las fases, entregables y objetivos afectados durante desarrollo de la presente tesis, las mismas de detallan en el *Anexo N° 02*.

### **3.7.3. Producto acreditable**

#### **1. Interfaces**

Se construyeron las interfaces del sistema web haciendo uso del lenguaje PHP con la plantilla de bootstrap de AdminLTE. Estas se presentan en el *ítem 4.1.5. Iteración #5: Diseño, sección Diseño de interfaces, en el Capítulo IV. Resultados.*

#### **2. Arquitectura**

De diseñó una arquitectura idónea para el funcionamiento del sistema web, el cual se detalla en el *ítem 4.1.5. Iteración #5: Diseño, sección Diseño de la arquitectura, en el Capítulo IV. Resultados.*

#### **3. Infraestructura tecnológica**

Considerando la arquitectura anteriormente descrita, se definen las características de cada uno de sus componentes en el *ítem 4.1.6. Iteración #7: Implementación y prueba, sección Diagrama de componentes, en el Capítulo IV. Resultados.*

### **3.7.4. Manual de usuario**

Se elaboró un manual de usuario con la finalidad de ayudar a los usuarios tanto nuevos como expertos (en la asociación) respecto al uso de la aplicación web implementada, el cual se muestra en el *Anexo N° 04.*

### **3.8. Plan de procesamiento y análisis de datos**

El procesamiento de los datos extraídos de los repositorios para el desarrollo del producto acreditable fue multivariado a través de la herramienta Rapidminer que automatizó los algoritmos a través de métodos descriptivos. Por otro lado, para el desarrollo metodológico de la presente investigación, el procesamiento posterior a la recolección de datos de la preprueba y posprueba a través de las encuestas y entrevistas mediante herramientas digitales fue también mecánico, utilizando un análisis estadístico a través de Microsoft Excel 2016 y *Google Sheets* para comparar los resultados a través de tabulación univariada y determinar el cumplimiento o no de los objetivos de la investigación.

### 3.9. Matriz de consistencia

TABLA VII: MATRIZ DE CONSISTENCIA

<b>PROBLEMA</b>	<b>OBJETIVOS</b>	<b>HIPÓTESIS</b>	<b>VARIABLES</b>
<b><u>FORMULACIÓN DEL PROBLEMA</u></b>	<b><u>OBJETIVO GENERAL</u></b>	<b><u>HIPÓTESIS</u></b>	<b><u>VARIABLES DE ESTUDIO</u></b>
¿Cómo se puede apoyar el proceso de conversión de <i>leads</i> en la asociación AIESEC en Perú?	Desarrollar un sistema web que apoye el proceso de conversión de <i>leads</i> en la asociación AIESEC en Perú.	La implementación de un sistema web basado en algoritmos de minería de datos apoya el proceso de conversión de leads en la asociación AIESEC en Perú	<b>VARIABLE INDEPENDIENTE</b> Aplicación web  <b>VARIABLE DEPENDIENTE</b> Proceso de conversión de <i>leads</i>
<b><u>OBJETIVOS ESPECÍFICOS</u></b>	<b><u>DESCRIPCIÓN DEL LOGRO DE LOS OBJETIVOS ESPECÍFICOS</u></b>		<b><u>INDICADORES</u></b>
Reducir el tiempo de respuesta a <i>leads</i> tras su registro en la plataforma.	Se reduce el número de horas que tarda la organización en contactar a un <i>lead</i> y ofrecerle programas de intercambio según su necesidad e intereses.		Tiempo de respuesta
Segmentar <i>leads</i> a través de técnicas de minería de datos.	Se obtienen clústeres que segmenten óptimamente los perfiles de consumidor en base a factores críticos que determinan la adquisición de un programa de intercambio.		Número óptimo de clústeres
Incrementar la calidad de información brindada a los directivos de la organización para la toma de decisiones estratégicas.	Se presenta información a través de tableros que permita mejorar el apoyo a la toma de decisiones tácticas y estratégicas en la asociación.		Nivel de aceptación de la información
Incrementar la calidad de la información presentada a <i>leads</i> en el acompañamiento digital del proceso de conversión	Se mejora la información brindada por la organización, estrechando así la relación entre la asociación y los <i>leads</i> a través del soporte digital en personalización de contenidos, <i>mailing</i> y mensajes de texto.		Nivel de aceptación de la información
Verificar la calidad de la aplicación	Se garantiza la calidad de la aplicación desarrollada frente a los indicadores de usabilidad, disponibilidad y testeabilidad.		Juicio de expertos

### **3.10. Consideraciones éticas**

Para el desarrollo de la presente investigación se tomó en cuenta las políticas internas de gestión de datos dentro de la asociación, los cuales se encuentran regulados por la Ley N° 29733 de Protección de Datos Personales del Perú; de este modo se garantizó la confidencialidad de los datos correspondientes a los *leads* obtenidos por los repositorios de datos utilizados. Asimismo, para asegurar la protección y bienestar de los miembros de la muestra (tanto estudiantes como miembros asociados) y la seguridad de sus datos brindados se consideró:

- ✓ Anonimato de cada una de las encuestas y entrevistas aplicadas tanto en la preprueba como posprueba.
- ✓ Resguardo de los datos registrados en las pruebas de validación del sistema por parte de los estudiantes, específicamente los datos sensibles de correo electrónico y número de celular.

Debido a las razones expresadas, los nombres presentes en las interfaces del sistema han sido cubiertos, garantizando la confidencialidad otorgada al autor de la investigación.

## **IV. RESULTADOS**

### **4.1. En base a la metodología utilizada**

#### **4.1.1. Iteración #1: Comprensión del negocio (CRISP-DM)**

##### **4.1.1.1. Determinación de los objetivos del negocio**

AIIESEC en Perú es una asociación civil sin fines de lucro que busca desarrollar liderazgo a través de intercambios al extranjero para un público de jóvenes entre 18 y 30 años con el fin de alcanzar el entendimiento intercultural que lleve a la paz mundial y el desarrollo del potencial humano; los destinos de impacto se distribuyen en más de 120 países y territorios alrededor del mundo y cuentan con tres programas para realizar el intercambio: Voluntario Global, para ejecutar proyectos de voluntariado internacional; Emprendedor Global, para realizar prácticas preprofesionales en startups; y Talento Global, para participar de pasantías profesionales en empresas. Dichos programas buscan generar impacto global a través de su relación con la agenda 2030 y los Objetivos de Desarrollo Sostenible (ODS).

Diariamente, AIIESEC busca enganchar a jóvenes peruanos para tomar estos programas a través de marketing físico y digital a lo largo de sus sedes en distintas regiones del país. A aquellos interesados se les solicita datos relevantes como su carrera y ciclo de estudios, sexo, mes y país de interés, entre otros.

Como asociación civil, AIIESEC no persigue el lucro; sin embargo, como todo organismo con estas características, genera ingresos para reinvertirlos en la mejora de sus procesos internos y programas que ofrece al público. Para mejorar esta rentabilidad la asociación aspira a llegar a más jóvenes que sean parte de sus programas, la mejora de la conversión entre los interesados (llamados *leads*, en marketing digital) y contribuyentes (equivalente a clientes), la aceleración del tiempo en que tarda esta

conversión, entre otros. En parte, uno de los principales problemas que enfrenta la asociación es llegar a su público de un modo efectivo. Los objetivos anuales se centran en las cifras descritas previamente como factores clave, incrementando su meta anualmente en base al entorno y situación actual de cada una de sus oficinas locales, resultando así en criterios cuantitativos de éxito.

#### **4.1.1.2. Evaluación de la situación**

El problema descrito es conocido por la asociación, la cual aplica análisis básicos para determinar cómo orientar sus estrategias de marketing físico o tipificar un perfil idóneo denominado *buyer persona*. Se han aplicado estudios de mercado para identificar las características de sus potenciales *leads*.

Actualmente, la asociación no cuenta con un sistema propio que recolecte los datos sensibles de las personas que se registran mediante su página web. Asimismo, AIESEC en Perú se encuentra en el proceso de migración entre plataformas tercerizadas de gestión de datos. Debido a ello, no cuentan con datos normalizados u ordenados y almacenados en una única base de datos o en una sola plataforma; no obstante, la data que se tiene actualmente, aunque desordenada, supera la decena de miles de registros, suficiente para una exploración y aplicación de modelos de minería de datos y así resolver el problema que afronta la asociación.

Respecto a la relación costo-beneficio de la propuesta de minería de datos, los costos no se verían incrementados debido al software libre que se utilizará como herramienta de *data mining*: RapidMiner; por otro lado, el repositorio del sistema propuesto y su respectiva base de datos será alojado en servidores independientes a los que actualmente cuenta la asociación.

Los beneficios que esta propuesta trae son muchos tales como la mejora en la identificación de diversos perfiles de consumidor y sus subgrupos o clústeres, la limpieza y unificación de todos los datos no normalizados para futuros análisis, la mejora en las cifras de los factores críticos de éxito de los procesos internos de AIESEC en Perú, entre otros.

#### 4.1.1.3. Determinación de los objetivos de DM

En la presente etapa se explica el alineamiento entre los objetivos de la organización y los del proceso de minería de datos, de modo que se demuestre una correlación entre ellos. Dicha correspondencia se muestra en la Tabla VIII.

TABLA VIII: OBJETIVOS DE DM

Objetivos de negocio	Objetivos de DM
Planificar campañas de marketing para atraer <i>leads</i> .	Determinar los perfiles y subgrupos de <i>leads</i> no convertidos y consumidores
Entrenar a los miembros de la asociación en acercamiento al consumidor y elaboración de propuestas atractivas.	respecto a su capacidad y probabilidad de adquisición de los programas
Aplicar técnicas de <i>lead</i> nurturing para convertir leads.	Establecer ratios de conversión respecto a perfiles y subgrupos de consumidores
Reducir el tiempo de procesamiento en el ciclo de conversión de leads	Establecer tiempo de conversión respecto a perfiles y subgrupos de consumidores

#### 4.1.1.4. Producción de un plan de proyecto

Además del cronograma general de la presente investigación anexo, se detallan las etapas y actividades específicas de la minería de datos y desarrollo del software.

TABLA IX: PLAN DEL PROYECTO

<b>Etapa</b>	<b>Actividad</b>	<b># días</b>
<b>Etapa 01: Comprensión del negocio</b>	1. Determinación de objetivos del negocio	1
	2. Evaluación de la situación actual	1
	3. Determinación de los objetivos de la DM	1
<b>Etapa 02: Comprensión de los datos</b>	4. Recolección de datos iniciales	1
	5. Descripción de los datos	1
	6. Exploración de los datos	2
	7. Verificación de la calidad de los datos	3
<b>Etapa 03: Preparación de los datos</b>	8. Selección de datos	2
	9. Limpieza de los datos	14
	10. Estructuración de los datos	1
	11. Integración de los datos	1
	12. Formateo de los datos	1
<b>Etapa 04: Modelado</b>	13. Selección de la técnica de modelado	3
	14. Generación del plan de prueba	3
	15. Construcción del modelo	5
	16. Evaluación del modelo	5
<b>Etapa 05: Diseño</b>	17. Elaboración de diagramas	5
<b>Etapa 06: Desarrollo</b>	18. Desarrollo de módulo de registro	12
	19. Desarrollo de módulo de conversión	12
	20. Desarrollo de módulo de reportes	5
<b>Etapa 07: Evaluación</b>	21. Evaluación de los resultados	5
	22. Proceso de revisión	1
<b>Etapa 08: Implantación y Prueba</b>	23. Elaboración de diagramas	3
	24. Elaboración de plan de implantación	2
	25. Elaboración de plan de monitorización y mantenimiento	2
	26. Elaboración de informe final	7
	27. Ejecución de pruebas unitarias y de integración	3
	28. Revisión de proyecto	4

## 4.1.2. Iteración #2: Comprensión de los datos (CRISP-DM)

### 4.1.2.1. Recolección de datos iniciales

Tal y como se explicó, de primera mano la asociación no contaba con una sola base de datos uniforme, por lo que se procedió a extraer los datos mediante dos actividades:

- Exportación a formato .xlsx desde la plataforma Podio en los múltiples tableros que almacenaban los datos a través de las herramientas de la aplicación. Esta técnica se utilizó para extraer los datos de todos los *leads* generados a través de los procesos internos de AIESEC en Perú.

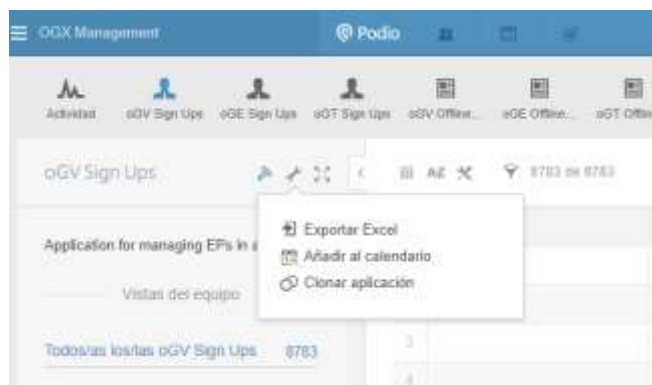


Fig. 2. Exportación de datos en la herramienta Podio

- Consulta a la base de datos utilizada a nivel global a través de un aplicativo de la asociación. Esta técnica se utilizó para extraer los datos de los consumidores convertidos tanto a través de los procesos internos de AIESEC en Perú como los registros directos a través del portal global de la organización.

El motivo de la extracción de los datos a través de distintos repositorios se debe a la falta de un único medio de almacenamiento para los datos, por lo que este método permitió la consulta y cotejamiento de ambas bases de datos para así tener información más precisa y completa para ejecutar los modelos de minería de datos; sin embargo, debido a que no tenían los mismos atributos, el cruce de datos se tornó aún más complicado. Todos los datos

recolectados se ubicaron en diferentes hojas de cálculo (formato .xlsx) agrupados según su naturaleza (principalmente programa de interés de los *leads*).

#### 4.1.2.2. Descripción de los datos

Todos los datos recolectados pertenecían a características de un *lead*, y tras unificar ambas fuentes, se encontraron los siguientes campos:

TABLA X: DESCRIPCIÓN DE LOS DATOS

<b>Campo</b>	<b>Tipo de dato</b>	<b>Descripción</b>
<b>EXPA ID</b>	Número entero	Representa el identificador único del <i>lead</i> en la base de datos.
<b>NOMBRE</b>	Cadena de texto	Nombres del <i>lead</i> .
<b>APELLIDO</b>	Cadena de texto	
<b>EMAIL</b>	Cadena de texto	Correo electrónico de registro y acceso a la plataforma.
<b>TELÉFONO</b>	Cadena de texto	Número de celular del <i>lead</i> para ser contactado.
<b>EDAD</b>	Número entero	Edad del <i>lead</i> a la fecha del registro, no se pide fecha de nacimiento.
<b>CAMPO DE ESTUDIO</b>	Cadena de texto	Carrera universitaria o técnica del <i>lead</i> .
<b>CICLO EN CURSO</b>	Cadena de texto	Ciclo en el que se encuentra el <i>lead</i> , se acepta el valor de egresado.
<b>EXPERIENCIA LABORAL</b>	Cadena de texto	Descripción de la experiencia laboral del <i>lead</i> , si existiese.
<b>FECHA DE VIAJE ESTIMADA</b>	Cadena de texto	Mes en el que el <i>lead</i> está interesado para tomar su intercambio.
<b>PROGRAMA DE INTERÉS</b>	Cadena de texto	Programa al cual el <i>lead</i> se registró para tomar su intercambio.
<b>DEPARTAMENTO</b>	Cadena de texto	Departamento del cual se registró el <i>lead</i> , no de su nacimiento.
<b>UNIVERSIDAD</b>	Número entero	ID de la Universidad o instituto del cual se registra el <i>lead</i> .

<b>MEDIO</b>	Cadena de texto	Medio por el cual conocieron sobre AIESEC y los llevó a inscribirse.
<b>COMITÉ</b>	Número entero	ID de la oficina local de AIESEC en Perú correspondiente a su registro.
<b>CONTACTADO</b>	Cadena de texto	Confirmación del contacto al <i>lead</i> .
<b>STATUS</b>	Cadena de texto	Etapas del <i>customer funnel</i> del <i>lead</i> : <i>Lead, Sign-up, In-Progress, Approved, Realized, Finished, Completed</i> .
<b>EP MANAGER</b>	Número entero	ID de la persona a cargo del <i>lead</i> dentro de la asociación.
<b>FECHA DE CONTACTO</b>	Fecha	Fecha en la cual se ejecutó el primer contacto al <i>lead</i> .
<b>FECHA DE APD</b>	Fecha	Fecha en la cual se confirmó el intercambio a través del pago y firma del convenio.
<b>FECHA DE RE</b>	Fecha	Fecha del primer día del intercambio.
<b>FECHA DE FI</b>	Fecha	Fecha del último día del intercambio.
<b>FECHA DE CO</b>	Fecha	Fecha de conclusión satisfactoria del programa, si es que lo fuese.
<b>DUPLICADO</b>	Cadena de texto	Identificador si es un dato duplicado o no en otra plataforma de la asociación
<b>PAÍS DE DESTINO</b>	Número entero	País donde se realizará el intercambio, si es que fuese confirmado.
<b>COMITÉ DE DESTINO</b>	Número entero	Oficina local responsable por el intercambio en el extranjero, si es que fuese confirmado.

Se recolectó un total de 12981 registros correspondientes a *leads* generados entre enero de 2017 a diciembre de 2018.

### 4.1.2.3. Exploración de los datos

Tras la unión de fuentes de información y la descripción de los datos, se procedió a explorarlos en el repositorio unificado a través de técnicas estadísticas para determinar su estructura y propiedades. A continuación, se presenta dicha exploración, cuya fuente son los datos obtenidos tras la extracción detallada en la fase de recolección inicial.

#### - Atributo edad

TABLA XI: DISTRIBUCIÓN DE FRECUENCIAS PARA EL ATRIBUTO EDAD EN EL CONJUNTO INICIAL DE LEADS

Edad	# registros
12	1
15	3
16	13
17	130
18	877
19	1000
20	1202
21	2164
22	1543
23	1131
24	1103
25	747
26	534
27	369
28	288
29	188
30	107
31	22
32	9
33	11
34	7
35	5
36	2
37	3
38	2
40	2
41	1
42	2

43	1
45	1
46	1
(en blanco)	28
<b>Total</b>	<b>11497</b>

# registros frente a Edad

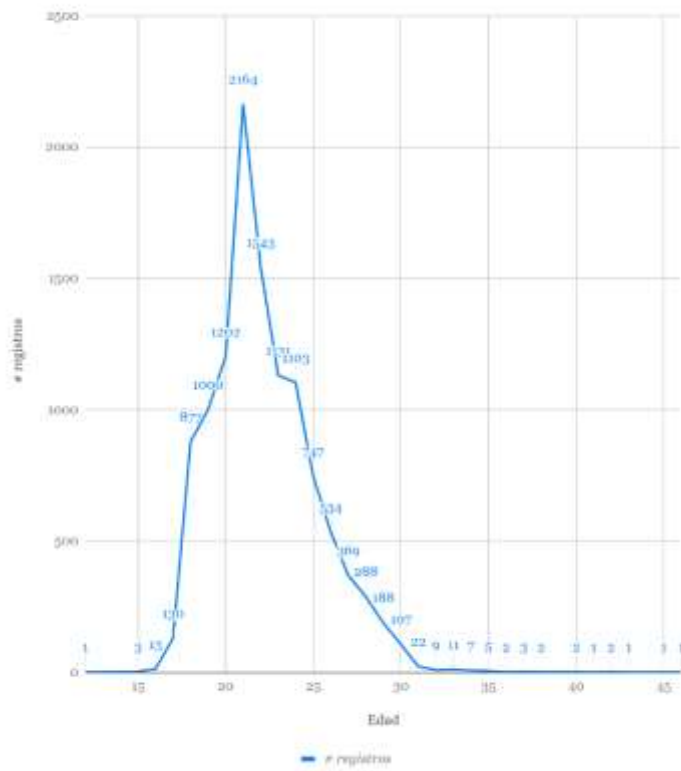


Fig. 3. Gráfico de líneas de la distribución de edades

Se observa que la mayor de distribución de datos se encuentra entre las edades de 19 a 24 años.

- Atributo departamento

TABLA XII: DISTRIBUCIÓN DE FRECUENCIAS PARA EL ATRIBUTO DEPARTAMENTO EN EL CONJUNTO INICIAL DE LEADS

<b>Departamento</b>	<b># registros</b>
Amazonas	8
Áncash	18
Arequipa	1133
Ayacucho	25
Cajamarca	44
Chimbote	9
Cusco	339
Huancayo	60
Huánuco	155
Huaraz	13
Ica	174
Juliaca	19
Junín	413
La Libertad	469
Lambayeque	362
Lima	6868
Loreto	44
Moquegua	1
Moquegua	3
Nueva Cajamarca	1
Otras ciudades	63
Otros departamentos	170
Piura	470
Puno	230
San Martín	251
Tacna	155
<b>Total</b>	<b>11497</b>

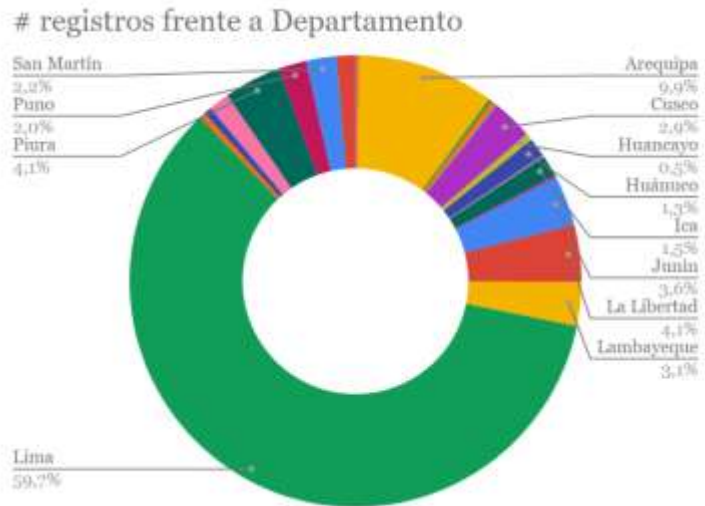


Fig. 4. Gráfico de anillo de la distribución de departamentos

Se observa que los departamentos con mayor volumen de datos son Lima y Arequipa.

- Atributo mes deseado

TABLA XIII: DISTRIBUCIÓN DE FRECUENCIAS PARA EL ATRIBUTO MES PROYECTADO PARA VIAJAR EN EL CONJUNTO INICIAL DE LEADS

Mes	# registros
Enero	5632
Febrero	853
Marzo	429
Abril	352
Mayo	300
Junio	306
Julio	828
Agosto	529
Setiembre	14
Octubre	234
Noviembre	206
Diciembre	1543
En más de 6 meses	1
Septiembre	270
<b>Total</b>	<b>11497</b>

# registros frente a Mes deseado

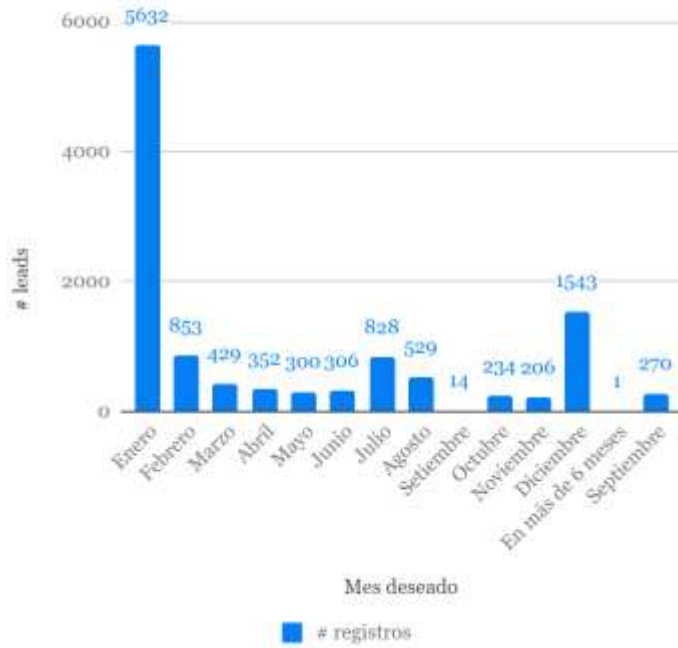


Fig. 5. Gráfico de columnas de la distribución de meses deseados para viajar

Se observa que los meses pico son julio, agosto, y de diciembre a febrero, fechas que coinciden con las vacaciones universitarias.

- Atributo ciclo

TABLA XIV: DISTRIBUCIÓN DE FRECUENCIAS PARA EL ATRIBUTO CICLO DE ESTUDIOS EN EL CONJUNTO INICIAL DE LEADS

Ciclo	# registros
1° - 3°	2651
4° - 6°	2338
7° - 10°	3102
Egresado	3406
<b>Total</b>	<b>11497</b>

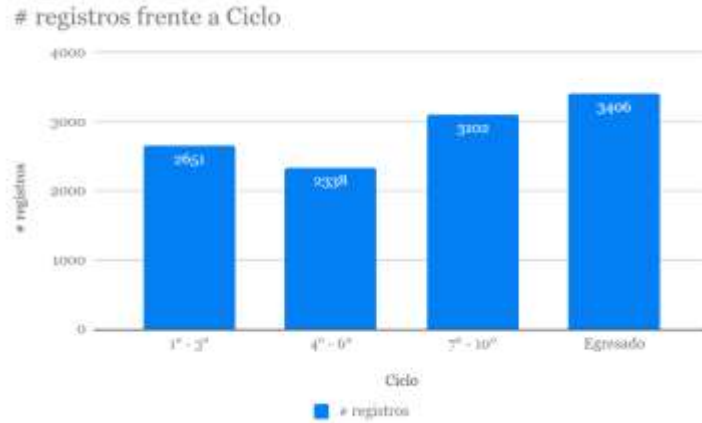


Fig. 6. Gráfico de columnas de la distribución de ciclo de estudios

Se observa que en este caso los datos están más distribuidos en partes iguales, pero son categorías y no identificadores.

- Atributo carrera

TABLA XV: DISTRIBUCIÓN DE FRECUENCIAS PARA EL ATRIBUTO CARRERA PROFESIONAL EN EL CONJUNTO INICIAL DE LEADS

Carrera	# registros
Administración	1745
Otros	1579
Ciencias de la Comunicación	995
Negocios Internacionales	714
Ingeniería Industrial	652
Marketing	531
Arquitectura y Urbanismo	480
Ingeniería Informática	405
Derecho	390
Contabilidad	363
Ingeniería Civil	349
Economía	332
Ciencias e Ingeniería	331
Artes Escénicas	299
Hotelería y turismo	254
Psicología	183
Ciencias Sociales	172
Idiomas	170
Ingeniería de Sistemas	155
Gestión y Alta Dirección	151
Publicidad y afines	124
Educación	121

Medicina Humana	117
Ingeniería Ambiental	109
Antropología	108
Ciencias de la Salud	104
Medicina Veterinaria	93
Diseño Gráfico	75
Letras y Ciencias Humanas	69
Arte y Diseño	55
Arqueología	41
Enfermería	34
Biología	30
Trabajo Social	30
Ingeniería Mecánica	29
Religión	22
Educación	21
Tecnología de información	18
Ingeniería Química	17
Arte y Diseño	15
Literatura	6
Finanzas	5
Ingeniería de Computación	2
Música	2
<b>Total</b>	<b>11497</b>

# registros frente a Carrera

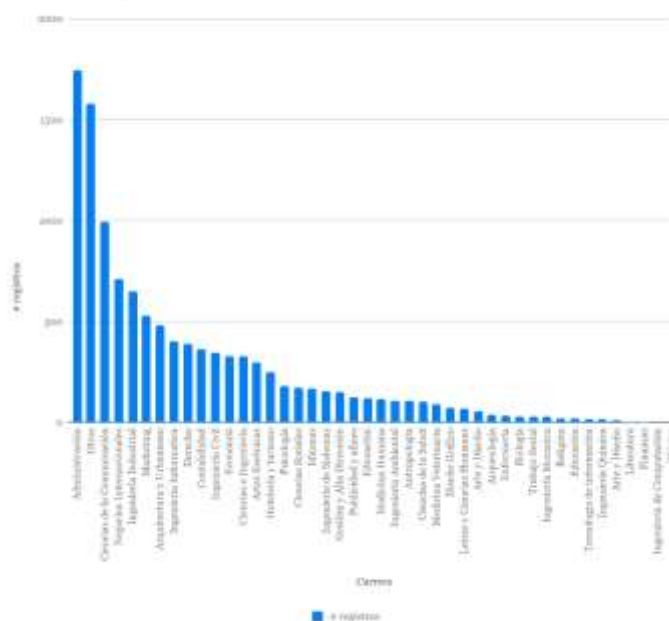


Fig. 7. Gráfico de columnas de la distribución de carreras

Se observa que las carreras con mayor número de registros son administración y ciencias de la comunicación. Asimismo, que hay carreras que representan la misma disciplina, pero con distintos nombres, y hay una gran cantidad de registros en la categoría “Otros”.

- Atributo programa

TABLA XVI: DISTRIBUCIÓN DE FRECUENCIAS PARA EL ATRIBUTO PROGRAMA DE INTERÉS EN EL CONJUNTO INICIAL DE LEADS

Programa	# registros
Emprendedor Global	1632
Talento Global	2766
Voluntario Global	7099
<b>Total</b>	<b>11497</b>



Fig. 8. Gráfico circular de la distribución de programas de interés

Se observa que la mayor cantidad de registros está distribuida en el programa Voluntario Global, seguido de Talento Global y con alrededor de mil datos menos, Emprendedor Global.

**4.1.2.4. Verificación de la calidad de los datos**

Existe un gran número de datos con valores vacíos o datos repetidos. Asimismo, se observa que hay campos relevantes como el sexo o la capacidad adquisitiva de inversión que no son solicitados ni almacenados en las bases de datos. No obstante, se cuenta con valores que

pueden cruzarse entre los dos grupos de datos para garantizar y corroborar la veracidad de algunos inconsistentes entre las dos fuentes, asegurando así la corrección y completitud de los datos. En consecuencia, se garantizó la consistencia de los valores individuales de los campos, distribución y cantidad de valores y disminución de ruido del proceso para eliminar registros duplicados. Por lo tanto, se cuenta con los datos suficientes para proceder a la siguiente fase.

### **4.1.3. Iteración #3: Preparación de los datos (CRISP-DM)**

#### **4.1.3.1. Selección de datos**

Si bien ya se había hecho un primer filtro de datos (en base a lo estipulado para esta investigación), se procedió a una nueva selección de datos, eligiendo aquellos más completos y con campos correctos. Asimismo, tras cruzar las bases de datos se eligieron aquellos no duplicados y se seleccionaron los campos más completos y que coincidían con la media de datos. Tras esto, la tabla de “Persona” quedó con los siguientes campos:

TABLA XVII: SELECCIÓN DE DATOS

<b>Campo</b>	<b>Tipo de dato</b>	<b>Motivo</b>
<b>NOMBRE</b>	Cadena de texto	Para identificar al <i>lead</i> y hacerle seguimiento.
<b>APELLIDO</b>	Cadena de texto	
<b>EMAIL</b>	Cadena de texto	Para funcionalidades anexas del sistema propuesto
<b>TELÉFONO</b>	Cadena de texto	
<b>EDAD</b>	Número entero	Para determinar rangos de edades de los <i>leads</i> y agruparlos.
<b>CAMPO DE ESTUDIO</b>	Cadena de texto	Para determinar distribución de carreras universitarias de los <i>leads</i> y agruparlos.
<b>CICLO EN CURSO</b>	Cadena de texto	Para determinar rangos y distribución del ciclo de los <i>leads</i> y agruparlos.
<b>FECHA DE VIAJE ESTIMADA</b>	Cadena de texto	Para determinar rangos de meses de los <i>leads</i> y agruparlos.
<b>DEPARTAMENTO</b>	Cadena de texto	Para identificar el origen del <i>lead</i> y futuras funcionalidades del sistema propuesto.
<b>STATUS</b>	Cadena de texto	Para determinar el agrupamiento o no en los reportes del sistema.
<b>FECHA DE APD</b>	Fecha	Para funcionalidades anexas al sistema propuesto.
<b>PAÍS DE DESTINO</b>	Número entero	Para determinar el agrupamiento del país de destino del intercambio y agruparlos.

#### 4.1.3.2. Limpieza de los datos

La normalización de datos, llenado de espacios en blanco, corrección de tipos y asignación de valores fue un

procedimiento que duró aproximadamente un mes en su totalidad debido a la gran cantidad de datos a limpiar. La herramienta utilizada fue Excel de Microsoft Office.

Fig. 9. Captura de datos incompletos en la tabla de leads interesados en el programa Voluntario Global

Fig. 10. Captura de datos incompletos en la tabla de leads interesados en el programa Talento Global

Para la normalización de datos se cruzaron todos los archivos exportados a través de los respectivos repositorios y utilizando funciones Excel para completar datos en blanco como BUSCARV, ALEATORIO.ENTRE, CONTAR.SI, CONCATENAR, MINUSC, tablas dinámicas, entre otras.

B	C	D	E
	Producto	Cliente?	Edad
uirre Lazarte	Talento Global	Sí	23

Fig. 11. Función BUSCARV para eliminación de duplicados tras la unificación de datasets

Se corroboró en repetidas veces la completitud e integridad de los datos previa la ejecución de algún modelo de minería de datos. Tras la limpieza total de datos, estos se redujeron de 12981 a 11973.

#### 4.1.3.3. Estructuración de los datos

Como se mencionó, existían datos no estructurados de la manera correcta, como el almacenamiento de la edad en lugar de la fecha de nacimiento o la ausencia del sexo del registrado, carreras de la misma disciplina, pero con múltiples nombres, entre otros. A partir de esta etapa se diseñó una base de datos relacional que soportará los modelos de minería y el almacenamiento de los datos en una plataforma online. El software gestor de base de datos elegido es MySQL. La tabla que recolecta todos los datos del registro se denomina *lead\_registro*. Para esto se generaron nuevos atributos, descritos a continuación:

TABLA XVIII: ATRIBUTOS GENERADOS PARA LA TABLA LEAD\_REGISTRO

Campo	Tipo de dato	Motivo
FECHANAC	date	Fecha de nacimiento.
CAT_INVERSION	int	Nombre de la categoría del monto inversión para el intercambio.
SEXO	char(1)	Sexo

#### 4.1.3.4. Integración de los datos

A partir de los datos seleccionados y los nuevos atributos de la tabla principal, se crean otras tablas que almacenarán las categorías o grupos que serán llamados a través de claves foráneas de identificadores únicos en la tabla principal. Estas tablas son:

- DESTINO\_REGION

Esta tabla contiene los nombres de las regiones que dividen las sedes de la asociación a nivel global en términos del negocio. Estos se dividen en cuatro: Americas, Asia-Pacific (AP), Europe y Middle-East and Africa (MEA)

TABLA XIX: DETALLE DE LA TABLA DESTINO\_REGION

<b>Campo</b>	<b>Tipo de dato</b>	<b>Descripción</b>
<b>IDDESTINO_REGION</b>	int(11)	Identificador único
<b>NOMBRE</b>	varchar(50)	Nombre de la región

- DESTINO\_PAIS

Esta tabla contiene los nombres en inglés y español de todos los países posibles de destino para el intercambio, así como identificadores únicos y clave foránea de la región a la que corresponden.

TABLA XX: DETALLE DE LA TABLA DESTINO\_PAIS

<b>Campo</b>	<b>Tipo de dato</b>	<b>Descripción</b>
<b>IDDESTINO_PAIS</b>	int(11)	Identificador único
<b>NOMBRE</b>	varchar(50)	Nombre del país en inglés
<b>NOMBRE_SPANISH</b>	int(11)	Nombre del país en español
<b>IDDESTINO_REGION</b>	varchar(50)	Clave foránea de la región

- CAT\_CARRERA

Esta tabla contiene los grupos de las carreras de educación superior, basándose en la elegibilidad para los programas especialistas de AIESEC. Estos se dividen en elegible para subproductos y carreras varias.

TABLA XXI: DETALLE DE LA TABLA CAT\_CARRERA

<b>Campo</b>	<b>Tipo de dato</b>	<b>Descripción</b>
<b>IDCAT_CARRERA</b>	int(11)	Identificador único
<b>DESCRIPCION</b>	varchar(50)	Nombre de la categoría

- CARRERA

Esta tabla contiene los nombres de las carreras de educación superior, así como identificadores únicos y clave foránea de la categoría a que pertenece.

TABLA XXII: DETALLE DE LA TABLA CARRERA

<b>Campo</b>	<b>Tipo de dato</b>	<b>Descripción</b>
<b>IDCARRERA</b>	int(11)	Identificador único
<b>NOMBRE</b>	varchar(50)	Nombre de la carrera
<b>IDCAT_CAR RERA</b>	int(11)	Clave foránea de la categoría de carrera

- PRODUCTO

Esta tabla contiene los nombres de los programas que ofrece la asociación, así como el precio de costo y el retorno neto por cada programa ofertado y los identificadores únicos. Estos se clasifican en Voluntario Global, Emprendedor Global y Talento Global.

TABLA XXIII: DETALLE DE LA TABLA PRODUCTO

<b>Campo</b>	<b>Tipo de dato</b>	<b>Descripción</b>
<b>IDPRODUCTO</b>	int(11)	Identificador único
<b>NOMBRE</b>	varchar(50)	Nombre del programa
<b>PRECIO</b>	decimal(10,2)	Monto de inversión para la adquisición del programa
<b>RETORNO</b>	decimal(10,2)	Monto de retorno neto por el monto aportado

- CAT\_INVERSION

Esta tabla contiene los nombres de los programas que ofrece la asociación, así como el precio de costo y el retorno neto por cada programa ofertado y los identificadores únicos. Estos se clasifican en Inversión básica, promedio y avanzada.

TABLA XXIV: DETALLE DE LA TABLA CAT\_INVERSION

<b>Campo</b>	<b>Tipo de dato</b>	<b>Descripción</b>
<b>IDPRODUCTO</b>	int(11)	Identificador único
<b>NOMBRE</b>	varchar(50)	Nombre del programa
<b>PRECIO</b>	decimal(10,2)	Monto de inversión para la adquisición del programa
<b>RETORNO</b>	decimal(10,2)	Monto de retorno neto por el monto aportado

- CAT\_CICLO

Esta tabla contiene los nombres de las categorías de los ciclos sobre los cuáles se evalúa la elegibilidad para un programa de intercambio u otro, así como su identificador único. Esta variación respecto al crudo de datos se debe a la intención de almacenar el ciclo para análisis más profundos pero aún así clasificar según este dato, como se hacía previamente en la asociación. Esta tabla servirá para hacer la comparación del primer factor: perfiles. Estos se dividen en ciclos básicos, avanzados y egresados.

TABLA XXV: DETALLE DE LA TABLA CAT\_CICLO

<b>Campo</b>	<b>Tipo de dato</b>	<b>Descripción</b>
<b>IDCAT_CICLO</b>	int(11)	Identificador único
<b>CICLOMIN</b>	int(11)	Ciclo mínimo de la categoría
<b>CICLOMAX</b>	int(11)	Ciclo máximo de la categoría
<b>DESCRIPCION</b>	varchar(50)	Descripción de la categoría

- DEPARTAMENTO

Esta tabla contiene los nombres de los departamentos del país desde el cual el *lead* hace su registro, así como los identificadores únicos.

TABLA XXVI: DETALLE DE LA TABLA DEPARTAMENTO

<b>Campo</b>	<b>Tipo de dato</b>	<b>Descripción</b>
<b>IDDEPARTAMENTO</b>	int(11)	Identificador único
<b>NOMBRE</b>	varchar(50)	Nombre del departamento

- **PERFIL\_DESEADO**

Esta tabla contiene los nombres de los perfiles de consumidor basado en las reglas de negocio. Esta clasificación se realizó tras un análisis de la asociación e intercambio de información respecto a su interés de hacer una doble tipificación del consumidor. Esta tabla sirve como referencia el primer algoritmo de clasificación de datos. Contiene las reglas de agrupamiento, así como su identificador único y descripción.

TABLA XXVII: DETALLE DE LA TABLA PERFIL\_DESEADO

<b>Campo</b>	<b>Tipo de dato</b>	<b>Descripción</b>
<b>IDPERFILESEADO</b>	int(11)	Identificador único
<b>IDCAT_INVERSION</b>	int(11)	Clave foránea de la categoría de inversión
<b>IDCAT_CARRERA</b>	int(11)	Clave foránea de la categoría de carrera
<b>CICLO</b>	varchar(10)	Lista de categorías de ciclos elegibles
<b>CAT_REGION_DESTINO</b>	varchar(10)	Lista de regiones elegibles
<b>CAT_MES_VIAJE</b>	varchar(10)	Lista de categoría de mes elegibles
<b>SEXO</b>	char(1)	Sexo
<b>DESCRIPCION</b>	varchar(250)	Descripción de las características de agrupamiento y nombre del perfil

- PRODUCTO\_PERFIL

Esta tabla contiene los nombres de los programas elegibles para determinado perfil en base al algoritmo propio de agrupamiento. También contiene un identificador único y claves foráneas del programa y perfil. Funciona como tabla accesoria intermedia para el módulo de conversión.

TABLA XXVIII: DETALLE DE LA TABLA PRODUCTO\_PERFIL

<b>Campo</b>	<b>Tipo de dato</b>	<b>Descripción</b>
<b>IDPRODUCTO_PERFIL</b>	int(11)	Identificador único
<b>IDPRODUCTO</b>	int(11)	Clave foránea del producto
<b>IDPERFILESEADO</b>	int(11)	Clave foránea del perfil

#### 4.1.3.5. Formateo de los datos

Para poder efectuar el modelo de agrupamiento se realizaron dos procesos a la tabla principal *lead\_registro*. El primer formateo de datos consistió en el cambio de campos de tipo cadena de texto (varchar) a identificadores únicos que funcionaron como claves foráneas, debido que ya se habían creado las nuevas tablas en la etapa anterior. Asimismo, se volvió a añadir los campos para almacenamiento de los datos de la confirmación del intercambio (denominado venta en términos generales de otros contextos), tales como la fecha, destino y mes de realización. Finalmente, se agregó un campo denominado clúster, para almacenar el valor resultante del modelo de agrupamiento a ejecutarse en la próxima fase de modelado. La tabla resultante es:

TABLA XXIX: FORMATEO DE LA TABLA LEAD\_REGISTRO

<b>Campo</b>	<b>Tipo de dato</b>	<b>Descripción</b>
<b>IDLEAD_REGISTRO</b>	int(11)	Identificador único del lead
<b>FECHA_REGISTRO</b>	date	Fecha de registro
<b>NOMBRE</b>	varchar(150)	Nombre del <i>lead</i>
<b>EMAIL</b>	varchar(150)	Correo electrónico de contacto
<b>TELÉFONO</b>	char(9)	Número de celular de contacto
<b>FECHANAC</b>	date	Fecha de nacimiento
<b>IDCAT_INVERSION</b>	int(11)	Clave foránea de la categoría de inversión
<b>IDCARRERA</b>	int(11)	Clave foránea de la carrera
<b>CICLO</b>	int(11)	Ciclo de estudios (considerando el egresado)
<b>IDDESTINODESEADO</b>	int(11)	Clave foránea del país de destino deseado para el intercambio
<b>MESDESEADO</b>	int(11)	Identificador del mes deseado
<b>SEXO</b>	char(1)	Sexo
<b>IDDEPARTAMENTO</b>	int(11)	Clave foránea del departamento de origen del registro
<b>IDPERFILDESEADO</b>	int(11)	Clave foránea del perfil resultante de la clasificación
<b>CLUSTER_PERFIL</b>	int(11)	Número del clúster resultante del algoritmo de agrupamiento
<b>STATUS_EP</b>	char(1)	Identificador del status del registro según el customer funnel
<b>FECHA_APD</b>	date	Fecha de confirmación del intercambio
<b>IDPRODUCTO_APD</b>	int(11)	Clave foránea del programa elegido para realizar el intercambio
<b>IDDESTINO_APD</b>	int(11)	Clave foránea del país de destino elegido para realizar el intercambio

El segundo formateo de datos consistió en la elaboración del primer algoritmo para agrupar los registros en base a la realidad del negocio sin ningún modelo de minería de datos, creando los perfiles de consumidor. Sobre esta primera segmentación se efectuarán los algoritmos de data mining a través de Rapid Miner, tal y como se justificó en la fase previa de esta investigación. El algoritmo de agrupamiento fue aplicado a la base de datos a través de un trigger al momento de la inserción de datos, asignando un valor al campo de perfil según las características del *lead* registrado.

#### **4.1.4. Iteración #4: Modelado (CRISP-DM)**

##### **4.1.4.1. Selección de la técnica de modelado**

Debido a la naturaleza del problema y objetivo de la minería descritos en la primera etapa, el problema de segmentación conlleva a elegir diferentes algoritmos entre k-means, x-means, DBSCAN, aleatorio, máquinas de vector soporte, entre otros. No obstante, debido a que los campos a analizar en esta fase son de tipo cualitativo (al tratarse de categorías), se opta por una variación del algoritmo k-means, puesto que está soportado por la herramienta elegida. Cabe resaltar que debido a la naturaleza cualitativa de los campos se deberá realizar una normalización de data y una modificación interna al criterio de agrupamiento del modelo.

##### **4.1.4.2. Generación del plan de prueba**

Al identificar 20 perfiles de consumidor, se separan los datos en 20 bloques diferentes ya clasificados por el valor de este campo y así separar los datos de prueba y de entrenamiento, minimizando la razón de error como medida de calidad de este modelo de minería de datos. La siguiente etapa de construcción de modelo se basará en el grupo de entrenamiento y se medirá la calidad de estos modelos con el conjunto de prueba. Para esto, se exportó

mediante herramientas de MySQL cada dataset con estos valores en formato .CSV que posteriormente fue convertido a .XLSX para ser leído por la herramienta.

#### 4.1.4.3. Construcción del modelo

Por cada perfil se ejecuta el algoritmo k-means al cual se le hicieron las modificaciones previamente mencionadas respecto a la normalización de datos y tipo de medidas del algoritmo de clusterización, de modo que se cambie de valores cuantitativos a cualitativos, considerando Bregman Divergences y Nominal Measures. Para la elaboración de este modelo se tomó como referencia base el auto-modelo de machine learning de Rapidminer respecto a clusterización de datos, normalizando los datos y agregando otras vistas y funciones de exportación para tener un modelo propio que se ajuste a la realidad de los datos. Este modelo fue el que se determinó para cada uno de los 20 perfiles.

Los campos que se consideraron para la minería fueron tres:

- Categoría de inversión
- País de destino
- Carrera

No obstante, debido a que esta subclasificación se hizo en base a los perfiles (los cuales ya tienen una clasificación previa y el primer criterio de agrupamiento es la categoría de inversión), los atributos distintos fueron solo los dos últimos. Para la elección del número de clúster a agrupar en base al algoritmo, se tomó la decisión de hacerlo mediante un análisis combinatorio del número de campos a evaluar. Este cálculo fue:

*Cálculo de espacio muestral = (#campos)!*

*Cálculo de espacio muestral = # clústeres*

*# clústeres = 3!*

*# clústeres = 6*

Habiendo obtenido este valor, se procedió a hacer un segundo procedimiento para validar la exactitud del número óptimo de clústeres, corroborando el índice de Davies–Bouldin, siendo 6 el que daba el valor más preciso para el número de clústeres respecto a las métricas evaluadas. Tras esto, se ejecutaron los algoritmos de agrupamiento.

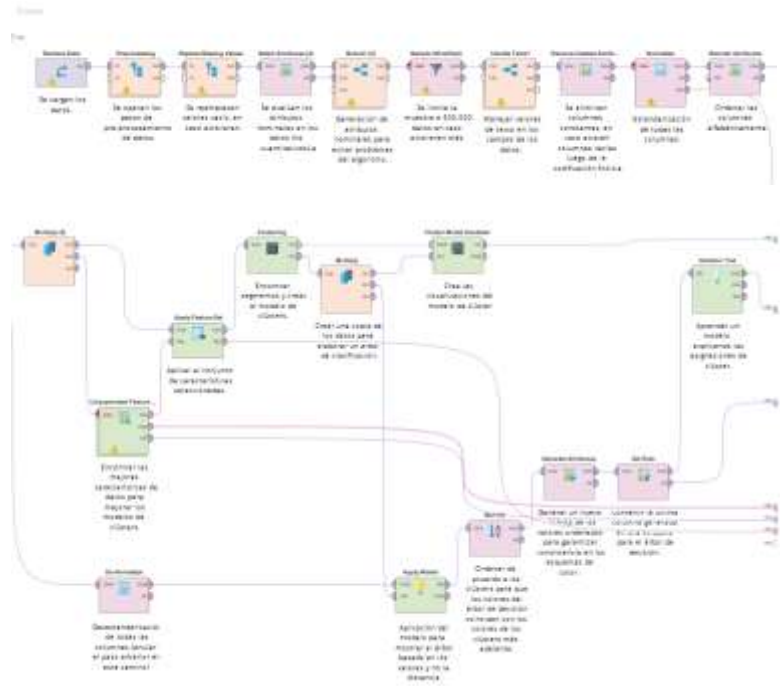


Fig. 12. Modelo de proceso en Rapidminer

#### 4.1.4.4. Evaluación del modelo

Se procedió a hacer una validación de este modelo mediante el juicio de expertos en minería de datos, quienes hicieron observaciones y posteriormente fueron levantadas en la versión final mostrada en este documento. Se contó con el soporte directo del servicio a la comunidad de Rapidminer.

Además, la elección de los tres campos a evaluar mediante un segundo proceso de segmentación de datos fue validado por los directivos nacionales de AIESEC en Perú, esperando conocer así más profundamente a sus consumidores.



#### 4.1.5.4. Diagramas de estados

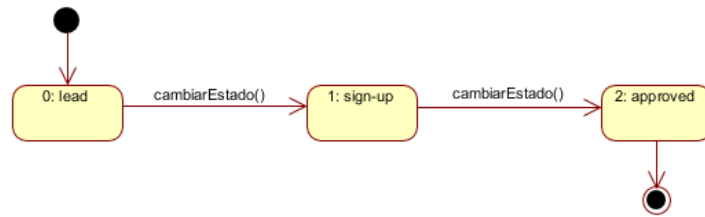


Fig. 16. Diagrama de estados de la clase *lead\_registro*

#### 4.1.5.5. Diseño de interfaces

Para el diseño de las interfaces del sistema se consideró principios de usabilidad heurística como un diseño minimalista y estético, presentando elementos claros y concisos; asimismo se consideró la prevención de errores a través de mensajes informativos y control intuitivo de fallos; asimismo se tomó en cuenta el diseño UX/UI para desarrollar interfaces responsivas, atractivas y con colores apropiados correspondientes a la organización. A continuación, se explica cada una de las funcionalidades del sistema y sus pantallas respectivas.

- Registro de *lead*, dentro del módulo cliente se diseñó una interfaz que recolecte todos los datos necesarios para la ejecución de los algoritmos de clasificación de perfiles y segmentación de *leads*.



Fig. 17. Registro de *leads* en el asistente (módulo cliente)

- Inicio de sesión para ingresar al sistema y los módulos de conversión y reportes. Para acceder son requeridos los campos de correo electrónico y clave.



Fig. 18. Inicio de sesión para acceso al sistema

- Menú principal, conteniendo en un solo tablero la información más relevante del mes respecto al progreso de metas operacionales, últimos registros, perfil destacado y distribución de *leads* y consumidores en base al producto elegido. Se diseñó de modo que los datos más relevantes estén a la vista del usuario y se eligieron los colores y logos corporativos para el tablero del menú principal.

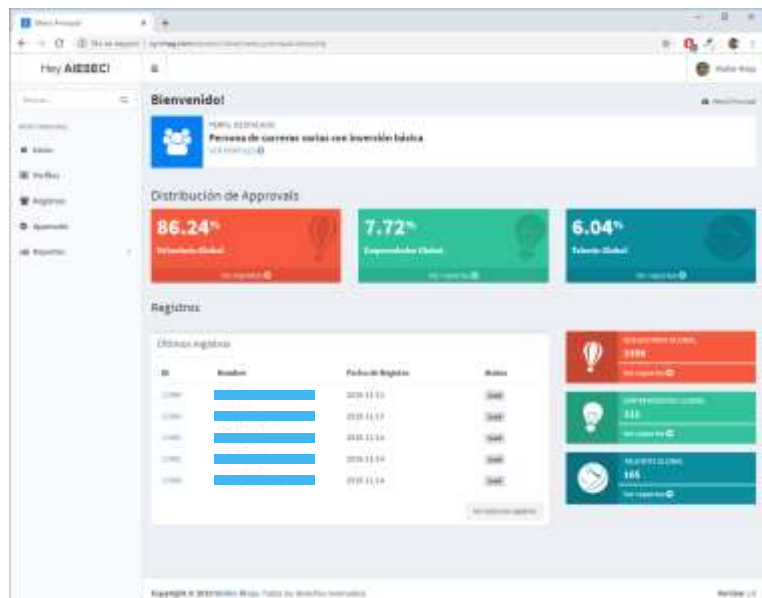


Fig. 19. Menú principal

- El listado de perfiles contiene los veinte (20) perfiles de consumidor resultantes de la clasificación explicada en la investigación. Esta interfaz puede ser accedida a través del menú principal o la barra de menús presente en el lado izquierdo. La presente pantalla redirecciona al detalle de los distintos perfiles existentes.

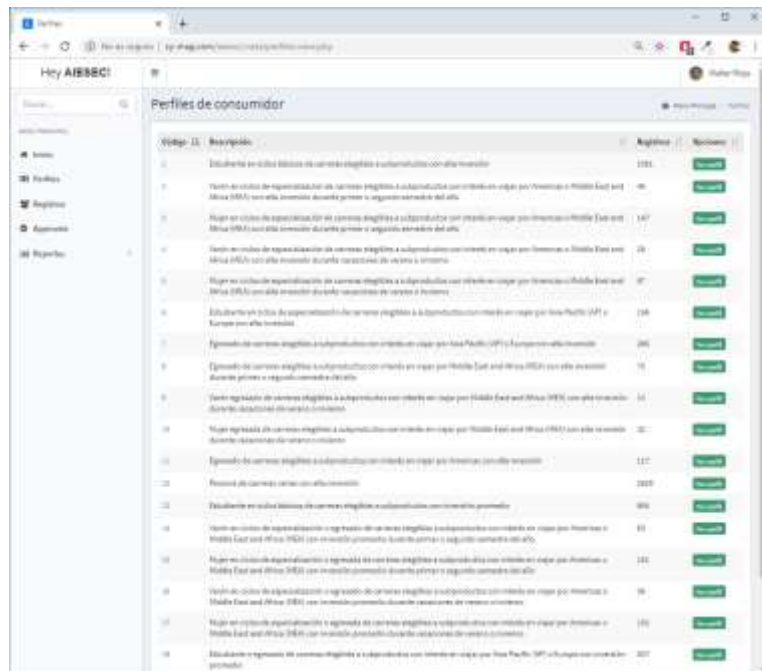


Fig. 20. Listado de perfiles

- Cada perfil cuenta con una interfaz de detalle donde se muestran reportes por carrera y departamento de origen, así como el listado completo de *leads* compatibles con dicho perfil y la lista de grupos (clústeres), cuyos detalles pueden ser accedidos a través de la misma pantalla (ver Fig. 21).
- Como se mencionó previamente, se cuenta también con una interfaz de detalle del clúster, donde se listan los registros compatibles con dicho segmento y un bloque adicional donde se encuentra la funcionalidad de envío de correos electrónicos personalizados, los cuales son enviados a todos los correos electrónicos de los *leads* pertenecientes a dicho clúster (ver Fig. 22).

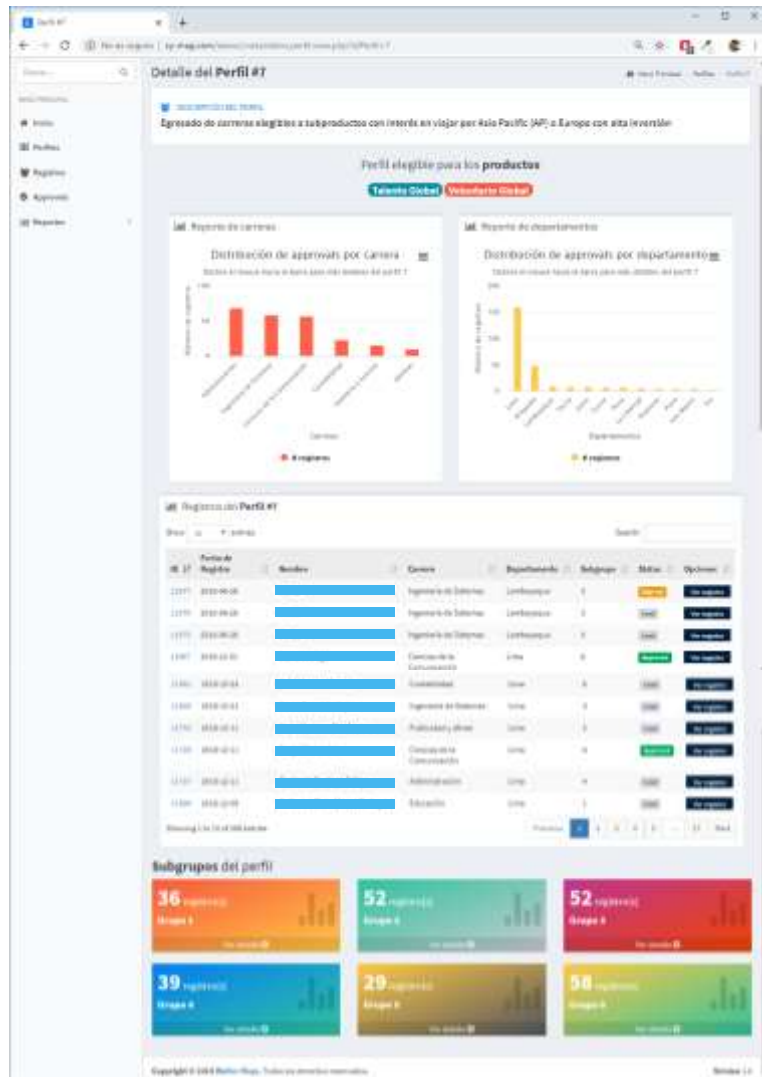


Fig. 21. Detalle de un perfil de consumidor

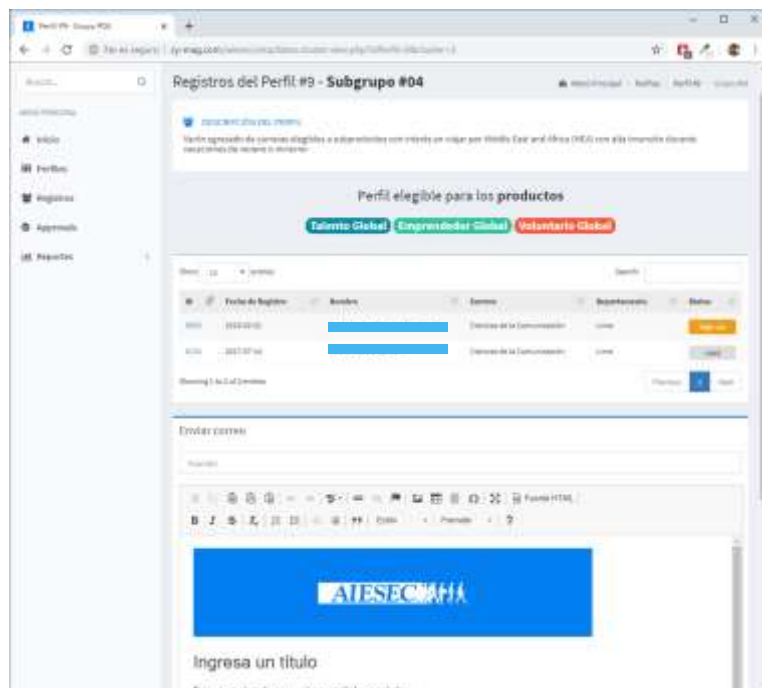


Fig. 22. Detalle de subgrupo (clúster) de perfil

- De modo similar a los perfiles de consumidor, se diseñó una interfaz con el listado de *leads* donde se muestran los datos más relevantes a primera vista, incluyendo el estado (*lead*, *sign-up* o *approved*), así como un botón de redireccionamiento al detalle del respectivo registro seleccionado.

Código	Fecha de Registro	Nombre	Servicio	Departamento	Perfil	Status	Opciones
1100	2020-11-25	[Barra azul]	Planetas	Managua	20	Lead	Ver perfil
1101	2020-11-25	[Barra azul]	El Escudo	San Marcos	10	Lead	Ver perfil
1102	2020-11-24	[Barra azul]	Agencia de Noticias	Managua	10	Lead	Ver perfil
1103	2020-11-24	[Barra azul]	Agencia de Noticias	Lombazona	1	Lead	Ver perfil
1104	2020-11-24	[Barra azul]	Asesoría	Lombazona	10	Lead	Ver perfil
1105	2020-11-24	[Barra azul]	Clases de la Compañía	Viña	10	Lead	Ver perfil
1106	2020-11-24	[Barra azul]	Asesoría	Managua	10	Lead	Ver perfil
1107	2020-11-24	[Barra azul]	Agencia de Noticias	Lombazona	1	Lead	Ver perfil
1108	2020-11-24	[Barra azul]	Agencia de Noticias	Lombazona	1	Lead	Ver perfil
1109	2020-11-24	[Barra azul]	Agencia de Noticias	Lombazona	1	Lead	Ver perfil
1110	2020-11-24	[Barra azul]	Sección	Lombazona	10	Lead	Ver perfil
1111	2020-11-24	[Barra azul]	Agencia de Noticias	Viña	1	Lead	Ver perfil
1112	2020-11-24	[Barra azul]	Agencia de Noticias	Viña	1	Lead	Ver perfil

Fig. 23. Listado de *leads*

- Para mostrar la información de cada *lead* se diseñó una interfaz que muestra todos los datos de registro, así como su estado según el *customer funnel* y detalles del perfil y subgrupo al que pertenecía. Dependiendo del estado en el que se encuentre, se lista información adicional correspondiente al proceso.

**Datos de Registro**

- Fecha de Registro: 2020-11-27
- Nombre: [Barra azul]
- Fecha de Actualización: [Barra azul]
- Sexo: [Barra azul]
- Departamento: [Barra azul]
- Categoría: [Barra azul]
- Perfil: [Barra azul]
- Monto a Iniciar: [Barra azul]

**Sobre el perfil**

- Perfil:** [Barra azul]
- Sub-grupo:** [Barra azul]
- Producto elegible:** [Barra azul]

Fig. 24. Detalle de *lead*

- Otra información requerida por la alta dirección corresponde a las tasas de conversión, por lo que se diseñó una interfaz de reportes clara, sencilla y gráfica donde se detallan los porcentajes de distribución por cifras totales y por perfil de consumidor. El reporte es además personalizable a través de un botón de división de fechas. Este reporte cuenta con semaforización según los indicadores óptimos planteados por la organización.

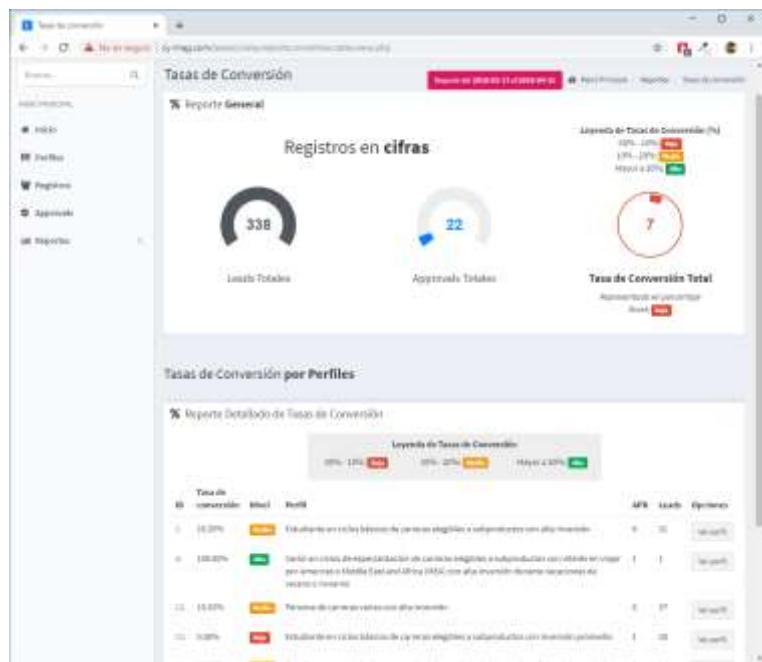


Fig. 25. Reporte de tasas de conversión

- En complemento al reporte mencionado previamente, se diseñó otro para el tiempo de procesamiento. Este está enfocado principalmente en el programa elegido por los consumidores, por lo que se diseñó este reporte dividiéndolo en dos bloques: tiempo de procesamiento por programa y otro por perfiles de consumidor. Asimismo, esta interfaz cuenta con una semaforización del indicador de tiempo según las metas planteadas por la organización, las cuales fueron brindadas por los directivos.

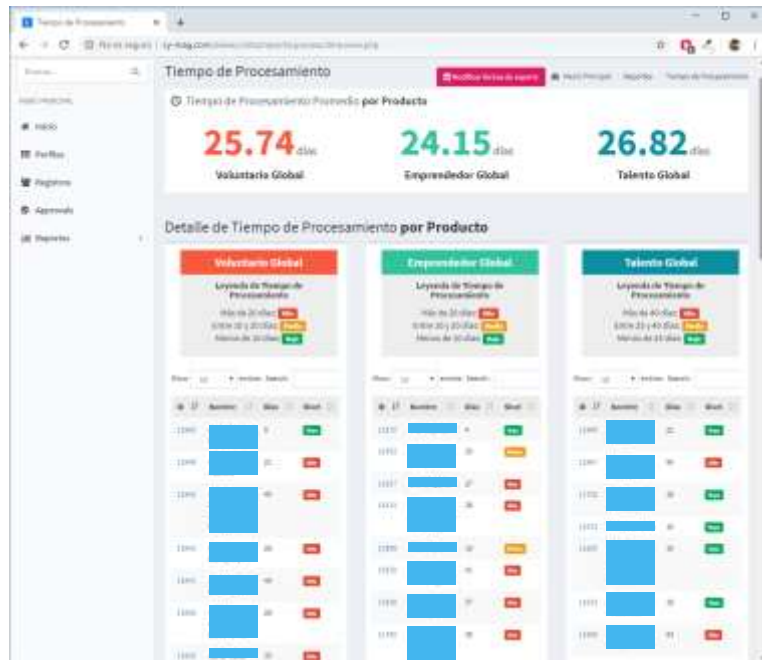


Fig. 26. Interfaz de registro en el asistente (módulo cliente)

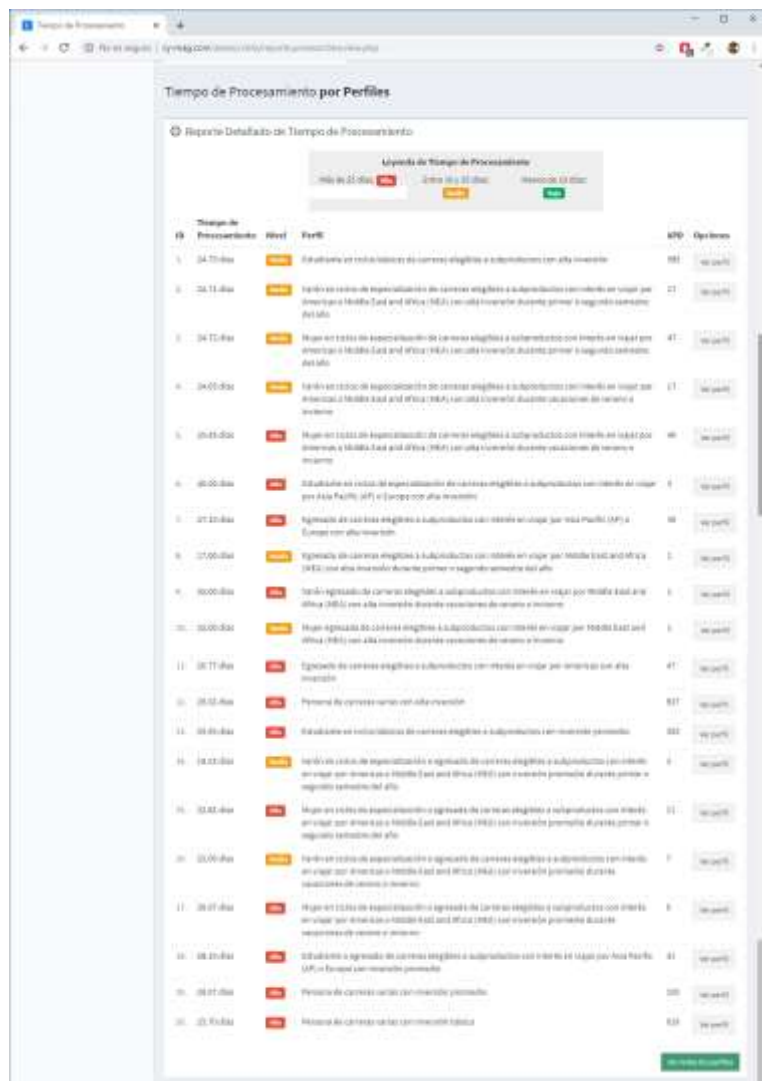


Fig. 27. Interfaz de registro en el asistente (módulo cliente)

- Finalmente, se diseñó una interfaz de reporte personalizado de aquellas personas que lograron convertir y llegaron a adquirir alguno de los programas ofrecidos por AIESEC. El término utilizado por la organización es *approval*, equivalente a consumidor final. Para este tablero se tomaron en cuenta los atributos definidos como críticos y relevantes para la investigación y el proceso de conversión tales como sexo, programa elegido, carrera y departamento/región de registro.

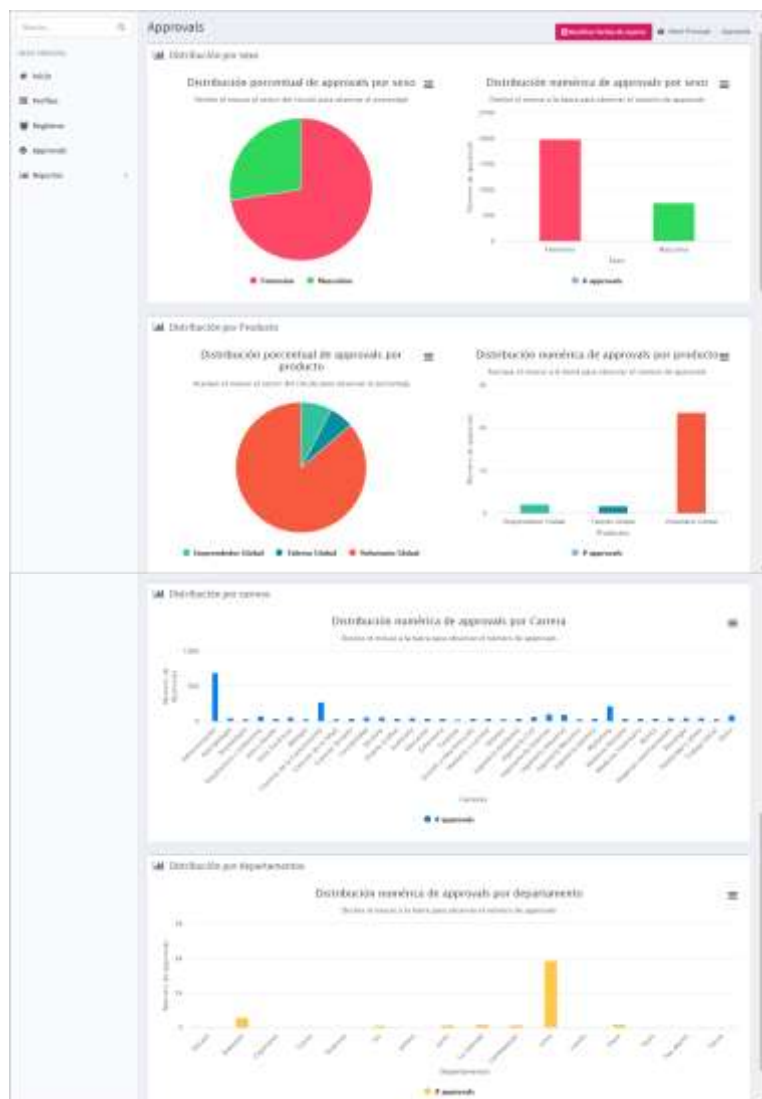


Fig. 28. Reporte de *approvals* (consumidores)

#### 4.1.5.6. Diseño de la arquitectura

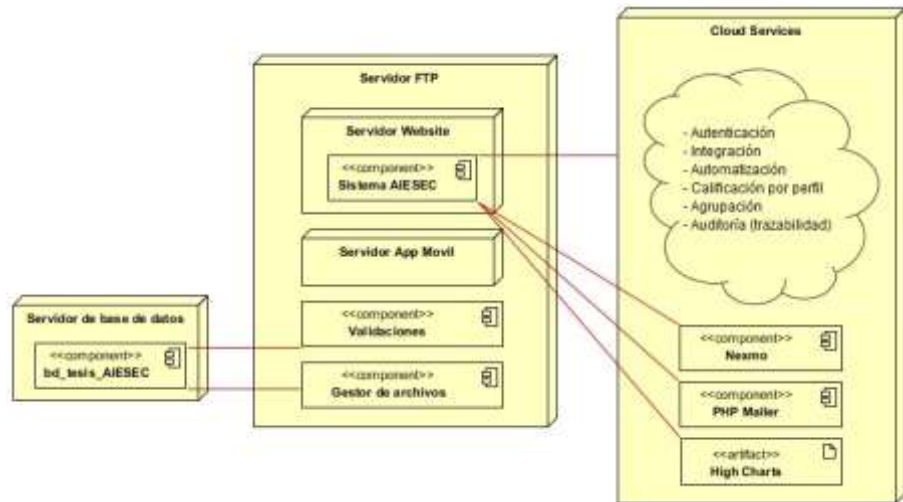


Fig. 29. Arquitectura de la aplicación

#### 4.1.6. Iteración #6: Evaluación (CRISP-DM)

##### 4.1.6.1. Evaluación de los resultados

Respecto a la relación de los objetivos del negocio con los de la minería, los resultantes de este procedimiento brindados por la herramienta permitieron visualizar los datos que posteriormente fueron estructurados a través de reportes en la aplicación, cumpliendo con los objetivos de la asociación mencionados en la primera etapa. Es aconsejable aplicar el modelo en tiempo real para realizar una nueva minería de datos en seis o doce meses próximos y analizar si los clústeres se mantienen con sus reglas o han cambiado; de modo que se tome una decisión de la implantación de los clústeres como reglas del negocio, tal y como sucedió con la segmentación por perfiles deseados.

##### 4.1.6.2. Proceso de revisión

El proceso completo de minería de datos aplicando los algoritmos de clusterización trajo resultados positivos aplicando el algoritmo de k-means. Otros algoritmos de aprendizaje no supervisado podrían ser probados (sus beneficios se detallan en el *Capítulo V: Discusión*), sin embargo, esto representaría una modificación en el modelo y la evaluación de otros atributos de la tabla *lead\_registro*,

distintos a los elegidos. Al no elegir aquellas características críticas para la decisión de adquirir o no un programa u otro, el resultado del modelo implementado sería la ponderación de otros atributos y su relación con el objetivo principal: convertir. Asimismo, se podría realizar un tercer procedimiento de clusterización, esta vez sobre la totalidad de datos para corroborar que las reglas del negocio rigen también sobre la naturaleza de los datos, buscando establecer una correlación directa y positiva entre la realidad de los registros de AIESEC en Perú y sus modelos internos tras estudios de mercado u otras técnicas de recolección de información.

#### 4.1.6.3. Determinación de futuras fases

Al haber obtenido resultados satisfactorios durante el proceso de minería de datos se decidió proceder a la etapa de implantación y prueba sin necesidad de correr una nueva iteración con parámetros distintos a los evaluados.

### 4.1.7. Iteración #7: Implementación y Prueba (CRISP-DM y RUP)

#### 4.1.7.1. Diagrama de componentes (RUP)

Para garantizar la buena capacidad, disponibilidad y usabilidad de cada uno de los aplicativos y tecnologías se consideraron como componentes del *backend* y *frontend* los que se muestran en la Fig. 30.

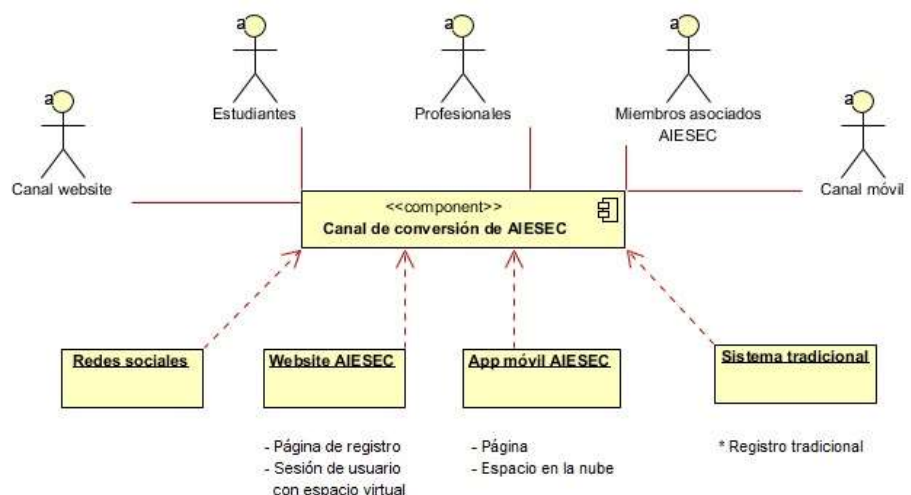


Fig. 30. Diagrama de componentes

#### **4.1.7.2. Plan de implantación (CRISP-DM)**

El modelo final elegido para este proceso de minería de datos se automatizó a través de la transformación de las reglas de agrupamiento de cada clúster en árboles de decisión y fueron integrados en el algoritmo inicial de segmentación en perfiles a través del *trigger* de inserción en la base de datos utilizando el lenguaje MySQL.

#### **4.1.7.3. Monitorización y mantenimiento (CRISP-DM)**

Tal y como se aconsejó en el ítem **4.1.6. Iteración #6: Evaluación, sección Evaluación de los resultados, en el presente Capítulo**, se recomienda realizar revisiones periódicas al modelo actual de minería de datos con el objetivo de garantizar el correcto funcionamiento de los clústeres generados. El tiempo propuesto para el monitoreo y mantenimiento es de doce (12) meses, debido a que se necesita una gran cantidad de datos para garantizar que la minería de datos concluye en modelos confiables para ser implementados en los sistemas de la asociación.

#### **4.1.7.4. Informe final (CRISP-DM)**

Los resultados logrados con el presente proyecto se detallan en el ítem **4.2. En base a los objetivos de la investigación, en el presente Capítulo IV**.

#### **4.1.7.5. Pruebas unitarias y de integración (RUP)**

Con el objetivo de garantizar el funcionamiento óptimo del sistema se consideraron cuatro (04) tipos de pruebas:

- Verificación de la calidad de información presentada a los directivos (detallada en el ítem **4.2.3**).
- Verificación de la calidad de información sugerida a los *leads* durante su proceso de conversión (detallada en el ítem **4.2.4**).
- Verificación de la calidad de la aplicación mediante juicio de expertos (detallada en el ítem **4.2.5**).
- Verificación de la calidad de la aplicación a través de pruebas automatizadas de portabilidad, usabilidad,

disponibilidad, velocidad de carga, eficiencia del código fuente, entre otros. Este ítem es el que se detalla en el presente inciso.

Se probaron distintas herramientas para las pruebas automatizadas. Obteniendo los resultados mostrados a continuación:

- A través del software online *PageSpeed Insights* se testeó la velocidad de diferentes interfaces tanto en plataformas web como móviles. En la mayoría de casos se obtuvo puntajes en rangos óptimos, generando resultados positivos; algunas capturas se detallan en las Fig. 31 y 32.

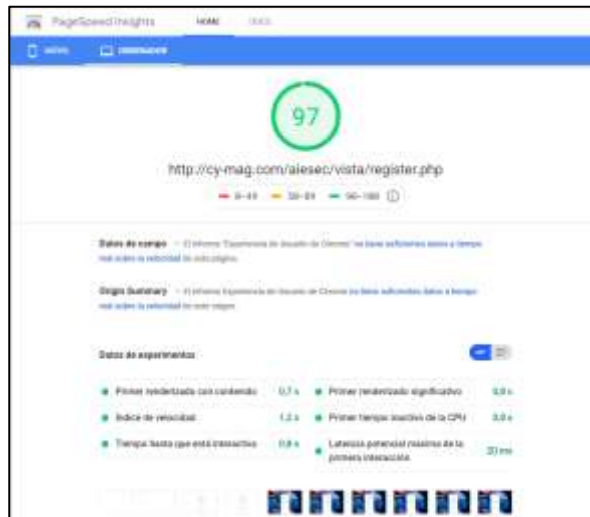


Fig. 31. Análisis de la velocidad de carga del formulario de registro de *leads*

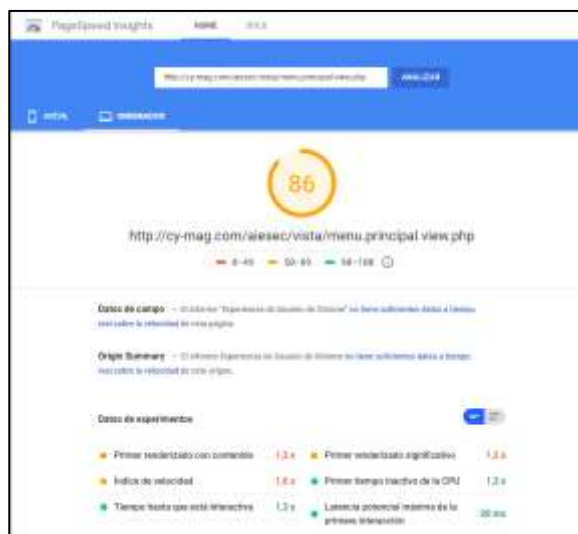


Fig. 32. Análisis de la velocidad de carga del menú principal del sistema

- A través del software online *Dr. Watson* se evaluaron la velocidad de descarga, la validez de enlaces, compatibilidad con el motor de búsqueda, entre otros factores. Tras testear el software se obtuvo resultados positivos, detallados en la Fig. 33.

Checking server response  
Server response

**Response code from http://cy-mag.com/aiesec/vista/register.php**  
OK (code 200)

Analysing HTML  
Comparing server-side assets  
Estimated download speed

These download times are estimates and should only be used as a general guideline. Many factors, such as your modem, quality of connection, ISP load, server responsiveness, and Internet routing can greatly impact the actual download times experienced by your visitors.

Object type	Number	Size in bytes	Estimated download times (seconds)					
			14.4	28.8	33.6	56k	128K	T1
HTML								
	1	17851	13.02	7.14	5.95	5.10	2.23	0.89
<b>Total</b>	<b>1</b>	<b>17851</b>	<b>13.02</b>	<b>7.14</b>	<b>5.95</b>	<b>5.10</b>	<b>2.23</b>	<b>0.89</b>

Analysing HTML errors  
Syntax and style analysis  
(clicking on the line # takes you to that line in the code)

Congratulations! I found no problems.

Computing word counts  
Word counts:

Words: 398  
Average word length: 6.6  
Unique words: 346

Checking spelling  
Spelling Check  
(clicking on the line # takes you to that line in the code)

No spelling problems

Fig. 33. Resultados del análisis del software Dr. Watson

- Se utilizó el servicio gratuito de la *World Wide Web Consortium (W3C)* para corroborar los estándares de accesibilidad de la aplicación y la validación de las hojas de estilo. Como se observa en la Fig. 34, no se encontraron errores en el código fuente.

**Nu Html Checker**

This tool is an ongoing experiment in better HTML checking, and its behavior remains subject to change.

Showing results for http://cy-mag.com/aiesec/vista/menu.principal.view.php

Checker Input

Show  source  outline  image report

Check by: address

http://cy-mag.com/aiesec/vista/menu.principal.view.php

**Document checking completed. No errors or warnings to show.**

Used the HTML parser. Externally specified character encoding was UTF-8.  
Total execution time: 442 milliseconds.

Fig. 34. Resultados de la prueba automatizada de caja blanca mediante W3C



## 4.2. En base a los objetivos de la investigación

### 4.2.1. Reducir el tiempo de respuesta a *leads* tras su registro en la plataforma

Para la asociación uno de los indicadores críticos de sus procesos es la demora en la respuesta a una persona desde su primer contacto con AIESEC (desde su registro mediante plataforma o inscripción offline). El sistema desarrollado requería disminuir dicho tiempo de respuesta tras el registro de un *lead*. El cumplimiento de este objetivo se muestra en la Tabla XXX.

TABLA XXX: COMPARACIÓN DE TIEMPOS DE RESPUESTA A LEADS TRAS SU REGISTRO

Tiempo de respuesta (preprueba)	Tiempo de respuesta (posprueba)
24 horas	< 10 segundos

El tiempo de respuesta se redujo al mínimo (se consideró menor a 10 segundos, siendo un valor que depende del servidor de correos) debido a la automatización de la respuesta tras el registro de un *lead*, puesto que se programó en el módulo de registro el envío inmediato de un correo electrónico al *lead*, ofreciéndole el programa más idóneo para él basado en sus datos de registro.



Fig. 36. Correo enviado automáticamente al *lead* con los programas a los que califica

Asimismo, se le envió un mensaje de texto al número telefónico de registro para apoyar aún más la respuesta positiva del *lead* e incrementar su probabilidad de adquirir un programa.



Fig. 37. Mensaje de texto enviado al *lead* tras su registro

Se demuestra que se cumplió el objetivo específico N° 01, así como también se redujo considerablemente el tiempo de contacto para aplicar otras técnicas de *lead nurturing* como el envío de correos personalizados, ofreciendo programas a la medida u otros fines que la asociación considere pertinente.

#### 4.2.2. Segmentar *leads* a través de técnicas de minería de datos

Previo a la aplicación de minería de datos se realizó una segmentación basándose en las reglas y políticas internas de AIESEC en Perú; los motivos de esto se detallan en el **Capítulo V: Discusión**.

El cumplimiento de este objetivo se detalla en el **ítem 4.1.4. Iteración #4: Modelado, sección Construcción del modelo, en el Capítulo IV. Resultados**. Tras la aplicación de minería de datos y encontrar las reglas que determinaban los clústeres de cada perfil aplicando la herramienta Rapidminer y validar el índice de Davies–Bouldin con el análisis combinatorio elegido, se exportaron los modelos mediante árboles de decisión. Este procedimiento se repitió por cada uno de los 20 perfiles de consumidor, siguiendo el ejemplo mostrado en la Fig. 38.

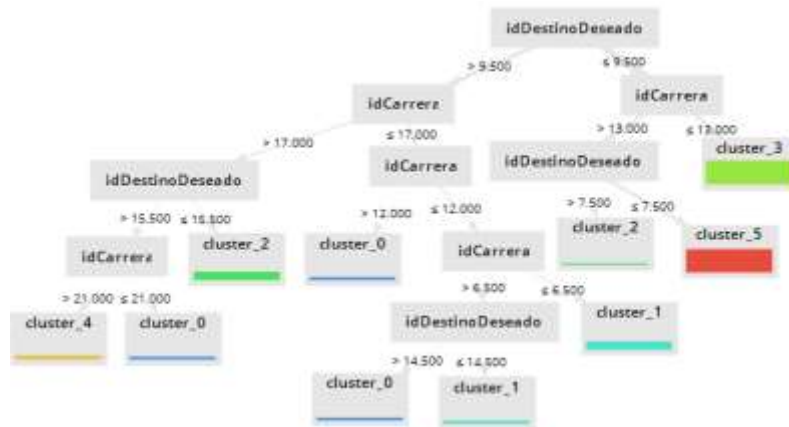


Fig. 38. Árbol binario para reglas de clusterización del Perfil #11 en Rapidminer.

Dichas reglas posteriormente se integraron al algoritmo inicial de clasificación de *leads* por perfiles tras su registro; en la Fig. 33 se muestran las reglas de clusterización integradas usando MySQL.

```

WHEN 11 THEN
CASE
WHEN NEW.idDestinoDeseado > 9.5 THEN
CASE
WHEN NEW.idCarrera > 17 THEN
CASE
WHEN NEW.idDestinoDeseado > 15.5 THEN
CASE
WHEN NEW.idCarrera > 21 THEN 4
WHEN NEW.idCarrera <= 21 THEN 0
END
END
WHEN NEW.idDestinoDeseado <= 15.5 THEN 1
END
END
WHEN NEW.idCarrera <= 17 THEN
CASE
WHEN NEW.idCarrera > 12 THEN 0
WHEN NEW.idCarrera <= 12 THEN
CASE
WHEN NEW.idCarrera > 6.5 THEN
CASE
WHEN NEW.idDestinoDeseado > 14.5 THEN 0
WHEN NEW.idDestinoDeseado <= 14.5 THEN 1
END
END
WHEN NEW.idCarrera <= 6.5 THEN 1
END
END
END
END
WHEN NEW.idDestinoDeseado <= 9.5 THEN
CASE
WHEN NEW.idCarrera > 13 THEN
CASE
WHEN NEW.idDestinoDeseado > 7.5 THEN 2
WHEN NEW.idDestinoDeseado <= 7.5 THEN 5
END
END
WHEN NEW.idCarrera <= 13 THEN 3
END
END
END

```

Fig. 39. Segmentación de clústeres del Perfil #11 en MySQL.

De este modo se demuestra que se logró segmentar los perfiles generados aplicando técnicas de clusterización, cumpliendo así con el objetivo específico N° 02.

### 4.2.3. Incrementar la calidad de información brindada a los directivos de la organización para la toma de decisiones estratégicas

El producto acreditable cuenta con distintos tableros y reportes con información crítica como tasas de conversión, tiempos de procesamiento y distribución de *leads* por ciudad de destino, carrera, sexo, programa de preferencia y otros. Se desarrolló así debido al objetivo de presentar información más precisa, rápida y confiable a los usuarios de los rangos más altos, apoyando así en mayor medida a la toma de decisiones estratégicas en AIESEC en Perú.

Se desarrollaron interfaces para mostrar el tiempo de procesamiento y tasa de conversión, ambas desde el registro hasta la adquisición de un programa de intercambio. Estas mostraban los indicadores semaforizados, dividiéndolos por programa: Voluntario Global, Emprendedor Global y Talento Global; así como por los perfiles de consumidor generados. En las Fig. 40 y 41 se presentan las capturas de dichas interfaces.



Fig. 40. Captura de pantalla de la interfaz de reporte de tiempo de procesamiento por programas.

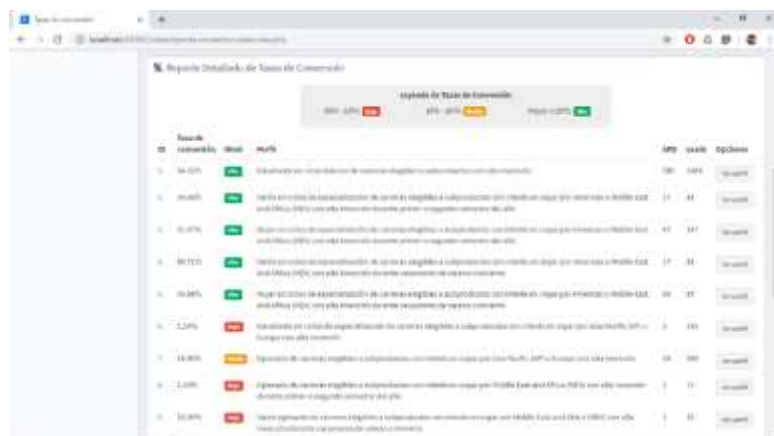


Fig. 41. Captura de pantalla de la interfaz de reporte de tasa de conversión por perfiles.

En la preprueba a los miembros asociados se preguntó también respecto a la confiabilidad de la información provista por el sistema, así como el grado en que esta brindaba soporte suficiente para la toma de decisiones estratégicas. Tras la demostración de la aplicación y sus funcionalidades se aplicó una posprueba con las mismas preguntas. La escala estimativa en ambos casos fue del 1 al 10, siendo este último el valor máximo. Los resultados se muestran en la Tabla XXXI.

TABLA XXXI: COMPARACIÓN DEL GRADO DE ACEPTACIÓN DE LA INFORMACIÓN POR LOS USUARIOS

Grado de aceptación (preprueba)	Grado de aceptación (posprueba)
6	8.5

De este modo se demuestra que la información mostrada en la aplicación desarrollada es mejor que la provista por los sistemas previamente utilizados, apoyando en mayor medida a la toma de decisiones por la alta dirección de AIESEC en Perú y cumpliendo con el objetivo específico N° 03.

#### **4.2.4. Incrementar la calidad de la información presentada a *leads* en el acompañamiento digital del proceso de conversión**

Tal y como se detalló en el cumplimiento del objetivo específico N° 1, el producto acreditable mejoró el tiempo de respuesta e incrementó los medios utilizados para el contacto; dichas funcionalidades fueron implementadas en la aplicación con el fin de mejorar la gestión de la relación con el consumidor para así incrementar su satisfacción respecto al acompañamiento dado durante su proceso y así tomar una decisión frente a la adquisición de algún programa.

Asimismo, en la interfaz donde se listaron los *leads* no convertidos de cada subgrupo (clúster) de perfil, se implementó la función para envío de correos electrónicos personalizados a todos los registros incluidos en dicho segmento. Estas interfaces, así como sus pruebas se muestran en las Fig. 42 y 43.



Se realizó una preprueba y posprueba a la muestra de estudiantes interesados en tomar un intercambio para determinar si todas las funcionalidades descritas mejoraban la calidad del contacto con AIESEC e incrementaban su satisfacción frente a dicho acompañamiento a través de medios digitales. La escala estimativa para este objetivo fue la misma que la anterior, siendo del 1 al 10 respecto al grado de aceptación de la información provista por la aplicación. La comparación del indicador correspondiente entre ambas pruebas se muestra en la tabla XXXII.

TABLA XXXII: COMPARACIÓN DEL GRADO DE ACEPTACIÓN DE LA INFORMACIÓN BRINDADA AL CONSUMIDOR

Grado de aceptación (preprueba)	Grado de aceptación (posprueba)
3	8.1

De este modo se demuestra que el objetivo específico N° 04 fue cumplido, puesto que la prueba realizada en estudiantes reflejó un claro incremento en el soporte digital y aumentó el nivel de satisfacción de los consumidores, siendo promotores y recomendándolo.

#### 4.2.5. Verificar la calidad de la aplicación

La calidad de la aplicación desarrollada fue evaluada a través de cuatro (04) indicadores: usabilidad, disponibilidad, portabilidad y testeabilidad. Estos tres (03) últimos fueron revisados y validados a través de herramientas online para realizar las pruebas de software, detallados en el *ítem 4.1.7. Iteración #7: Implantación y prueba, sección Pruebas unitarias y de integración*; mientras que para el criterio de usabilidad fue verificado mediante el juicio de expertos. Para esto, se aplicó una lista de cotejos (ver *Anexo N° 03*) basada en la norma ISO 9241-11: 2018 [64], evaluando la usabilidad mediante los factores de facultad d aprendizaje, recuerdo en el tiempo, comprensibilidad, tolerancia a errores, atraktividad y satisfacción. Se eligió a 05 expertos por su experiencia en revisión de calidad de software, testeo, seguridad y comprensión de *dashboards*. Sus respuestas pueden verse en la Fig. 44.

Posterior a la recolección del juicio de expertos, se procedió a evaluar la confiabilidad de dicha evaluación, por lo que se calculó el coeficiente alfa de Cronbach, instrumento eficaz para dicho cometido [62]. Su cálculo se muestra en la Fig. 45.

EXPERTOS	ÍTEMS																								Total	
	Facilidad de Aprendizaje					Recuerdo en el tiempo		Comprensibilidad					Tolerancia a errores			Atractividad				Satisfacción						
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24		
1	4	3	4	3	4	4	4	3	4	4	4	4	3	3	4	4	3	4	4	4	4	4	4	4	4	90
2	4	4	1	3	3	4	4	3	3	4	4	3	2	3	3	4	4	4	4	4	4	3	3	3	3	80
3	4	4	3	4	4	3	4	4	4	4	4	4	4	3	3	3	4	4	4	4	4	4	4	4	4	91
4	4	3	4	4	4	4	4	4	4	4	4	4	4	3	4	4	4	4	3	4	4	4	4	4	4	93
5	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	3	4	4	4	4	4	4	95
ESTADÍSTICOS																										
Varianza	0	0.3	1.7	0.3	0.2	0.2	0	0.3	0.2	0	0	0.2	0.7	0.2	0.3	0	0.2	0.3	0	0	0.2	0.2	0.2	0.2	0.2	

Fig. 44. Matriz de datos del juicio de expertos.

<b>K</b>	24
$\sum V_i$	5.9
<b>Vt</b>	33.7
<b>Sección 1</b>	1.04
<b>Sección 2</b>	0.82
<b> S2 </b>	0.82
<b><math>\alpha</math></b>	0.86

Fig. 45. Cálculo estadístico del coeficiente alfa de Cronbach.

De acuerdo con [62], el valor óptimo del alfa de Cronbach es determinado por su cercanía a la unidad. Tras realizar los cálculos estadísticos mostrados previamente, se pudo determinar como efectivo el instrumento aplicado al obtener  $\alpha=0.86$ , siendo este un valor adecuado-satisfactorio del alfa de Cronbach. Por lo tanto, se consideró el juicio de expertos aplicado como válido.

Expertos	Promedio
1	3.75
2	3.33
3	3.79
4	3.88
5	3.96
<b>Promedio</b>	<b>3.74</b>
<b>Máximo posible</b>	<b>4</b>
<b>Resultado</b>	<b>3.74</b>
<b>Porcentaje</b>	<b>94%</b>

Fig. 46. Promedio de calificaciones de expertos.

En la Fig. 46 se observa el promedio de las calificaciones brindadas por los expertos, concluyendo en un 3.74 como valor resultante. Este representa un 94% del máximo posible según la escala mostrada en la lista de cotejos. Dicho esto, se considera como cumplido el objetivo específico N° 05, puesto que la aplicación desarrollada fue calificada como válida.

#### **4.3. Impactos esperados**

##### **4.3.1. Impactos económicos**

La aplicación desarrollada no solo representa una herramienta de soporte para el proceso de conversión de *leads* sino además apoya la gestión táctica y estratégica de AIESEC respecto a sus áreas operativas; al facilitar el acceso a información crítica a la alta gerencia se apoya el proceso de toma de decisiones en la asociación. Asimismo, la aplicación permite una mejor gestión de la relación que se tiene con los consumidores antiguos, actuales y futuros. Todo esto en el mediano y largo permitirá incrementar la rentabilidad de AIESEC en Perú, impactando así positivamente en la economía de la asociación.

##### **4.3.2. Impactos sociales**

Como se describió previamente a la asociación, AIESEC en Perú trabaja en torno a los 17 ODS y la agenda 2030 que incluye a la erradicación de la pobreza, lucha contra el cambio climático y reducción de las desigualdades. La implementación de la aplicación propuesta en la presente investigación permite brindar, a un público juvenil, opciones de programas de intercambio que aportan un mayor conocimiento de otras culturas, desarrollo de habilidades interpersonales y otras aptitudes de liderazgo necesarias en el mundo de hoy en día.

##### **4.3.3. Impactos en tecnología**

El aporte de la investigación se fundamenta en la innovación tecnológica de bajo costo para la organización, pues el algoritmo principal de la aplicación fue desarrollado a la medida. Se espera que la implementación de esta aplicación y su futura integración con los

sistemas utilizados por la asociación representen un antecedente para próximas investigaciones en el uso de minería de datos aprovechando herramientas de bajo costo en empresas sociales, ONGs y MYPEs que manejen grandes volúmenes de datos.

#### **4.3.4. Impactos en la formación de cadenas productivas**

Para AIESEC en Perú, se espera que la implementación futura de la aplicación web desarrollada permita una transformación en los procesos internos de la organización, dando pie a un incremento en la productividad de los miembros y la innovación en subprocesos de la conversión de *leads*, reinventándose y planteando evoluciones que incrementen la eficacia y eficiencia del proceso completo y la integración de otras herramientas tecnológicas que mejoren el ecosistema digital de la asociación en distintas partes del país.

## V. DISCUSIÓN

El presente capítulo presenta el análisis de la hipótesis y objetivos planteados en esta investigación y su relación con los resultados de la aplicación desarrollada para AIESEC en Perú.

El producto acreditable buscó apoyar el proceso de conversión de *leads* en la asociación desde dos perspectivas: por un lado, automatizar, agilizar y personalizar los puntos de contacto a través de medios digitales; y por el otro la mejora en la calidad de información presentada para la alta dirección y apoyar la toma de decisiones estratégicas. Esto debido a la necesidad de incrementar la información y aprovechar los datos históricos de AIESEC para determinar las necesidades de los consumidores.

[11] organizó paquetes promocionales con el objetivo de incrementar las ventas en la empresa sobre la cual se realizó su investigación; sin embargo, se optó por no llegar a este nivel puesto que se consideró que en lugar de segmentar el consumo de productos era mejor agrupar a los clientes por perfiles de consumidor. Se tomó esta decisión debido a las políticas y procesos internos de la asociación, donde ya existían reglas previas para segmentar el público objetivo de cada programa de intercambio. Estas características por programa fueron provistas por la alta dirección cuando se le fue solicitado y se muestran en la Fig. 47.



Mes	Abril - Noviembre	Setiembre - Noviembre	Mayo - Junio
Precio	S/ 1,190.00	S/ 1,350.00	S/ 1,500.00
Carreras	Varias	Marketing, Enseñanza, Ciencias de la Computación, Administración de Empresas	Marketing, Enseñanza, Ciencias de la Computación, Administración de Empresas
Destinos	América	América - África	América - Asia
Sexo	Mujeres	Hombres	Hombres
Retorno	S/ 803.08	S/ 1,101.69	S/ 1,207.63
Etapas de Estudios	I-VII Ciclo	VIII - X Ciclo	Egresados

Fig. 47. Características del público objetivo de cada programa de intercambio

Por tanto, se decidió elaborar un algoritmo a medida. Este modelo, tras la realización de modificaciones para aumentar el alcance de los reportes por programa, sirvió para realizar la primera clasificación de *leads*. Sobre esto se ejecutó la segmentación detallada en capítulos previos.

[15] provee la comparación de las técnicas Holt Winters, ARMA y ARIMA para brindar soporte en la predicción de ventas, similar a [10] que predecía valores futuros comparando técnicas de minería de datos. Ambos antecedentes, a pesar de contar con funcionalidades que representaban aportes significativos para un sistema, no estaban orientadas al objetivo de segmentación de consumidores planteado en la presente investigación. Para cumplir dicho cometido se analizó los resultados obtenidos por [8], [9] y [12] respecto a la comparación entre técnicas de aprendizaje supervisado y no supervisado. Tras la contrastación de los resultados y sus impactos en las respectivas realidades problemáticas afrontadas se aproximó un modelo con la intención de minimizar las limitaciones sustentadas por los autores correspondientes. De este modo se eligió el algoritmo *k-means* por los siguientes motivos: en primer lugar, el objetivo era obtener clústeres que agrupen *leads* con características similares; en segundo lugar, dicha técnica es utilizada para la segmentación de mercado y entendimiento del cliente; en tercer lugar, se desconocían las características que delimitaban la adquisición del programa en cada perfil de consumidor (muchos registros compatibles para un programa terminaban consumiendo otro). La determinación del número óptimo de clústeres se validó del mismo modo que [14] a través de un índice matemático que regula la idoneidad de la cantidad de grupos.

Para mejorar la calidad de la información ya procesada se usó como referencia a [13] así como otros indicadores correspondientes a requisitos no funcionales debido a la necesidad de apoyar la toma de decisiones a través de tableros útiles y precisos en su contenido, ya que se concluyó en dicho antecedente que el sistema desarrollado mejoraba el nivel de información necesario para apoyar en mayor medida a la gestión estratégica de la empresa.

Los resultados coinciden con lo concluido en [7], puesto que la aplicación de herramientas de BI representó un aporte significativo para AIESEC en Perú y sus procesos internos, brindándole nuevas perspectivas y automatizando un gran número de tareas operativas realizadas en varios niveles de la asociación para dar pie a un nuevo posicionamiento de la figura estratégica de la alta dirección.

Finalmente, cabe destacar que, a diferencia de los antecedentes consultados, en esta investigación se realizó una validación de la calidad del software desarrollado a través del juicio de expertos de TI y de los directivos de la asociación como parte del desarrollo del objetivo planteado.

## VI. CONCLUSIONES

1. La implantación de la aplicación web basada en la segmentación de perfiles de consumidor contribuyó a una mejora en el proceso de conversión de *leads*.
2. Se redujo el tiempo de respuesta a *leads* al mínimo, permitiendo una ventaja de tiempo a los miembros asociados para elaborar un programa y paquete que se ajuste a las necesidades del consumidor.
3. Se segmentó el total de *leads* a través de la aplicación de un algoritmo basado en la segmentación de *k-means* y las políticas internas de la organización, resultando en veinte (20) perfiles de consumidor y seis (06) subgrupos por cada uno de ellos.
4. La presentación de los datos procesados y tableros resaltando la semaforización de indicadores relevantes para la asociación incrementó la calidad de información crítica para la toma de decisiones estratégicas por los altos directivos de la asociación.
5. La personalización del *lead nurturing* es altamente relevante en la experiencia del consumidor, el cual se mostró satisfecho respecto a las funcionalidades brindadas por la aplicación implementada.
6. Se consiguió diseñar una aplicación de calidad considerando los criterios de testeabilidad, disponibilidad y usabilidad.

## VII. RECOMENDACIONES

1. Para trabajos posteriores es importante interconectar los sistemas actuales con el producto acreditable propuesto para poder realizar investigaciones cuasiexperimentales o de experimento puro, para así comparar diferentes grupos y determinar si el software desarrollado es clave en los resultados.
2. Evaluar y ponderar otras características de clientes que sean obtenidas por otros medios fuera de lo solicitado en un formulario de registro para validar que los factores utilizados para la segmentación son los más determinantes para el agrupamiento.
3. La conexión con otras herramientas de minería de datos podría implementarse para soluciones que tengan el objetivo de añadir *machine learning* al sistema desarrollado y así revalidar en tiempo real los modelos resultantes de la minería de datos.
4. Debido a los resultados mostrados por la alta dirección de AIESEC en Perú, el análisis de datos históricos podría replicarse en otras entidades debido a la presencia global de la asociación y uniformizar así la ventaja generada por las herramientas utilizadas.
5. El estudio de mercado realizado por la asociación debe ser complementado con el estudio de los datos históricos con los que cuenta AIESEC para la toma de decisiones estratégicas.

## VIII. LISTA DE REFERENCIAS

- [1] Microsoft Corporation, “*A la Vanguardia: Lecciones en Tecnología y Desarrollo de las PYMEs Líderes del Perú*”. Perú: Microsoft. 2014.
- [2] P. Drucker, “*Do NGOs really need Business Intelligence Solutions?*,” The New Spaces, 2016. [Online]. Available: <https://bit.ly/2Xk9SrJ>. [Accessed: Sep 27, 2018].
- [3] Ernst & Young Global Limited, “*Big data: Changing the way businesses compete and operate*,” London, United Kingdom, 2014. [Online]. Available: <https://go.ey.com/2XfQcFq>. [Accessed: Aug 30, 2019].
- [4] *EY Customer Analytics Platform*, EY, London, United Kingdom, 2019. [Online]. Available: <https://go.ey.com/2QnZCgL>. [Accessed: Aug 31, 2019].
- [5] T. Smith, B. Stiller, J. Guszczka and T. Davenport, “*Analytics and AI-driven enterprises thrive in the Age of With*,” Deloitte Insights, Deloitte Touche Tohmatsu Limited, London, United Kingdom, July 25, 2019. Available: <https://bit.ly/2Qlule7>. [Accessed: Sep 12, 2019].
- [6] Social Media y Contenidos, “*Tasa de conversión en e-Commerce: cifras, informes, sueños y realidad*,” 2014. [En línea]. Disponible en: <https://bit.ly/3571BN2>. [Accedido: ago 5, 2018]
- [7] A. Deligianni, “*Data-driven innovation in NGOs*,” M.S. Thesis, ICT in Business, Leiden Univ., Leiden, The Netherlands, 2018. [Online]. Available: <http://bit.ly/32mYq0f>. [Accessed: Oct 16, 2019].
- [8] J. Naber, “*Achieving Customer Loyalty from Email Campaigns by Using Data Mining Techniques*,” M.S. Thesis, Faculty of Business Economics, Hasselt Univ., Hasselt, Belgium, 2017. [Online]. Available: <http://bit.ly/2nSUtBt>. [Accessed: Oct 16, 2019].
- [9] K. Manero, R. Rimiru and C. Otieno, “Customer Behaviour Segmentation among Mobile Service Providers in Kenya using K-Means Algorithm,” *International Journal of Computer Science Issues (IJCSI)*, vol. 15, no. 5, pp. 67-76, Sept. 2018. [Online]. Available: <http://bit.ly/2oDNhJV>.
- [10] J. Carpio, “*Modelo de predicción de la morosidad en el otorgamiento de crédito financiero aplicando metodología CRISP-DM*,” trabajo de fin de grado, Esc. de Ing. de Sistemas, Univ. Andina "Néstor Cáceres Velásquez", Juliaca, Puno, Perú, 2016.

- [11] M. Grández., *"Aplicación de minería de datos para determinar patrones de consumo futuro en clientes de una distribuidora de suplementos nutricionales,"* trabajo de fin de grado, Esc. de Ing. Informática y de Sistemas, Univ. San Ignacio de Loyola, Lima, Perú, 2017.
- [12] H. Leon, *"Desarrollo de un modelo algorítmico basado en árboles de decisión para la predicción de la permanencia de un paciente en un proceso psicoterapéutico,"* trabajo de fin de grado, Esc. de Ing. Informática, Pontificia Univ. Católica del Perú, Lima, Perú, 2018.
- [13] C. Quepuy, *"Sistema de Gestión y Soporte de toma de decisiones basado en algoritmos de Bayes y Clúster para mejorar los procesos analíticos del área comercial de una empresa educativa,"* trabajo de fin de grado, Esc. de Ing. de Sistemas y Computación, Univ. Católica Santo Toribio de Mogrovejo, Chiclayo, Lambayeque, Perú, 2016.
- [14] L. Gastulo, *"Sistema de gestión de descansos médicos basado en patrones de comportamiento para contribuir a la toma de decisiones del departamento de enfermería de una empresa prestadora de servicios de salud,"* trabajo de fin de grado, Esc. de Ing. de Sistemas y Computación, Univ. Católica Santo Toribio de Mogrovejo, Chiclayo, Lambayeque, Perú, 2018.
- [15] L. Fernández, *"Aplicación de técnicas de minería de datos para pronóstico de producción de espárragos,"* trabajo de fin de grado, Esc. Acad. Prof. de Ing. de Sistemas, Univ. Señor de Sipán, Pimentel, Lambayeque, Perú, 2019.
- [16] F. Solms, "What is Software Architecture," *ACM International Conference Proceeding Series*, 2012. doi: 10.1145/2389836.2389879.
- [17] *Systems and software engineering — Architecture description*, ISO/IEC/IEEE 42010:2011.
- [18] I. Sommerville, *Ingeniería del Software*, 9na ed. Estado de México, México: Pearson Educación, 2011.
- [19] A. Segura, "Arquitectura de software de referencia para objetos inteligentes en internet de las cosas," *Revista Latinoamericana de Ingeniería de Software*, 2016. ISSN 2314-2642.
- [20] Microsoft Corporation, "La Arquitectura Orientada a Servicios (SOA) de Microsoft aplicada al mundo real Microsoft," 2016.

- [21] M. Valipour, B. Amirzafari, K. Maleki and N. Daneshpour, “A Brief Survey of Software Architecture Concepts and Service Oriented Architecture,” in *2nd IEEE International Conference on Computer Science and Information Technology*, 2009, pp. 34-38, doi: 10.1109/ICCSIT.2009.5235004.
- [22] P. Mell and T. Grance, “The NIST Definition of Cloud Computing,” NIST Special Publication 800-145, Gaithersburg, MD, USA, Sept 2011. [Online]. Available: <https://bit.ly/2CMeIEP>. [Accessed: Oct 28, 2019].
- [23] T. Erl, *Cloud Computing: Concepts, Technology & Architecture*. Upper Saddle River, NJ, USA: Prentice Hall, 2013.
- [24] F. Hu et al., “A review on cloud computing: Design challenges in architecture and security,” *Journal of Computing and Information Technology*, vol. 19, no. 1. pp. 25–55, 2011.
- [25] I. Odun-Ayo, S. Misra and F. Agono, “Cloud Computing Architecture: A Critical Analysis” in *18th International Conference on Computational Science and Applications (ICCSA)*, 2018, doi: 10.1109/ICCSA.2018.8439638.
- [26] P. Kotler and G. Armstrong, *Principles of Marketing*, 17th ed. Harlow, United Kingdom: Pearson, 2017.
- [27] “Internet usage statistics: The Internet Big Picture,” *Internet World Stats*, 2015. [Online]. Available: <https://bit.ly/2NPkYBW>. [Accessed: Nov 1, 2015].
- [28] S. Kingsnorth, *Digital Marketing Strategy An integrated approach to online marketing*. Philadelphia, PA, USA: Kogan Page Limited, 2016. [Online]. Available: <https://bit.ly/35747QI>. [Accessed: Oct 30, 2019].
- [29] “Internet Usage Statistics: The Internet Big Picture,” *Internet World Stats*, 2019. [Online]. Available: <https://bit.ly/2QkrXo5>. [Accessed: Nov 14, 2019].
- [30] A. Herrero, “Las claves del nuevo marketing digital,” *Titular.com*, 2018. [En línea]. Disponible en: <https://bit.ly/2XjpUSx>. [Accedido: 28-oct-2019].
- [31] P. Kotler, H. Kartajaya and I. Setiawan, *Marketing 4.0: Moving from Traditional to Digital*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2017.
- [32] “Lead”, *At Internet*, n.d. [En línea]. Disponible en: <https://bit.ly/2XkdkCG>. [Accedido: 14-oct-2019].

- [33] B. Eisenberg, “Jeff Bezos’ Secret to High Conversion Rates,” *Buyer Legends*, 2013. [Online]. Available: <https://bit.ly/357dgJ0>. [Accessed: Sep 14, 2019].
- [34] S. Manola, “How to Calculate, Track, and Analyze Conversion Rates,” *AB Tasty*, 2019. [Online]. Available: <https://bit.ly/2qjKxlv>. [Accessed: Sep 19, 2019].
- [35] “Lead nurturing,” *At Internet*, n.d. [En línea]. Disponible en: <https://bit.ly/2CNBEU3>. [Accedido: 17-oct-2019].
- [36] Marketo, Inc., “What is lead nurturing?,” n.d. [Online]. Available: <https://bit.ly/2qYCMkU>. [Accessed: Oct 18, 2019].
- [37] J. Han, M. Kamber and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Waltham, MA, USA: Elsevier Inc., 2012. [Online]. Available: <https://bit.ly/33RqauD>. [Accessed: Oct 31, 2019].
- [38] J. Gironés Roig, “Data mining,” *Univ. Iberta de Catalunya*, 2013, PID\_00203552. [En línea]. Disponible en: <https://bit.ly/37d5Ju3>. [Accedido: 25-oct-2019].
- [39] S. Weiss and N. Indurkha, *Predictive Data Mining - A Practical Guide*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1998.
- [40] M. García, L. Quintales, F. García-Peñalvo y M. Martín, “*Aplicación de Técnicas de Minería de Datos en la Construcción y Validación de Modelos Predictivos y Asociativos a Partir de Especificaciones de Requisitos de Software*,” Univ. De Salamanca, 2001. [En línea]. Disponible en: <https://bit.ly/2COvJhq>. [Accedido: 26-oct-2019].
- [41] L. Rokach and O. Maimon, *Data Mining with Decision Trees*, 2nd ed. Hackensack, NJ, USA: World Scientific Publishing Co. Pte. Ltd., 2015. [Online]. Available: <https://bit.ly/32PnrAG>. [Accessed: Oct 25, 2019].
- [42] A. Jain, J. Mao and K. Mohiuddin, “Artificial neural networks: A tutorial,” *Computer*, vol. 29, no. 3, pp. 31-44, 1996. [Online]. Available: <https://bit.ly/2XkPypX>. [Accessed: Oct 27, 2019].
- [43] D. Olson and D. Delen, *Advanced Data Mining Techniques*, Berlin, Germany: Springer – Verlag Berlin Heidelberg, 2008.
- [44] P. Esling and C. Agon, Time-series data mining, *ACM Computing Surveys*, vol. 45, no. 1, 2012, doi: 10.1145/2379776.2379788.

- [45] A. Zimek and E. Schubert, "Outlier Detection," in *Encyclopedia of Database Systems*, New York, NY, USA: Springer Science+Business Media, LLC, 2017, doi: 10.1007/978-1-4899-7993-3\_80719-1.
- [46] J. Dean, *Big Data, Data Mining and Machine Learning: Value Creation for Business Leaders and Practitioners*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2017.
- [47] J. Krzysztof, P. Witold, W. Roman and A. Lukasz, *Data Mining: A Knowledge Discovery Approach*. New York, NY, USA: Springer Science+Business Media, LLC, 2007.
- [48] N. Mabroukeh and C. Ezeife, "A Taxonomy of Sequential Pattern Mining Algorithms," *ACM Computing Surveys (CSUR)*, vol. 43, no. 1, 2010, doi: 10.1145/1824795.1824798.
- [49] "Orange," *Univ. of Ljubljana*, n.d. [Online]. Available: <https://orange.biolab.si/>. [Accessed: Oct 30, 2019].
- [50] "Weka 3: Machine Learning Software in Java," *Univ. of Waikato*, n.d. [Online]. Available: <https://www.cs.waikato.ac.nz/ml/weka/>. [Accessed: Oct 30, 2019].
- [51] "Lightning Fast Business Impact with AI," *Rapidminer, Inc.*, 2019. [Online]. Available: <https://rapidminer.com/>. [Accessed: Oct 30, 2019].
- [52] "The Forrester Wave™: Multimodal Predictive Analytics And Machine Learning Solutions, Q3 2018," *Forrester Research, Inc.*, 2018. [Online]. Available: <https://bit.ly/3574OJO>. [Accessed: Oct 30, 2019].
- [53] "Magic Quadrant for Data Science and Machine Learning Platforms," *Gartner, Inc.*, 2019. [Online]. Available: <https://gtmr.it/2QoabQY>. [Accessed: Oct 30, 2019].
- [54] "Software de data mining: realiza análisis de datos más efectivos," *IONOS España S.L.U.*, 2018. [En línea]. Disponible en: <https://bit.ly/2q5XM9z>. [Accedido: 31-oct-2019].
- [55] U. Fayyad, G. Piatetsky-Shapiro and P. Smyth, "From Data Mining to Knowledge Discovery in Databases," *AI Magazine*, vol. 17, no. 3, 1996.
- [56] H. Camargo y M. Silva, "Dos caminos en la búsqueda de patrones por medio de Minería de Datos: SEMMA y CRISP," *Tecnol Journal of Technology*, vol. 9, no. 1, pp. 11-18, 2010.

- [57] A. Azevedo and M. Santos, “KDD, SEMMA and CRISP-DM: A Parallel Overview,” in *IADIS European Conf. Data Mining 2008*, pp. 182-185.
- [58] P. Chapman et al., “CRISP-DM 1.0 Step-by-step data mining guide,” SPSS, USA, 2000. [Online]. Available: <https://bit.ly/32QIVxZ>. [Accessed: Oct 14, 2018].
- [59] G. Piatetsky-Shapiro, “CRISP-DM, still the top methodology for analytics, data mining, or data science projects,” *KDnuggets*, 2014. [Online]. Available: <https://bit.ly/2Qo0CI1>. [Accessed: Nov 1, 2019].
- [60] P. Lagos, “CommonKADS y el Lenguaje de Modelado Unificado æ UML,” *Ingeniería informática*, no. 8, 2002, ISSN 0717-4195.
- [61] A. Martínez y R. Martínez, “Guía a Rational Unified Process,” *Escuela Politécnica Superior de Albacete–Universidad de Castilla la Mancha*, 2014. [En línea]. Disponible en: <https://bit.ly/2prX9a0>. [Accedido: 01-nov-2019].
- [62] R. Hernández, C. Fernández y M. Baptista, *Metodología de la Investigación*, 6ª ed. México D.F., México: McGRAW-HILL/Interamericana Editores, S.A. de CV., 2014.
- [63] C. Bernal, “Proceso de la investigación científica,” en *Metodología de la investigación científica: administración, economía, humanidades y ciencias sociales*, 3ra. ed. Bogotá D.C., Colombia: Pearson Educación, 2010.
- [64] *Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts*, ISO 9241-11: 2018.

**IX. ANEXOS**

**ANEXO N° 01. CONSTANCIA DE APROBACIÓN DEL PRODUCTO  
ACREDITABLE DE LA ENTIDAD DONDE SE EJECUTÓ LA TESIS**



Lima, Perú  
28 de octubre, 2019

**CONSTANCIA DE APROBACIÓN DE PRODUCTO  
ACREDITABLE**

Por medio de la presente certificamos la aprobación de la aplicación web presentada por el **Sr. WALTER MIGUEL RIOJA CURO** como producto acreditable de su trabajo de investigación de fin de grado, cumpliendo los requisitos establecidos y objetivos planteados por este.

Se expide la presente constancia a petición del interesado para los fines que crea conveniente.

Atentamente,

ASOCIACIÓN CIVIL LIDERAZGO  
E INTERCAMBIO  
R.U.C. 20543380319

**Maria Camila Padilla Puello**  
Presidenta y Representante Legal  
C.E 001671476

---

ASOCIACIÓN CIVIL LIDERAZGO E INTERCAMBIO  
Avenida República de Panamá 3563 - Of. 701B, San Isidro - Lima, Perú

Fig. 48. Carta de aprobación del producto acreditable

## ANEXO N° 02. ANÁLISIS DE RIESGOS

### 1. Datos generales

- **Tesista** : Walter Miguel Rioja Curo
- **Fecha inicial** : 04 de setiembre de 2019
- **Fecha final** : 02 de noviembre de 2019

### 2. Alcance del proyecto

- Se desarrolló una aplicación web para apoyar el proceso de conversión de *leads* en la asociación AIESEC en Perú haciendo uso de una arquitectura *cloud* y minería de datos. El software permite clasificar *leads* a través de subgrupos de perfiles de consumidor, segmentando a los interesados en adquirir alguno de los programas de intercambio con el objetivo de convertirlos en consumidores.

### 3. Interesados (Stakeholders)

Durante el desarrollo de la presente tesis se ha identificado a los siguientes interesados:

- **Internos**

TABLA XXXIII: INTERESADOS INTERNOS

<b>Interesado</b>	<b>Participación</b>
Junta directiva nacional de AIESEC en Perú	Alta dirección de la asociación civil, proveyendo información relevante de la situación problemática para la tesis.
Miembros de oficinas locales de la asociación	Miembros del sector táctico y operacional de la asociación, probando el sistema para determinar el cumplimiento de los objetivos específicos.

- **Externos**

TABLA XXXIV: INTERESADOS EXTERNOS

<b>Interesado</b>	<b>Participación</b>
Mgr. Juan Antonio Torres Benavides	Asesor de la presente investigación, siendo consultor y apoyo durante todo el desarrollo de esta.
Mgr. Marlon Eugenio Vílchez Rivas	Docente de la asignatura de Seminario de Tesis II, revisando los aspectos metodológicos y la forma de la tesis.

### 4. Beneficios

Los beneficios que se van a obtener con el producto que se ha desarrollado son:

- Generar perfiles y segmentos de *leads* en base a características comunes, facilitando su gestión y contacto.
- Reducir el tiempo de respuesta a un *lead* tras su registro en la plataforma.
- Mejorar la información mostrada a los gerentes para apoyar la toma de decisiones estratégicas, así como a los *leads* con el objetivo de agilizar la conversión.

## 5. Análisis de riesgos

Para el análisis de riesgos correspondiente se tomó la decisión de considerar aquellos presentes durante el desarrollo del proyecto frente a la implantación del software, así como las amenazas del funcionamiento de la aplicación. Tras realizar la criticidad de ellos y la definición de controles que gestionen los riesgos se obtuvo lo siguiente:

- **Matriz de riesgos**

Entre los riesgos identificados durante el desarrollo del proyecto se encontró:

TABLA XXXV: MATRIZ DE RIESGOS DEL PROYECTO

Código del riesgo	Descripción del riesgo	Principio afectado	Causa raíz	Activos afectados	Estimación probabilidad	Objetivo afectado	Estimación Impacto	Probabilidad * impacto	Nivel de riesgo
RE1 - 001	Divulgación de datos personales sensibles de consumidores.	Confidencialidad Integridad	Fugas de información	Base de datos	4	Alcance	3	12	<b>ALTO</b>
						Tiempo	3	12	
						Costo	5	20	
						Calidad	5	20	
						<b>Total probabilidad * impacto</b>		<b>64</b>	
RE1 - 002	Incompatibilidad entre los requisitos iniciales y las funcionalidades de la aplicación desarrollada.	Disponibilidad	Mala extracción de requerimientos	Aplicación	2	Alcance	5	10	<b>MEDIO</b>
						Tiempo	5	10	
						Costo	5	10	
						Calidad	5	10	
						<b>Total probabilidad * impacto</b>		<b>40</b>	
RE1 - 003	Interrupción del funcionamiento del servidor de aplicaciones o datos.	Disponibilidad	Cortes eléctricos, desastres naturales	Base de datos Aplicación	3	Alcance	3	9	<b>BAJO</b>
						Tiempo	2	6	
						Costo	3	9	
						Calidad	3	9	
						<b>Total probabilidad * impacto</b>		<b>33</b>	
RE1 - 004	Demoras prolongadas en la respuesta del sistema tras peticiones de la aplicación.	Disponibilidad	Saturación del servidor	Aplicación	4	Alcance	3	12	<b>MEDIO</b>
						Tiempo	4	16	
						Costo	1	4	
						Calidad	3	12	
						<b>Total probabilidad * impacto</b>		<b>44</b>	

• **Matriz de controles**

Con el objetivo de gestionar los riesgos identificados se plantearon los siguientes controles:

TABLA XXXVI: MATRIZ DE CONTROLES DE RIESGOS

Código del control	Amenaza / Oportunidad	Descripción del control	Riesgo(s) controlado	Nivel de riesgo	Tipo de acción	Responsable	Plan de mitigación
CT – 001	Fuga de información / Seguridad de datos sensibles	Fortalecimiento de la seguridad de contraseñas y segregación de funciones según rol y rango para la visualización de datos.	RE1 – 001	ALTO	Mitigación de riesgo	Tesista	<ul style="list-style-type: none"> <li>✓ Sugerencia de políticas de seguridad más exigentes en la organización.</li> <li>✓ Implementación de permisos en la aplicación según la jerarquía.</li> <li>✓ Protección de datos sensibles presentados en el desarrollo de la investigación.</li> </ul>
CT – 002	Inconsistencia del sistema / Software de calidad	Verificación periódica de las funcionalidades del software y los requisitos especificados.	RE1 – 002	MEDIO	Mitigación de riesgo	Tesista	<ul style="list-style-type: none"> <li>✓ Ejecución de reuniones periódicas con los directivos de la organización para validación de las funcionalidades de la aplicación.</li> <li>✓ Implicación del tesista con la realidad problemática para determinar otras posibles funciones o problemas abordados.</li> </ul>
CT – 003	Interrupción del funcionamiento del servidor / Continuidad de los procesos	Migración a un servidor de base de datos y aplicaciones tercerizado.	RE1 – 003	BAJO	Transferencia	Tesista	<ul style="list-style-type: none"> <li>✓ Contratación de servicios de un proveedor de servidores de aplicaciones y base de datos externo.</li> <li>✓ Migración de datos reales de la asociación a través de sus distintas fuentes.</li> <li>✓ Limpieza de datos para garantizar consistencia en la nueva base de datos.</li> </ul>
			RE1 – 004	MEDIO			

## ANEXO N° 03. LISTA DE COTEJOS PARA USABILIDAD DEL SOFTWARE

### PRUEBA DE USABILIDAD

#### 1. Datos

- **Nombre:**
- **Cargo:**

#### 2. Introducción

El estándar ISO/IEC 9241-11:2018 define la usabilidad como “el grado en el que un producto puede ser utilizado por usuarios específicos para conseguir objetivos específicos con efectividad, eficiencia y satisfacción en un determinado contexto de uso”. Por otro lado, en el estándar ISO/IEC 25010:2011 se define la usabilidad como una contribución relativamente independiente a la calidad del software asociado con el diseño y la evaluación de la interfaz del usuario y su interacción. Bajo estas premisas se elaboró la siguiente evaluación de usabilidad en donde se fijaron algunos parámetros de medición para el software.

Para que la evaluación sea efectiva se recomienda realizar previamente las siguientes tareas:

- Estudiar previamente el software para familiarizarse con ella.
- Entender todos los criterios de usabilidad utilizados en la prueba.

#### 3. Objetivo

Evaluar el software en base a los criterios de usabilidad propuestos en el estándar ISO 9421.

#### 4. Evaluación

	No aplica	Muy en desacuerdo	En desacuerdo	De acuerdo	Muy de acuerdo
VALOR	0	1	2	3	4

CRITERIOS DE USABILIDAD	NA	MED	ED	DE	MDE
<b>FACILIDAD DE APRENDIZAJE</b>					
El ingreso datos y los resultados mostrados en el software son fáciles de entender					
El usuario puede explorar con facilidad el software, sin llegar a un punto del cual no pueda salir					
El software permite volver a pasos anteriores para modificar los datos previamente ingresados					
El software provee indicaciones que permitan al usuario completar las acciones que se consideren complejas					
El usuario puede identificar fácilmente las figuras, tablas, elementos clickeables y el tipo de acción que debe ejecutar cada elemento del software					
<b>RECUERDO EN EL TIEMPO</b>					

El software tiene la capacidad de ser usado sin tener que hacer mucho esfuerzo para recordar su funcionamiento					
Es muy fácil recordar como ejecutar las acciones disponibles dentro del software					
<b>COMPREENSIBILIDAD</b>					
Los mensajes emitidos dentro del software expresan una idea clara y concisa					
Los párrafos contenidos en el software son entendibles y mantienen uniformidad a través de todas sus secciones					
El tamaño de las imágenes e íconos usados dentro del software es adecuado para saber qué representan					
El lenguaje utilizado dentro del software es amigable, familiar y cercano					
Los tipos y tamaños de letra usados dentro del software son legibles y distinguibles					
<b>TOLERANCIA A ERRORES</b>					
El software ayuda a los usuarios a evitar errores					
El software ayuda a los usuarios a reconocer y recuperarse de los errores					
El software informa adecuadamente sobre los errores que podrían presentarse en el proceso					
<b>ATRACTIVIDAD</b>					
Los títulos de las figuras y tablas son íntegramente descriptivos y distintivos					
El software no despliega excesiva información dentro de su interfaz					
El software utiliza notificaciones o emergentes para mostrar información importante					
La posición de los elementos (botones, mensajes, información, etc.) dentro del software es el adecuado					
La apariencia del software es visualmente agradable y sencilla					
<b>SATISFACCIÓN</b>					
El software presenta contenido que es de auténtica utilidad para el usuario					
El software puede ser considerado una herramienta que no provoque frustración al utilizarlo					
El software cumple con todas las expectativas que el usuario tiene sobre este					
El usuario volvería a usar el software con más frecuencia para sus labores					



**Hey AIESEC!**

**MANUAL DE USUARIO**



## **Hey AIESEC!**

**MANUAL DE USUARIO**

### **Especificaciones Técnicas**

Navegador web  
recomendado: Google  
Chrome.

Desarrollado por Walter  
Miguel Rioja Curo.

Programado con  
herramientas y lenguajes  
PHP nativo, JavaScript,  
HTML 5 y CSS.

Basado en framework de  
Bootstrap AdminLTE.

Año de lanzamiento:  
2019

Ruta de acceso al  
sistema:  
[cy-mag.com/aiesec](http://cy-mag.com/aiesec)

Contacto:

[walter.riojac@gmail.com](mailto:walter.riojac@gmail.com)  
/ 945 723 623



## MANUAL DE USUARIO POR FORMULARIOS Y COMPONENTES

### 1. Inicio de sesión



Para acceder al sistema, es necesario ingresar un correo electrónico y contraseña. Son los mismos proporcionados por la organización, por lo que debe consultar en su oficina local para obtener estos datos.

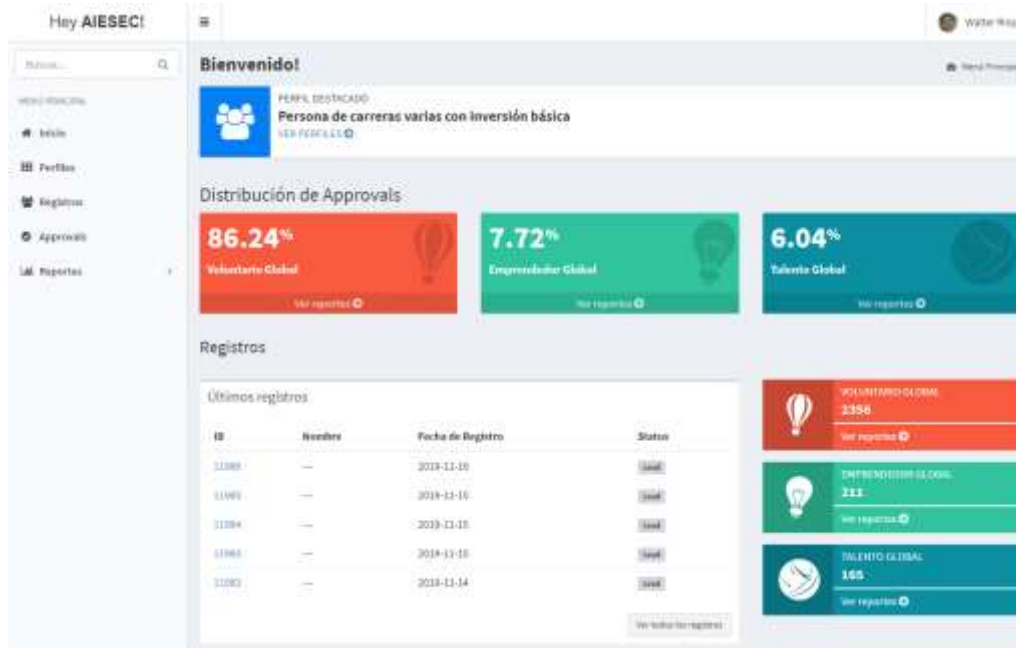
### 2. Barra de navegación



La aplicación cuenta con una barra de navegación ubicada en el lado izquierdo de la pantalla. Esta muestra todas las interfaces a las que se puede acceder para probar las funcionalidades del software. Cada una de ellas será detallada en el presente manual de usuario en los ítems siguientes.




### 3. Menú principal



El menú principal muestra la información más relevante respecto a últimos leads con su respectivo status, perfil más destacado y distribución absoluta y relativa de los approvals según el producto elegido. Para ver detalles de cada uno solo es necesario dar clic en cualquiera de los links que detallan su redireccionamiento.

### 4. Perfiles

 [Perfiles](#)

Esta interfaz brinda el listado de todos los perfiles de consumidor registrados en el sistema basado en la investigación realizada.

Código	Descripción	Registros	Opciones
1	Estudiante en ciclos básicos de carreras elegibles a subproductos con alta inversión	1081	<a href="#">Ver perfil</a>
2	Van en ciclos de especialización de carreras elegibles a subproductos con interés en viajar por Americas o Middle East and Africa (MEA) con alta inversión durante primer o segundo semestre del año.	48	<a href="#">Ver perfil</a>
3	Viajar en ciclos de especialización de carreras elegibles a subproductos con interés en viajar por Americas o Middle East and Africa (MEA) con alta inversión durante primer o segundo semestre del año.	147	<a href="#">Ver perfil</a>
4	Van en ciclos de especialización de carreras elegibles a subproductos con interés en viajar por Americas o Middle East and Africa (MEA) con alta inversión durante vacaciones de verano o invierno.	29	<a href="#">Ver perfil</a>

Para acceder a cada uno de ellos solo es necesario dar clic en el botón *Ver Perfil*.

[Ver perfil](#)



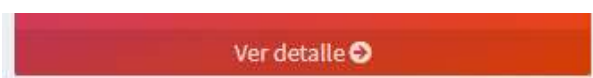
## 5. Detalle del perfil



Tras hacer clic al detalle del perfil que haya elegido, se muestra una pantalla con dos partes. Primero, un detalle del perfil junto al producto al que califica, así como gráficos que pueden verse en diferentes pantallas o guardarse en el escritorio, tal y como se muestra en la figura. La segunda parte presenta un listado de todos los leads dentro de ese perfil, al que se puede acceder para ver sus detalles dando clic en el botón *Ver registro*.



Para ver el detalle de los subgrupos que comparten características similares, solo es necesario dar clic en el botón *Ver detalle*, dentro de cada caja de grupo.





## 6. Detalle del subgrupo

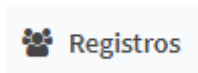
The screenshot displays the 'Registros del Perfil #8 - Subgrupo #03' interface. At the top, there is a navigation bar with 'Inicio', 'Perfil', 'Perfil #8', and 'Grupo #03'. Below this, a section titled 'DESCRIPCIÓN DEL PERFIL' contains the text: 'Egresado de carreras elegibles a subproducto con interés en viajar por Middle East and Africa (MEA) con alta inversión durante primer o segundo semestre del año'. A message states 'Perfil elegible para los productos' with buttons for 'Talento Global' and 'Voluntario Global'. A table lists leads with columns for 'Fecha de Registro', 'Nombre', 'Correo', 'Secretaría', and 'Estado'. Below the table is an 'Enviar correo' section with a text area for 'Asunto', a rich text editor for the body, and an 'Enviar' button.

Esta es una de las interfaces más importantes del sistema. En ella podrá revisar la lista de leads no convertidos del perfil que haya elegido, nuevamente con la opción de ver los productos a los que califica y un campo de texto donde podrá escribir un correo donde ofrezca un producto a medida, promoción u otras opciones que considere apropiadas con el objetivo de convertir a los leads pertenecientes a dicho grupo. Para enviar el correo solo se requiere dar clic en el botón *Enviar*, después de agregar el asunto y modificar el cuerpo del correo. Tras ello, se le solicitará una verificación a través de una notificación emergente. Brinde su confirmación para proceder al envío de correos electrónicos.

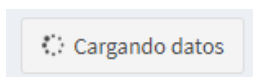




## 7. Registros



En esta pantalla se listan todos los leads (independientemente de su estado según el customer funnel), mostrando información más relevante. Al hacer click en esta función desde el menú se cargará una interfaz con un mensaje de espera para la carga de datos. Dependiendo de su velocidad de internet, esta interfaz tardará en cargar.



Una vez haya cargado, se listará en bloques de 100 todos los leads junto a datos relevantes como carrera, departamento, perfil y status. El orden por defecto es en base a la fecha de registro, sin embargo, se puede ordenar por la característica de preferencia al hacer clic sobre ella. Para buscar algún nombre en específico solo es necesario ingresar su nombre en la opción de *Búsqueda*. Asimismo, para observar los detalles de este lead solo es necesario dar clic en el botón *Ver registro*.

Código	Fecha de Registro	Nombre	Carrera	Departamento	Perfil	Status	Opciones
11880	2018-12-25	---	Ciencias de la Comunicación	Lima	25	Lead	Ver registro
11881	2018-12-25	---	Marketing	Lima	6	Lead	Ver registro
11884	2018-12-25	---	Ciencias de la Salud	Jurín	25	Lead	Ver registro
11885	2018-12-25	---	Ciencias de la Comunicación	Lima	25	Approved	Ver registro

Tras dar clic en el lead elegido, se mostrará información según su status respecto al producto elegido, perfil y subgrupo al que pertenece, así como sus datos personales ingresados.

**Datos de Lead**

**Carrito de Registros**

Fecha de Registro: 2018-12-25

Nombre: [Empty field]

Fecha de Radicación: 2018-12-25

Ella: [Empty field]

Departamento: [Empty field]

Carrera (Elegida): [Empty field]

Perfil: [Empty field]

Status: [Empty field]

Nombre y Apellido: [Empty field]

Carrera: [Empty field]

Cita: [Empty field]

Código de Lead: [Empty field]

Ver registro: [Empty field]

**Sobre el perfil:**

Perfil: [Empty field]

Subgrupo: [Empty field]

Producto elegido: [Empty field]

**Datos de Approval:**

Fecha de APT: 2018-12-25

Producto: [Empty field]

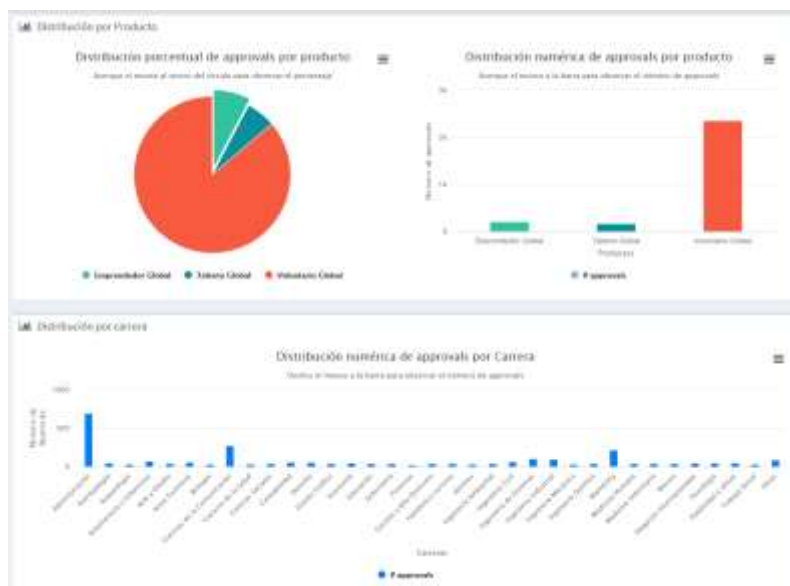
Fecha de Apto: [Empty field]



## 8. Approvals

### Approvals

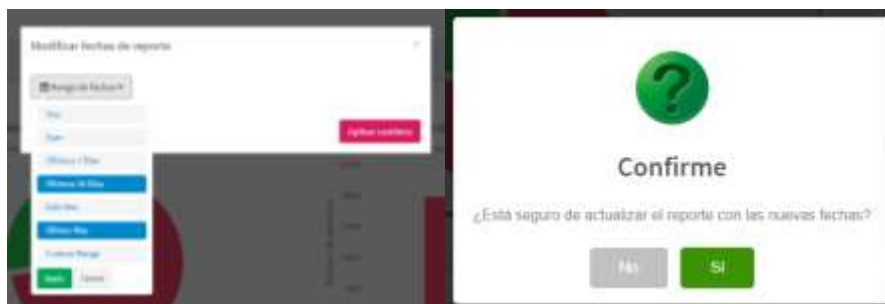
En esta pantalla se brinda un reporte de todos los approvals en base a sexo, producto elegido, carrera y región de origen. Las gráficas tienen las mismas propiedades explicadas previamente para exportar el reporte en el formato deseado.



Para modificar el reporte en base a las fechas solo es necesario dar clic en el botón magenta *Modificar fechas de reporte*.

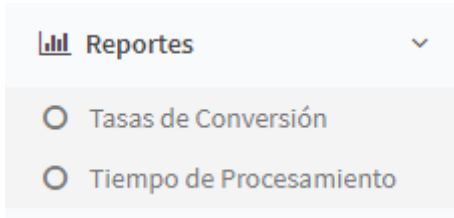
 **Modificar fechas de reporte**

Tras elegir la fecha deseada, solo es necesario solicitar aplicar los cambios y confirmar en el mensaje emergente. El reporte se actualizará automáticamente.





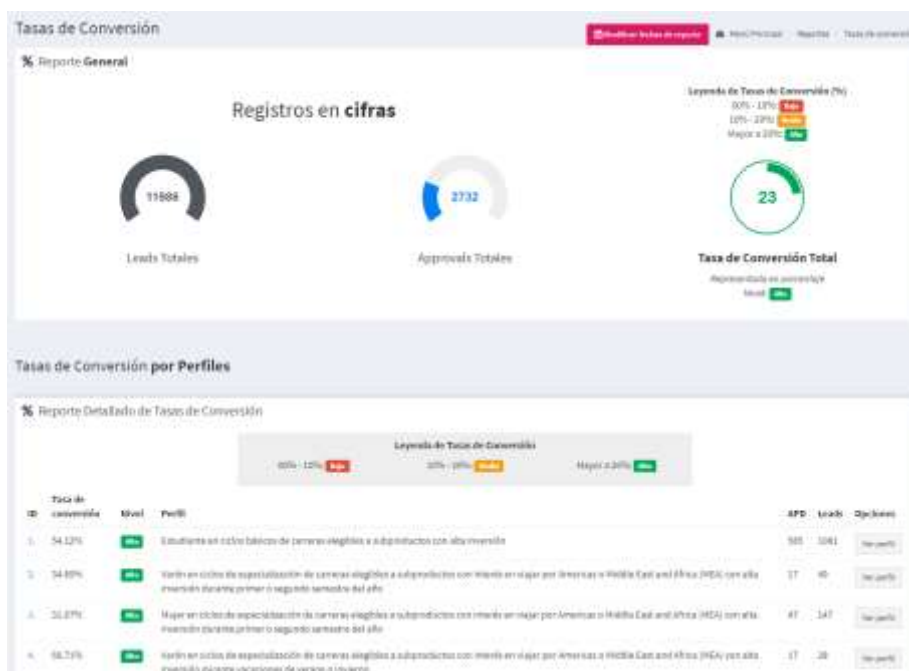
## 9. Reportes



### 9.1. Tasas de conversión

Este reporte cuenta con dos partes. Por un lado, muestra la tasa de conversión general de lead/approval, semaforizada según los indicadores establecidos como meta de la gestión de MC respectiva. Por el otro, muestra una tasa de conversión detallada por perfiles de consumidor con los mismos criterios de evaluación. De este modo, podrá analizar qué perfiles están convirtiendo mejor y tomar decisiones al respecto.

Del mismo modo que el reporte de approvals, se puede modificar en base a las fechas de preferencia para analizar y tras la confirmación el reporte se actualizará automáticamente.



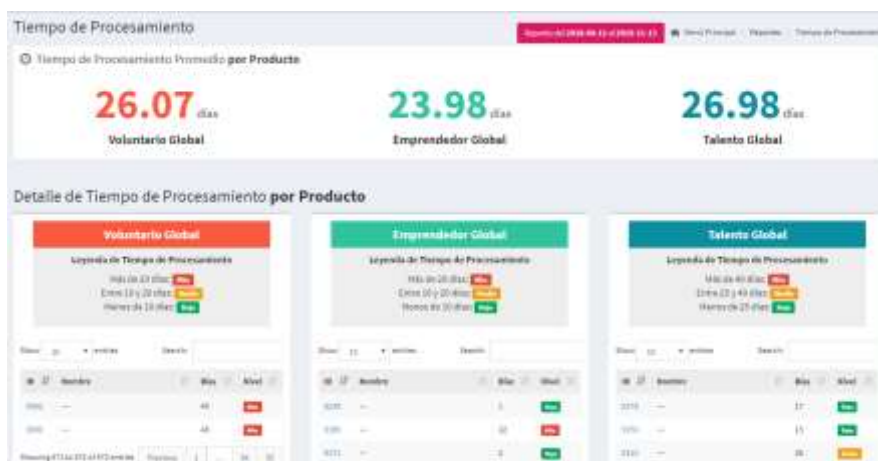
Para ir al perfil de preferencia solo es necesario dar clic en el botón *Ver Perfil* presente en la última fila de la tabla de tasa de conversión por perfiles.





## 9.2. Tiempo de procesamiento

Este último reporte cuenta con tres bloques. El primero, muestra el tiempo de procesamiento promedio por producto; el segundo, el detalle de esto, listando los approvals y su tiempo respectivamente semaforizado; y el tercero, detallando el mismo indicador, pero por perfiles. Asimismo, se puede modificar el reporte por fechas.

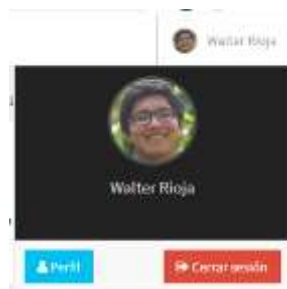


Del mismo modo que el anterior reporte, se cuenta con navegabilidad entre esta interfaz y el detalle del perfil al dar clic en el botón de *Ver perfil*.



## 10. Cerrar sesión

No olvide cerrar su sesión tras finalizar su uso del sistema. Solo es necesario dar clic en su nombre y después en el botón *Cerrar sesión*.



## ANEXO N° 05. REPORTES DE GESTIÓN 2017-2018

La información correspondiente a datos de la asociación fue brindada por la alta dirección de AIESEC en Perú en las fechas solicitadas a través de reportes de gestión. Dichos documentos no pueden ser divulgados, sin embargo, se adjunta la evidencia del envío de los archivos correspondientes por parte de la presidenta de la organización.



Fig. 49. Captura de pantalla del correo conteniendo los reportes de gestión 2017-2018