

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ

Colegio de Ciencias e Ingenierías

**Aplicación de Institutional Research para apoyar la toma de
decisiones en la USFQ**

**Esteban Daniel Burbano Benalcázar
Cristina Belén Gallardo Villamarín**

Ingeniería Industrial

Trabajo de fin de carrera presentado como requisito
para la obtención del título de
Ingeniero Industrial

Quito, 18 de mayo de 2021

UNIVERSIDAD SAN FRANCISCO DE QUITO USFQ
Colegio de Ciencias e Ingenierías

HOJA DE CALIFICACIÓN
DE TRABAJO DE FIN DE CARRERA

**Aplicación de Institutional Research para apoyar la toma de decisiones en
la USFQ**

Esteban Daniel Burbano Benalcázar
Cristina Belén Gallardo Villamarín

Nombre del profesor, Título académico

Danny Navarrete, M. Sc.

Quito, 18 de mayo de 2021

© DERECHOS DE AUTOR

Por medio del presente documento certifico que he leído todas las Políticas y Manuales de la Universidad San Francisco de Quito USFQ, incluyendo la Política de Propiedad Intelectual USFQ, y estoy de acuerdo con su contenido, por lo que los derechos de propiedad intelectual del presente trabajo quedan sujetos a lo dispuesto en esas Políticas.

Asimismo, autorizo a la USFQ para que realice la digitalización y publicación de este trabajo en el repositorio virtual, de conformidad a lo dispuesto en la Ley Orgánica de Educación Superior del Ecuador.

Nombres y apellidos: Esteban Daniel Burbano Benalcazar

Código: 00137260

Cédula de identidad: 1723337637

Lugar y fecha: Quito, 18 de mayo de 2021

Nombres y apellidos: Cristina Belén Gallardo Villamarín

Código: 00132719

Cédula de identidad: 1719159921

Lugar y fecha: Quito, 18 de mayo de 2021

ACLARACIÓN PARA PUBLICACIÓN

Nota: El presente trabajo, en su totalidad o cualquiera de sus partes, no debe ser considerado como una publicación, incluso a pesar de estar disponible sin restricciones a través de un repositorio institucional. Esta declaración se alinea con las prácticas y recomendaciones presentadas por el Committee on Publication Ethics COPE descritas por Barbour et al. (2017) Discussion document on best practice for issues around theses publishing, disponible en <http://bit.ly/COPETHeses>.

UNPUBLISHED DOCUMENT

Note: The following capstone project is available through Universidad San Francisco de Quito USFQ institutional repository. Nonetheless, this project – in whole or in part – should not be considered a publication. This statement follows the recommendations presented by the Committee on Publication Ethics COPE described by Barbour et al. (2017) Discussion document on best practice for issues around theses publishing available on <http://bit.ly/COPETHeses>.

RESUMEN

El Institutional Research (IR) es el departamento encargado del aseguramiento de la calidad de la educación y la toma de decisiones en muchas instituciones de educación superior a nivel mundial. El IR tiene dos definiciones más aceptadas. La primera como un departamento de apoyo para la toma de decisiones. La segunda, como inteligencia organizacional de tres niveles: técnico y analítico, inteligencia de problemas e inteligencia contextual. Estados Unidos al ser el primero en el desarrollo del IR, ha servido como un referente para el resto del mundo, registrando hasta la actualidad 9 asociaciones de IR en todas las regiones exceptuando Latinoamérica.

Las funciones, partes beneficiadas y proyectos del IR son variadas, sin embargo, en el presente trabajo se diseñó un sistema de visualización de métricas clave para la USFQ donde se rastre y anticipe los cambios y tendencias de dichas métricas referentes a estudiantes, personal y publicaciones para apoyar a la toma de decisiones. Para esto se realizó un benchmarking de 100 departamentos de IR de universidades de Norteamérica para definir los elementos de este, así como también un análisis de la mejor herramienta de Business Intelligence para la elaboración de un factbook donde se muestre esta información institucional, siendo Tableau la herramienta seleccionada. Finalmente, se desarrolló un modelo de regresión logística, como parte de uno de los proyectos del IR que prediga la deserción de los estudiantes en base a la información previa a su ingreso a la universidad, donde con una precisión de 69% presentar características como estudiante transferido, hombre, que estudiará medicina veterinaria, animación digital, biología, matemática, artes musicales, física o agronomía aumenta las probabilidades de ser un estudiante desertor.

Palabras clave: Institutional Research, educación superior, deserción, Machine Learning, Business Intelligence. Regresión logística, Backward Elimination.

ABSTRACT

The Institutional Research (IR) is the department in charge of the quality assurance of education and decision-making in many higher education institutions worldwide. IR has two more accepted definitions. The first as a support department for decision making. The second, as three-level organizational intelligence: technical and analytical, issues intelligence, and contextual intelligence. The United States, being the first in the development of the IR, has served as a benchmark for the rest of the world, registering to date 9 IR associations in all regions except Latin America.

The functions, benefited parties and projects of the IR are varied, however, in the present work a system of visualization of key metrics for the USFQ was designed to track and anticipate the changes and trends of those metrics that refers to students, staff and publications to support decision-making. For this, a benchmarking of 100 IR departments of North American universities was carried out to define the elements of the department, as well as an analysis of the best Business Intelligence tool for the elaboration of a factbook where this institutional information is shown, Tableau was the selected tool. Finally, a logistic regression model was developed as part of one of the IR projects that predicts student dropout based on information prior to entering the university. The results showed that with a precision of 69%, a student with characteristics such as transferred, male, that will study veterinary medicine, digital animation, biology, mathematics, music, physics, or agronomy increases the chances of a dropout.

Key words: Institutional Research, educación superior, deserción, Machine Learning, Bussines Intelligence. Logistic Regression, Backward Elimination

TABLA DE CONTENIDO

1. INTRODUCCIÓN	11
1.1. OBJETIVOS	13
1.1.1. <i>Objetivo General</i>	13
1.1.2. <i>Objetivos Específicos</i>	13
1.2. ALCANCE	13
2. REVISIÓN DE LITERATURA	14
2.1. INSTITUTIONAL RESEARCH	14
2.2. ANALÍTICA DE DATOS	20
3. METODOLOGÍA	22
3.1. BENCHMARKING INSTITUTIONAL RESEARCH	22
3.1.1. <i>Herramientas de Business Intelligence</i>	23
3.1.2. <i>El Factbook</i>	26
3.1.3. <i>Identificación de indicadores</i>	26
3.1.4. <i>Identificación de indicadores</i>	26
3.2. ANALÍTICA DE DATOS	27
3.2.1. <i>Backward Elimination</i>	29
3.2.2. <i>Partición de los datos</i>	29
3.2.3. <i>Métricas de evaluación</i>	30
4. RESULTADOS	32
4.1. INTITUTIONAL RESEACH	32
4.1.1. <i>Benchmarking del IR de las 100 Universidades</i>	32
4.1.2. <i>La herramienta BI</i>	34
4.1.3. <i>Factbook</i>	37
4.1.4. <i>Dashboards KPIs</i>	38
4.2. ANALÍTICA DE DATOS	38
4.2.1. <i>Descripción de los conjuntos de datos y su preprocesamiento</i>	39
4.2.2. <i>La variable de respuesta</i>	40
4.2.3. <i>Problema desbalanceado</i>	40
4.2.4. <i>Backward Elimination</i>	40
4.2.5. <i>El modelo</i>	41
4.2.6. <i>Resultados del modelo</i>	42
4.2.7. <i>Coeficientes</i>	44
5. CONCLUSIONES Y RECOMENDACIONES	45
5.1. INSTITUTIONAL RESEARCH	45
5.2. ANALÍTICA DE DATOS	47
6. LIMITACIONES	49
7. SIGUIENTES PASOS	50
ANEXOS:	54
ANEXO A: CRITERIOS PARA LA IDENTIFICACIÓN DE MÉTRICAS DENTRO DE LAS UNIVERSIDADES.	54
ANEXO B: CRITERIOS PARA LA IDENTIFICACIÓN DE MÉTRICAS DENTRO DE LAS UNIVERSIDADES.	54
ANEXO C: ORGANIZACIÓN DE MÉTODOS DE MINERÍA DE DATOS DE ACUERDO CON LA NATURALEZA DE LOS DATOS.	55
ANEXO D: PLANTILLAS DE REQUERIMIENTO DE DASHBOARD	55
ANEXO E: PLANTILLAS DE REQUERIMIENTO DE DASHBOARDS POR VISTA	56
ANEXO F: PROCESO DE BACKWARD ELIMINATION	57
ANEXO G: ESTRUCTURA DEL LIBRO DE DATOS Y LOS INDICADORES QUE LO CONFORMAN EL MISMO	58

ANEXO H: CUADRANTE MÁGICO DE GARTNER DE HERRAMIENTA BI 2021	60
ANEXO I: MODELO PARA EL CÁLCULO DEL NIVEL DE MADUREZ DEL DEPARTAMENTO DE IR PROPUESTO POR LA AIR	60
ANEXO J: ESTRUCTURA DEL LIBRO DE DATOS Y LOS INDICADORES QUE LO CONFORMAN EL MISMO	61

ÍNDICE DE TABLAS

Tabla 1 Funciones comunes de los departamentos de Institutional Research.....	15
Tabla 2 Principales proyectos de los departamentos de Institutional Research	16
Tabla 3 Partes beneficiadas con la implementación IR.....	16
Tabla 4 Niveles de inteligencia organizacional.....	18
Tabla 5 Capacidades o funcionalidades críticas de la plataforma de BI.	25
Tabla 6 Estados de la matriz de confusión.	31
Tabla 7 Métricas	31
Tabla 8 Precio de los productos de Microsoft Power BI	35
Tabla 9 Precio de los productos de Tableau.	36
Tabla 10 Precio de los productos de Qlik Sense.	36
Tabla 11 Resumen de todas las variables	39
Tabla 12 Variables seleccionadas por el backward elimination.....	41
Tabla 13 Métricas algoritmo sin código de balanceo.....	43
Tabla 14 Métricas algoritmo con código de balanceo.....	44

ÍNDICE DE FIGURAS

Figura 1	Proceso de minería de datos, Shmueli (2020).....	28
Figura 2	Países de donde provienen las universidades incluidas en el benchmarking... 33	
Figura 3	Herramientas que ocupan las universidades en la página web de IR.....	34
Figura 4	Estructura inicial del factbook	37
Figura 5	Matriz de confusión algoritmo sin código de balanceo. . ¡Error! Marcador no definido.	
Figura 6	Matriz de confusión algoritmo con código de balanceo.	43

APLICACIÓN DE INSTITUTIONAL RESEARCH PARA APOYAR LA TOMA DE DECISIONES EN LA USFQ

1.Introducción

El aumento de estudiantes en la educación superior implica un reto para el desarrollo de las instituciones en los ámbitos tanto académicos como administrativos (Pizarro, 2017). Estos retos se agudizan para países de la región, es decir aquellos en vías de desarrollo. De acuerdo con el informe de Seguimiento de la Educación en el Mundo 2017/8 de la UNESCO, en Latinoamérica y el Caribe se reportó que desde el año 2000 hasta el 2015 hubo un incremento del 24% en la tasa bruta de matriculación en educación superior. A pesar de presentar avances, la calidad de esta sigue siendo un desafío pendiente, particularmente en los grupos más vulnerables (Pizarro, 2017).

De acuerdo con Saavedra a nivel mundial existe una variedad de opciones sobre la gestión y toma de decisiones para el aseguramiento de la calidad en las instituciones de educación superior (2015). Sin embargo, una de las tendencias que ha tomado importancia en los últimos años, gracias a la globalización son las prácticas del departamento de Institutional Research (IR) (Saavedra, 2015). El autor menciona que la práctica y desarrollo de IR son capaces de proporcionar datos fiables e información válida para los tomadores de decisiones sobre las áreas de interés de la institución. Así como también responder a las demandas de rendición de cuentas e informar sobre las medidas relacionadas para evaluar la eficiencia y efectividad de los procesos y programas institucionales (Howard, 2001, como se citó en Saavedra, 2015). El detalle de las funciones y proyectos del departamento se tratarán a mayor profundidad en la sección de revisión de literatura.

A pesar de que las tareas relacionadas a este departamento se han practicado desde que existen instituciones de educación superior, el término se puso en boga cuando estos

departamentos fueron establecidos en todas las instituciones de Estados Unidos a finales de los años 1950s (Reichard, 2012, como se citó en Angel & Calderon, 2015). Para el año 1966 la primera asociación de IR fue fundada con el interés común de sus miembros en este departamento y el estudio de la operaciones y resultados de la educación post secundaria (Coughlin & Howard, 2011, p.64). Desde entonces hasta la actualidad de acuerdo con Botha, 9 asociaciones han sido establecidas en las diferentes regiones del mundo (2018) (Remítase al Anexo A).

Debido a que existen diferencias significativas en los niveles de madurez del departamento en las regiones, Estados Unidos es considerado como un buen referente, puesto que de acuerdo con Botha, “un indicador del nivel de madurez de las IR en una región es la duración, extensión y sofisticación de la acción colectiva de los investigadores institucionales expresada en la historia y las actividades de las organizaciones profesionales” (2018, p.3). Además, argumenta que el desarrollo del IR en los Estados Unidos, (y en menor medida, Europa y Australasia) ha tenido una influencia significativa en el desarrollo del departamento en otras regiones (2017, p.6).

Latinoamérica, es la única región del mundo sin una asociación u oficinas de IR. A pesar de que existe evidencia de que muchas prácticas del departamento son aplicadas en las instituciones de educación superior, estas se deben principalmente al cumplimiento de mandatos gubernamentales (Saavedra, 2015). Ecuador no es la excepción de este fenómeno, donde la llegada de entidades de acreditación, eventualmente llevaron a la implementación de registro de datos y divulgación de los mismos. (Saavedra, 2015) Desde el año 2010 se tiene registro de que la nueva reforma de educación ha implicado esfuerzos por mejorar los sistemas de manejo de datos, no obstante, todavía hay una tarea por implementar mejores métodos y principalmente otorgar acceso público de estos datos. (Gazzola, 2010, como se citó en Saavedra, 2015) De acuerdo con Botha, ha habido

indicios en Ecuador y Argentina de existir un diálogo para la formación de una asociación formal del IR en Latinoamérica, pero por el momento un ente que garantice prácticas estandarizadas del IR es inexistente (2017).

1.1. Objetivos

1.1.1. Objetivo General

Diseñar un sistema de visualización de métricas clave para la USFQ mediante el cual se pueda rastrear y anticipar los cambios y tendencias de dichas métricas, basado en las directrices y prácticas del departamento de Institutional Research.

1.1.2. Objetivos Específicos

- Desarrollar un benchmarking de los departamentos de Institutional Research de 100 universidades de Estados Unidos.
- Definir cuál es la mejor herramienta de Business Intelligence en para la visualización de las métricas establecidas.
- Establecer la estructura del libro de datos y KPIs que se desarrollarán en este proyecto con un enfoque en las áreas de interés establecidas.
- Crear tableros interactivos donde se muestren los KPIs previamente seleccionados.
- Crear un modelo de analítica de datos, que permita predecir la deserción de un estudiante a partir de información previa al inicio de sus estudios en la universidad.

1.2. Alcance

Para el desarrollo del presente proyecto se establecieron las siguientes áreas de interés dentro de la USFQ en las cuales se aplicarán las prácticas del IR.

- Estudiantes: información de estudiantes registrados y los estudiantes que pasaron por el proceso de admisión. Los tipos de estudiantes a considerar serán regulares y modalidad de estudio presencial de pregrado y posgrado
- Personal: información relacionada con personal académico y universitario

- Investigaciones: información relacionada con publicaciones realizadas por la USFQ

Así como también se definió las prácticas y directrices del departamento de IR a implementar en el proyecto, basado en Delaney (1997), los cuales se tratarán a mayor profundidad en la sección de revisión de literatura.

- Análisis de tendencia: rastrear y anticipar los cambios y las tendencias
- Reporte de datos: generar información estadística precisa sobre la universidad
- Proyectos especiales: realizar estudios y proporcionar un estudio específico

2. Revisión de Literatura

2.1. Institutional Research

Las actividades relacionadas con el Institutional Research (IR) han existido desde los inicios de educación superior. El término solo se puso de moda a partir de los finales de la década de 1950, cuando comenzaron a establecerse oficinas de los departamentos de IR en todas las instituciones Universitarias de EE. UU. (Webber & Calderon, 2015). Según Delaney, los teóricos y profesionales afirman que el IR es un departamento esencial para realizar una toma de decisiones de forma eficaz, dentro de colegios y universidades, llegando a ser una pieza fundamental en el funcionamiento y desarrollo de estas. El autor también menciona que, en la actualidad, la educación superior está atravesando nuevos retos como: el aumento de su competencia directa e indirecta, el incremento de costos de funcionamiento, la creación de nuevas regulaciones del estado, a la tendencia de disminución de las tasas de matrícula y graduación de ciertos segmentos de mercado (1997). Para enfrentar de mejor manera todos estos retos, tener un departamento de IR dentro de las instituciones de educación superior se han vuelto cada vez más esencial y así poder asegurar la sostenibilidad de la organización a mediano y largo plazo (Delaney, 1997).

Para poder entender de mejor manera el impacto que tiene el IR dentro de las organizaciones se debe conocer sus objetivos, funciones y principales proyectos. Según Howard (2001), los principales objetivos del IR dentro de las instituciones superiores son los siguientes:

- Ser un apoyo y apalancamiento en la toma de decisiones universitarias.
- Proveer datos confiables e información valida sobre las diferentes áreas de la institución.
- Reportar acerca del comportamiento de los indicadores para asegurar la eficiencia y efectividad de los procesos y programas institucionales.
- Recopilar, mantener, analizar y difundir datos precisos, consistentes y oportunos.
- Apoyar a los electores públicos en la evaluación y el seguimiento de la calidad de la institución.

De igual manera, Delaney (1997) realizó un estudio enfocado en el departamento del IR de varias instituciones de educación superior, en el cual participaron 127 universidades. Encontró que los departamentos poseen las siguientes funciones en común, las cuales se pueden apreciar en la tabla 1:

Tabla 1: Funciones comunes de los departamentos de Institutional Research

Funciones IR	Descripción de función Institutional Research
Mantenimiento de Datos	Desarrollar, mantener y analizar información institucionalmente relevante en apoyo de la actividad de planificación, evaluación y toma de decisiones de la universidad.
Reporte de Datos	Dar información estadística precisa sobre la universidad a agencias externas. Informar a la comunidad universitaria sobre los resultados de los estudios a través de informes, resúmenes y actualizaciones que se publican periódicamente.
Solicitudes de Datos Externos	Coordinar la respuesta de la Universidad a encuestas estadísticas nacionales, cuestionarios y otras solicitudes de información no rutinarias. Presentación de datos.
Integridad de los Datos	Coordinar, estandarizar y verificar los datos universitarios publicados al público. Revisar las publicaciones de la universidad para verificar la coherencia de la información.
Seguimiento de Objetivos	Hacer un seguimiento del desempeño de la Universidad con respecto a las metas y los objetivos a largo plazo.
Análisis Comparativos	Participar en estudios comparativos de instituciones pares y analizar y presentar resultados

Representación Externa	Participar en consorcios y organizaciones profesionales relevantes en nombre del colegio.
Análisis de Tendencia	Diseñar un sistema mediante el cual se puedan rastrear y anticipar los cambios y las tendencias.
Proyectos Especiales	Realizar estudios y proporcionar la información solicitada por la administración, los departamentos y los comités sobre temas relevantes para la universidad.

Fuente: Delaney A. (1997)

De igual manera, Delaney (1997) en su investigación habla acerca de los principales proyectos que estas instituciones manejan dentro de los departamentos de IR, y las dividió dentro de 8 categorías las cuales se las puede apreciar en la tabla 2:

Tabla 2: Principales proyectos de los departamentos de Institutional Research

Proyectos del IR	Descripción de proyectos de los departamentos de Institutional Research
Informes	Estadísticas institucionales, informes administrativos internos y externos; Investigación
Planificación y análisis de políticas	Estudios de planificación y análisis de políticas, proyecciones de pronóstico / estadísticas, investigación longitudinal e investigación de mercado y encuestas.
Estudios financieros	Análisis de costos, planificación presupuestaria y proyecciones financieras.
Estudios de la facultad	Evaluaciones de la facultad, estudios de carga de trabajo de la facultad y análisis de sueldos.
Encuestas a estudiantes	Encuestas a estudiantes y exalumnos
Estudios académicos	Revisión del programa académico, evaluación del programa académico, evaluación de las pruebas de ubicación y evaluación de los resultados.
Estudios de gestión de matrícula	Estudios de admisión, ayudas económicas y retención.
Otros proyectos	Estudios de utilización del espacio, estudios de transferencia y otros proyectos misceláneos

Fuente: Delaney A. (1997)

Dentro del estudio se encuentra también que las partes beneficiadas con la implementación IR serían las que se aprecian en la tabla 3:

Tabla 3: Partes beneficiadas con la implementación IR.

Partes Beneficiadas	Partes beneficiadas con la implementación IR
Docentes, personal administrativo y estudiantes:	Toma de decisiones, planificación, mejores presentaciones, informes y propuestas
Futuros estudiantes y padres	Comprensión de propiedades y características. Proporcionar una forma de medir la calidad de educación proporcionada.
Instituciones Externas	Comparar desempeño entre instituciones.

Público en General	Comprender la naturaleza y la calidad de la universidad.
--------------------	--

Fuente: Delaney A. (1997)

Como se mencionó anteriormente, el papel principal del IR ha cambiado con el tiempo desde enfatizar y requerir principalmente estadísticas descriptivas, factbook (libros de datos) e informes, a más análisis y evaluación, tanto cuantitativos como cualitativos (Volkwein 2008: 8, como se citó en Nel, 2016). Por lo tanto, los primeros elementos mencionados, se los puede considerar como los básicos para la construcción de un departamento de IR. De acuerdo con Kallina, en una entrevista con la Asociación de Institutional Research (AIR) sobre el factbook institucional para la mayor parte de la comunidad universitaria, este es el medio principal de interacción con la oficina de IR (2018).

Cada universidad define al factbook de diferente manera, por ejemplo, para la universidad de Tufts (posición 30 en el ranking de mejores universidades nacionales de EE. UU. 2021) es una compilación de información básica que incluye historia, programas, estudiantes, personal e instalaciones. Para la universidad de Swarthmore (tercera en la edición 2021 de las mejores universidades nacionales de artes liberales en EE. UU.) el factbook contiene información histórica completa y comúnmente solicitada sobre Swarthmore, y sirve como referencia para el público interno y externo. Por lo que se concluye que la definición de lo qué es el factbook y la información institucional que se presenta en el mismo será dependiente de la universidad y la funcionalidad que se defina darle.

La definición del IR ha ido cambiando a lo largo del tiempo. Sin embargo, de acuerdo con Volkwein, en la actualidad existen dos definiciones que son las más aceptadas (2018). La primera señala que el IR es un departamento de apoyo a las decisiones y lo define como: “un conjunto de actividades que brindan apoyo para la planificación institucional, la formación de políticas y la toma de decisiones” (Saupe,

1990, como se citó en Volkwein, 2008). La segunda definición indica la descripción del IR como inteligencia organizacional (Fincher, 1978 como se citó en Volkwein, 2008).

Profundizando más desde la definición del IR como inteligencia organizacional, Terenzini (1993) desarrollo más a fondo este concepto e identificó 3 niveles de inteligencia organizacional. Los cuales son:

Tabla 4: Niveles de inteligencia organizacional

Niveles	Niveles de inteligencia organizacional
Nivel 1: Técnico y Analítico (Technical and analytical)	-Nivel de inteligencia requerida para producir las cifras o números que describen el perfil básico de una institución: admisiones, matrículas, títulos otorgados, carga de trabajo del profesorado y finanzas. -Implica tener habilidades en el manejo de hojas de cálculo, conocimiento de estadísticas, paquetes de software como SPSS (programas de estadística) y habilidades de investigación de encuestas, entre otros. -Nivel en el que se construye el resto de los niveles.
Nivel 2: Inteligencia de Problemas (Issues Intelligence)	-Incluye conocimiento no solo sobre los aspectos técnicos del trabajo, sino también sobre los problemas particulares que enfrenta la institución: diversidad de estudiantes y profesores, prioridades de asignación de recursos, necesidad de evaluación y mejora del programa, establecimiento de metas de inscripción y construcción de capital. - Requiere conocer a las personas clave de la institución y trabajar con ellas para abordar estos problemas.
Nivel 3: Inteligencia Contextual (Contextual Intelligence)	-Implica conocer la institución no solo internamente, sino también externamente: su historia, su cultura, su evolución y su entorno externo. -La suma total de la inteligencia contextual comprende todas las tendencias relevantes del entorno externo: financieras, sociales, políticas y demográficas.

Fuente: Elaboración propia basado en Volkwein (2008)

Al investigar no existe una forma estandarizada de cómo implementar el IR. Esto se debe a que cada universidad genera su estructura de acuerdo a sus necesidades y consideraciones. Por lo cual se debe ocupar una herramienta que nos permita realizar una exploración y análisis de lo que actualmente se tiene en el mercado. Una de las técnicas más efectivas para hacerlo es un benchmarking. Scott define al benchmarking como un proceso de investigar y comparar productos, servicios, procesos o resultados con otras organizaciones. Esto ayuda a mejorar los resultados o entender más lo que sucede en el mercado permitiendo identificar, adaptar e implementar enfoques hacia mejores prácticas. (2015).

Por otro lado, Scott (2011) señala que dentro de la educación superior para desarrollar con éxito un benchmarking tenemos la siguiente lista de factores a tomar en cuenta:

1. Determinar qué áreas o departamentos comparar;
2. Identificar a que universidades se realizara el benchmarking;
3. Determinar el tipo de benchmarking a utilizar;
4. Preparar documentos y plantillas del benchmarking que incluyan el propósito y el alcance de la investigación;
5. Diseñar el proceso de benchmarking;
6. Realizar el proceso de benchmarking previamente establecido;
7. Revisar los resultados;
8. Comunicar resultados y recomendaciones;
9. Implementar estrategias de mejora.

Por otro lado, las instituciones de educación superior utilizan diferentes métricas por lo que es importante que estas sean evaluadas para ver si se están cumpliendo los objetivos. Suryadi ha identificado que, para evaluar el desempeño de las instituciones de educación superior, existen tres aspectos o criterios principales, los cuales son el académico, investigación y actividades de apoyo. (2017, p.1690) (Remítase al Anexo B) Siendo estos el criterio para la identificación de métricas dentro de las universidades. De igual manera es esencial tomar en cuenta los indicadores que se requieren por ley en el Ecuador. Según el Consejo de Aseguramiento de la calidad de la Educación Superior (2019) en su informe sobre Modelo de Evaluación Externa de Universidades y Escuela Politécnicas señala que el 14 de junio del 2019, el Pleno del Consejo de Aseguramiento de la Calidad de la Educación Superior aprobó el Modelo de Evaluación Externa de Universidades y Escuelas Politécnicas, el cual busca el aseguramiento de la calidad de la educación superior en el Ecuador.

Este sistema busca la autoevaluación permanente de las instituciones de educación superior realizan sobre el cumplimiento de sus propósitos. Por ende, se creó el modelo

2019 con el propósito de evaluar con fines de acreditación a las universidades y escuelas politécnicas. El cual evalúa 20 estándares en base a las tres funciones sustantivas de la universidad ecuatoriana: docencia, investigación e innovación y vinculación con la sociedad. Donde cada eje puede ser evaluado en función de tres dimensiones; planificación, ejecución y resultados. De igual manera el modelo de evaluación posee indicadores cuantitativos y cualitativos con sus respectivas formas de valoración. Las cuales con su calificación otorgará la nota de acreditación de la institución, de tal manera de que, se debe monitorear estos indicadores de forma constante y ver la manera de incrementar su calificación.

2.2. Analítica de Datos

Otra de las funciones del departamento de IR es el estudio de las tendencias de variables para explicar su comportamiento y factores, el cual busca rastrear y anticipar los cambios de las tendencias en métricas clave de la institución (Delayney, 1997). Por esta razón también se decidió crear un modelo de analítica de datos para evaluar qué factores influyen sobre la deserción de los estudiantes en base a la información previa al ingreso de la universidad. Estudios recientes han concluido que la deserción de los estudiantes universitarios generalmente se incrementa durante los primeros años de la carrera del individuo, por lo que analizarlo previamente a su ingreso a la universidad, permitiría establecer medidas preventivas y no esperar a que falle en el primer año, puesto que estos periodos iniciales son cruciales para un estudiante (Sandoval, 2020).

Según Jo, de forma tradicional las empresas han utilizado técnicas estadísticas para extraer información de los datos. Sin embargo, para realizar un análisis estadístico eficaz requiere mucho tiempo y una formación matemática que pocos administradores de bases de datos tienen (1995). Mientras que las técnicas de Machine learning se han desarrollado como respuesta a una urgente necesidad de automatizar los procesos de

análisis y descubrimiento de información por parte de las empresas (Jo, 1995). Por lo general, los algoritmos de aprendizaje automático del machine learning buscan automatizar (al menos parcialmente, si no totalmente) la generación de hipótesis, así como su prueba, permitiendo realizar un análisis de los datos de una forma mucho más rápida y efectiva (Jo, 1995). De igual manera según Aldowah et al, el machine learning está ganando cada vez más importancia en el sector de la educación, pues ayuda a analizar la tendencia de los datos en un menor tiempo y de manera automatizada.

En machine learning existen dos lenguajes de aprendizaje. El lenguaje supervisado y el no supervisado. El lenguaje supervisado a su vez se divide en clasificación cuando se trata de una variable categórica y predicción cuando se trata de una variable numérica (Shmueli, 2021). En el Anexo C, se puede visualizar la organización de métodos de minería de datos de acuerdo con la naturaleza de los datos. El problema de terminar o no la carrera entra en el grupo de una variable categórica, por lo que para predecir las causas que con llevaría a desertar a los alumnos se elaborara el modelo de regresión logística. Sandoval (2020) señala que este modelo es un método tradicional usualmente usado en el campo de la educación y particularmente cuando las variables predictoras son continuas. El modelo se basa en calcular la probabilidad de que una variable categórica tome cierto valor dentro de un conjunto de datos dados por las variables predictoras.

Para la selección de variables se analizaron 3 estudios acerca de la deserción de los estudiantes en los primeros años, uno de ellos aplicados al contexto ecuatoriano. De acuerdo con Alban & Mauricio, tras haber analizado 67 artículos sobre la deserción en educación superior, se menciona que “la edad, género, etnia y desempeño en los exámenes de admisión son los factores más comúnmente usados” (2019). Por otro lado, “la información personal (edad, sexo, dirección, lugar de nacimiento, antecedentes de

inmigración), estudios (curso de estudio, tipo de inscripción), y formación previa (tipo y lugar de titulación de acceso a la universidad, previa experiencia académica), son los factores utilizados para su modelo de sistema de detección temprano (Berens et al., 2018). Finalmente, y dentro un contexto ecuatoriano, en [20] utiliza cuatro variables predictoras: régimen, nivel de curso a la que entra en la carrera, nota de aplicación e índice de vulnerabilidad (analiza la situación socioeconómica del estudiante), donde la nota de aplicación y el régimen fueron las que aportaron con información útil para el modelo (Sandoval, 2020).

3. Metodología

El presente proyecto se desarrolla en dos fases. En la primera se realizó un benchmarking del departamento de IR en diferentes universidades con el fin de diseñar un sistema de visualización de métricas clave para la USFQ, para así poder rastrear y anticipar cambios y tendencias del mercado. En la segunda fase se realizó un modelo de analítica de datos. En el cual se utilizó la metodología propuesta de Shmueli, para evaluar los principales factores que causan la deserción de un estudiante de primer año partiendo de información previa a su ingreso a la universidad.

3.1. Benchmarking Institutional Research

Se inició la investigación con un benchmarking, con el fin de entender a mayor profundidad las aplicaciones y las funcionalidades del IR en otras universidades. Esto ya que se encontró información insuficiente al respecto en la revisión de literatura. Por lo cual se utilizó los primeros siete pasos de la metodología propuesta por Scott (2011) para lograrlo:

1. Determinar el área o departamento al cual se realizaría la investigación, en este caso se seleccionó el departamento de IR.

2. Identificar las universidades a las que se les realizaría el benchmarking. Para el estudio se escogió a las 100 primeras universidades de Estados Unidos pertenecientes al ranking QS 2021. Tomando en consideración a las universidades pertenecientes al Ivy League, ya que son las más prestigiosas del mundo, las cuales son las siguientes: Brown, Columbia, Cornell, Dartmouth, Harvard, Pensilvania, Yale, Princeton. De igual manera se incluyó cuatro universidades de artes liberales en el estudio similares a la USFQ: Hope, Parkside, Holy Cross, Westleyan.
3. Determinar el tipo del benchmarking a utilizar. Para lo cual se escogió un benchmarking comparativo, ya que se realizaron investigaciones y comparaciones con competidores dentro de la misma industria, en este caso es la de la educación superior.
4. Preparar los documentos y plantillas del benchmarking a realizar. En este caso se creó una plantilla de Excel para registrar la información del IR de cada universidad. Los campos incluidos son: nombre, tipo de herramienta y comentarios adicionales del IR.
5. Diseñar el proceso de benchmarking en el cual se definió que se investigaría universidad por universidad y se llenaría la información previamente detallada en la plantilla.
6. Ejecutar el benchmarking previamente establecido.
7. Revisar y analizar los resultados obtenidos, y realizar una investigación más exhaustiva más relevantes.

3.1.1. Herramientas de Business Intelligence

Para la selección de las herramientas de Business Intelligence (BI), se realizaron dos análisis. El primero fue durante el proceso de benchmarking, donde se registró el tipo de herramienta que cada una de las 100 universidades de Estados Unidos empleaba dentro

de sus páginas web de IR para la visualización de sus indicadores. El segundo análisis consistió en la evaluación de las herramientas por medio de reportes de expertos acerca de estas herramientas de BI. Según Amara hasta la fecha Gartner, Forrester y Fuld & Company se desempeñan en las evaluaciones para el software de BI (2008). De los cuales se tuvo acceso al informe de Gartner del 2021 titulado “Magic Quadrant for Analytics and Business Intelligence Platforms” donde se comparan a los diferentes proveedores dentro del mercado de las herramientas de BI.

En dicho reporte se muestra el cuadrante mágico donde se comparan a los proveedores según los criterios y la metodología estándar de Gartner (2021). Esta evaluación consiste en clasificar a los proveedores de BI en 4 tipos, los cuales son líderes, visionarios, jugadores de nicho y retadores (Gartner, 2021). Gartner define a los líderes como aquellos que se desempeñan bien actualmente y están bien posicionados para el futuro. Los visionarios son quienes entienden hacia donde se dirige el mercado o podrían tener una visión para cambiar las reglas de este, pero por el momento no lo ejecutan correctamente. Los jugadores de nicho o bien enfocan su atención en un segmento pequeño o están desenfocados y no presentan innovaciones ni superan a los demás. Finalmente, los retadores se desempeñan bien en el momento actual o dominan un gran segmento, pero no demuestran una comprensión de hacia dónde se dirige el mercado (2021).

En este análisis se incluyen a aquellos proveedores que presentan al menos 8 de las 12 características de la plataforma consideradas en el análisis que se muestran en la tabla 5, también si tiene una presencia razonable en el mercado, es decir que tienen ingresos anuales superiores a \$20 millones y que demuestran que sus soluciones se utilizan y respaldan en todos los aspectos de la empresa y va más allá de las implementaciones departamentales (Amara, 2008).

Tabla 5: Capacidades o funcionalidades críticas de la plataforma de BI.

Capacidades críticas de la plataforma	Definición
Seguridad	Capacidades que habilitan la seguridad de la plataforma, la administración de usuarios, la auditoría del acceso a la plataforma y la autenticación.
Capacidad de administración	Capacidades que rastrean el uso de la plataforma de BI y administran cómo se comparte la información y por quien.
Análisis en la nube	La capacidad de respaldar la creación, implementación y gestión de análisis en la nube, en función de los datos almacenados tanto en la nube como en las instalaciones
Conectividad de la fuente de datos	Capacidades que permiten a los usuarios conectarse, consultar e ingerir datos, mientras optimizan el rendimiento.
Preparación de datos	Soporte para arrastrar y soltar, combinación de datos impulsada por el usuario de diferentes fuentes, y la creación de modelos analíticos (como medidas, conjuntos, grupos y jerarquías definidos por el usuario).
Catálogo	la capacidad de generar y seleccionar automáticamente un catálogo de contenido analítico en el que se pueden realizar búsquedas, lo que facilita que los consumidores analíticos sepan qué contenido está disponible.
Insights automatizados	Un atributo central de la analítica es la aplicación de técnicas de ML para generar hallazgos automáticamente para los usuarios finales (por ejemplo, identificando los atributos más importantes en un conjunto de datos).
Visualización de datos	Soporte para cuadros de mando altamente interactivos y exploración de datos mediante la manipulación de imágenes de gráficos.
Narración de datos	La capacidad de combinar la visualización interactiva de datos con técnicas narrativas para empaquetar y entregar contenido analítico en una forma convincente y fácil de entender para su presentación a los tomadores de decisiones.
Consulta de lenguaje natural	Esto permite a los usuarios hacer preguntas y consultar datos y contenido analítico utilizando términos que se escriben en un cuadro de búsqueda o se hablan.
Generación de lenguaje natural	La creación automática de descripciones lingüísticamente ricas de respuestas, datos y contenido analítico. Dentro del contexto de análisis, a medida que el usuario interactúa con los datos, la narrativa cambia dinámicamente para explicar los hallazgos clave o el significado de los gráficos o tableros.
Informes	La capacidad de crear y distribuir informes de varias páginas con diseño de cuadrícula y píxeles perfectos para los usuarios de forma programada.

Fuente: Gartner (2021)

Mediante la información expuesta se realizó la comparación de las diferentes herramientas considerando los resultados principales del cuadrante mágico de Gartner en base a las capacidades mencionadas previamente, además de los costos y la facilidad de uso. Esto con el fin de seleccionar la mejor herramienta y realizar la implementación de la visualización de KPIs relevantes en la plataforma y para facilitar su uso en el departamento de mejora continua de la USFQ.

3.1.2. El Factbook

Para la construcción del factbook, en español el libro de datos, de igual manera, mediante el proceso de benchmarking, se recolectó y resumió la estructura de los diferentes factbooks mostrados en las páginas web de los departamentos de IR, así como los datos y figuras que se guardan en ellos, pues el factbook es donde se recolectan cifras y figuras de la institución y uno de los elementos básicos dentro de un departamento de IR desde sus inicios, como se discutió en la sección de revisión de literatura. Este análisis y recolección de datos fue restringido a las ocho universidades perteneciente al Ivy League y las cuatro universidades de Estados Unidos que presentaban características similares a la USFQ como se mencionó previamente en la sección 3.1. La información recolectada estuvo enfocada a recolectar toda aquella información sobre la estructura en general del factbook y la información pública que contenían estos elementos.

3.1.3. Identificación de indicadores

Para realizar la identificación de los indicadores que se implementaron en el factbook de la USFQ, se enlistaron de los principales que se encontraron dentro del benchmarking de los departamentos de IR. Se buscaba que estos se adapten al contexto de la USFQ y que formen parte de las tres categorías propuestas: estudiantes, profesores e investigación.

3.1.4. Identificación de indicadores

Para crear los dashboards se utilizó la metodología propuesta por Chantilly Jaggernouth en su charla de “Secretos de diseño para un no diseñador”, en la conferencia anual de Tableau del 2019. En la cual propone que para poder crear dashboards de forma efectiva se deben seguir los siguientes 4 pasos:

- 1) **Reunir los requisitos:** Primero se debe tener claro el objetivo y que es lo que se quiere tener dentro de cada dashboard. Para lo cual se propone el uso de dos

plantillas de requerimientos de dashboards que serán de utilidad para definir el contenido como: audiencia, modo de visualización, uso, notas y filtros de los dashboards, base de datos, filtros, Cálculos entre otros las cuales se las pueden ver el Anexo D y Anexo E (Jaggernouth, 2019).

- 2) **Crear un bosquejo:** Luego se debe crear un bosquejo que tenga todos los requerimientos previamente establecidos. Se debe priorizar a los elementos de mayor importancia y se los debe crear en una con las dimensiones establecidas a la vista escogida. Para lo cual se debe tomar en cuenta los colores de la organización (Jaggernouth, 2019).
- 3) **Creación de plantilla de estructura de construcción de dashboard:** Se debe crear la plantilla en un software de diseño gráfico y las vistas previamente establecidas para el dashboard.
- 4) **Llenar la plantilla de estructura de construcción de dashboard:** Se colocan todas las vistas en la plantilla de estructura de construcción de dashboard en base a la información del bosquejo colocando todos los elementos en las ubicaciones establecidas (Jaggernouth, 2019).
- 5) **Reutilizar el formato de las plantillas:** Se repite el proceso para los siguientes dashboards, manteniendo el formato, tamaño y tipografía empleada (Jaggernouth,2019).

3.2 Analítica de Datos

Dentro de analítica de datos, o minería de datos una de las metodologías más utilizadas es la de Shmueli. En la figura 1 se puede observar que esta metodología está compuesta por 10 pasos que se siguieron para la construcción del modelo con el objetivo de predecir si el estudiante desertará o no en base a la información de ingreso a la universidad.

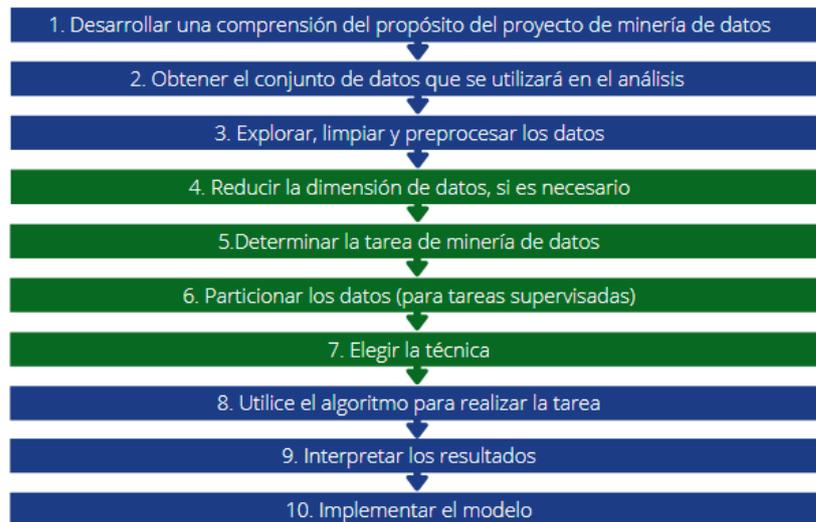


Figura 1: Proceso de minería de datos, Shmueli (2020).

Establecido el objetivo y comprensión del propósito del proyecto de minería de datos, se deberá obtener el conjunto de datos referente a los estudiantes a analizar, es decir regulares, modalidad de estudio presencial y de pregrado. Después se deberá limpiar el conjunto de datos, identificando datos atípicos o valores nulos. Con un conjunto de datos limpio y sin valores nulos, se reducirá la dimensión del conjunto de datos al identificar las variables predictoras. Para esto se recurrirá a la metodología de backward elimination explicado en la sección 3.2.1.

Después, se establece la tarea de minería de datos, la misma que será de clasificación. Antes de la implementación del modelo se requiere particionar los datos en 3 subconjuntos de datos: entrenamiento, validación y prueba. De acuerdo con Shmueli el conjunto de entrenamiento tiene la función de entrenar al modelo, después con el de validación se analiza su poder de predicción y su función principal es identificar el mejor modelo, en el caso de emplear más de un algoritmo (2020). Por último, el set de prueba sirve para analizar el comportamiento del modelo frente a dato nuevos. Esta partición se explicará en la sección 3.2.2. Finalmente, se utiliza el algoritmo para ejecutar la tarea de clasificación, se interpretará los resultados en base a las métricas establecidas en la sección 3.2.3, y por último se deberá implementar el modelo, que no es parte del alcance

de este proyecto, pues esto se deberá realizar con los estudiantes nuevos del siguiente periodo.

3.2.1. Backward Elimination

En el estudio de Karnan y Kalyani indican que cuando se desarrolla un modelo de aprendizaje automático, es de suma importancia seleccionar las variables que aporten y tengan una influencia significativa en el modelo (2010). Caso contrario el modelo puede presentar resultados erróneos a causa de un sobreajuste. Un método de utilidad para encontrar las variables que aportan al modelo es el backward elimination. Este proceso se lo puede apreciar de mejor manera en Anexo F. Este algoritmo ayuda a comprobar el rendimiento del modelo eliminando de forma iterativa las características con peor rendimiento de forma individual hasta que el rendimiento general del modelo tenga un rango aceptable. La métrica de rendimiento que se utiliza en el modelo para poder evaluar el rendimiento de las funciones es el valor p dentro de un modelo de mínimos cuadrados ordinarios. Para poder ingresar las variables, primero se debe convertir las variables categóricas a variables ficticias numéricas para ingresar al algoritmo. Una variable ficticia es una variable que toma valores de 0 y 1, donde los valores indican la presencia o ausencia de su valor asignado. Este algoritmo identificará las variables predictoras significativas en el modelo.

3.2.2. Partición de los datos

La partición de datos fue realizada en entrenamiento, validación y prueba en una proporción de 60%, 20% y 20% respectivamente, de la misma manera que se realizó en el modelo de Mduma (2019) sobre la aplicación de ML para la reducción la tasa de deserción. Además, Gholamy menciona que los mejores resultados, cuando se tiene un conjunto de datos mayor a 1000 observaciones son mejores cuando se mantiene una relación de 80%:20% para el conjunto de datos de entrenamiento más validación y de

prueba. (s/f) El modelo debe ser entrenado utilizando el 60% del conjunto de datos, es decir datos de entrenamiento. Después, el conjunto de datos de validación (20%) se utiliza para la evaluación del desempeño del modelo y ver los resultados del modelo. Finalmente, y para una validación del comportamiento del modelo frente a nuevos datos (esta fase es opcional, debido a que se empleará un solo algoritmo), se combinan los conjuntos de datos de entrenamiento y validación (80%) para su entrenamiento y se evalúa utilizando el 20% de datos restante que el modelo no ha visto aún, esto para analizar el desempeño de este frente a nuevos datos (Mduma, 2019).

3.2.3. Métricas de evaluación

Para poder evaluar el desempeño es decir el poder predictivo del algoritmo se utilizaron varias métricas para poder medir la efectividad de la predicción del modelo. Según Shmueli (2020) las mejores métricas para evaluar algoritmos supervisados son las siguientes:

Matriz de confusión o matriz del error: Esta matriz se utiliza para describir el desempeño de un modelo de clasificación dentro de un conjunto de datos que se conocen los valores verdaderos. Permite tener información acerca de los tipos de errores que está teniendo el algoritmo. Es una tabla con 4 combinaciones diferentes de valores predichos y reales. Para el algoritmo los valores posibles que puede tomar la variable son los siguientes: el algoritmo hace una predicción de que si el estudiante se queda o retirará de la universidad evaluando información previa a su ingreso, donde la variable respuesta puede tomar los siguientes valores:

El estudiante retirado (0)

El estudiante continua (1)

Los resultados del algoritmo de cada estudiante pueden coincidir o no con su estado real para lo cual se puede tener cuatro estados como se muestra en la tabla 6:

Tabla 6: Estados de la matriz de confusión.

Estados Matriz de confusión	Descripción de Estados Matriz de confusión
Verdadero Positivo (TP)	El estudiante que continua en la universidad (1) fue identificado correctamente como estudiante que continua en la universidad.
Falso positivo (FP)	El estudiante que no continua en la universidad (0) fue identificado incorrectamente como estudiante que continua en la universidad.
Verdadero negativo (TN)	El estudiante retirado (0) fue identificado correctamente como estudiante retirado de la universidad.
Falso negativo (FN)	El estudiante no retirado de la universidad (1) fue identificado incorrectamente como estudiante retirado de la universidad.

Fuente: Elaboración Propia basado en Shmueli (2020)

De igual manera partiendo de la matriz de confusión se generan las siguientes métricas de la tabla 7:

Tabla 7: Métricas

Métrica	Descripción	Formula
Accuracy	Porcentaje del total de elementos clasificados de forma correcta. Asume que la probabilidad de que ocurran el error tipo I & II es el mismo	$\text{Accuracy} = \frac{(TP + TN)}{(N + P)}$
Métricas de evaluación de variable respuesta positiva: El estudiante continua (1)		
Sensitivity (1)	Elementos correctamente identificados como positivo de un total de verdaderos positivos. Indica si se ha reconocido una clase correctamente. Se lo usa principalmente cuando se tiene un costo alto de un falso positivo.	$\text{Sensitivity } 1 = \frac{(TP)}{(TP + FN)}$
Precision(1)	Número de elementos correctamente identificados como positivos del total elementos identificados como positivos.	$\text{Precision } 1 = \frac{(TP)}{(TP + FP)}$
F1 Score(1)	Es una media que evalúa la precisión y sensibilidad. Es una mejor medida para usar si necesitamos buscar un equilibrio entre Precision y Recall Y hay una distribución de clases desigual	$F1 = \frac{2 * \text{Precision}1 * \text{Sensitivity}1}{(\text{Precision}1 + \text{Sensitivity}1)}$
Métricas de evaluación de variable respuesta positiva: El estudiante se retiró (0)		
Sensitivity (0)	Elementos correctamente identificados como negativo de un total de verdaderos negativos. Indica si se ha reconocido una clase correctamente. Se lo usa principalmente cuando se tiene un costo alto de un falso negativo.	$\text{Sensitivity } 0 = \frac{(TN)}{(TN + FP)}$
Precision (0)	Número de elementos correctamente identificados como negativos del total elementos identificados como positivos.	$\text{Precision } 0 = \frac{(TN)}{(TN + FN)}$
F1 Score (0)	Es una media que evalúa la precisión y sensibilidad. Es una mejor medida para usar si necesitamos buscar un equilibrio entre Precision y Recall Y hay una distribución de clases desigual	$F1 = \frac{2 * \text{Precision}0 * \text{Sensitivity}0}{(\text{Precision}0 + \text{Sensitivity}0)}$

Fuente: Shmueli (2020)

4. Resultados

Dado que el trabajo investigativo se dividió en dos fases los resultados se muestran de la misma manera. En la primera fase se presentan los resultados encontrados referente al análisis al departamento de IR; mientras que la segunda parte se exponen los resultados encontrados del modelo de analítica de datos.

4.1. Intitutional Reseach

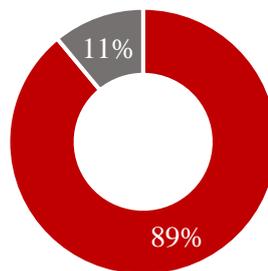
En base a la literatura analizada sobre los elementos y características básicas de un departamento de IR, se estableció que la USFQ debe empezar en el nivel 1 que es la técnica y la analítica, el cual implica principalmente el correcto manejo de datos y la visualización de estadística de estos y el desarrollo de sistemas y estructuras para la generación de datos (De Lisle, 2014). Este tipo de inteligencia es aquel que se requiere para producir los hechos y cifras que describen el perfil básico de una institución como admisiones, matrículas, títulos otorgados, carga de trabajo del profesorado y finanzas (Volkwein, 2008). El autor también menciona que “La inteligencia técnica y analítica también incluye las herramientas de nivel de entrada que necesita, en forma de hojas de cálculo, conocimiento de estadísticas, paquetes de software como SPSS y habilidades de investigación de encuestas, entre otros” (2008).

Adicionalmente, y en base a los resultados del benchmarking del IR de las 100 universidades, también se identificaron y estructuró estos elementos mínimos que permitirán posicionar al departamento de IR como nivel 1. También, y para analizar el cambio del nivel de madurez del departamento se sugiere emplear el modelo de evaluación presentado en el Anexo I.

4.1.1. Benchmarking del IR de las 100 Universidades

Tras analizar las 100 universidades de Norteamérica, de manera general se encontraron los siguientes hallazgos:

Países de donde provienen las Universidades Analizadas



■ Estados Unidos ■ Canadá

Figura 2: Países de donde provienen las universidades incluidas en el benchmarking.

- Se comprobó que todos los departamentos de IR manejan factbooks, ya sean interactivos o no.
- Las universidades deciden qué información se hace pública y cual no. Es decir, el hecho de no encontrar información o indicadores en una cierta universidad no significa que no la manejen, puede simplemente significar que no la muestren.
- Los indicadores dentro de las áreas de estudiantes y profesores que se muestran en los factbooks de las páginas web del IR muestran bastante información y similar entre universidades, sin considerar aspectos financieros.
- El tipo de universidad también influye sobre la diversidad de indicadores en el área de actividades de soporte. Sin embargo, hay que considerar un posible sesgo en este hallazgo, puesto que pueda que manejen ciertos indicadores, pero simplemente no se los hizo públicos.
- Se adquirió un mayor contexto y familiarización acerca del IR en práctica.
- Se consiguió mejor criterio para escoger una herramienta de BI y como mostrar el factbook interactivo.
- Se identificaron los elementos constantes del departamento de IR, siendo en efecto el factbook y la estructura de este (se analiza en la sección 4.1.3)
- Se confirmó que de manera general que los factbooks cambian en forma, mas no en forma.
- La mayoría de las páginas web de los departamentos de IR muestran sus factbooks de manera interactiva.

Es importante señalar que los departamentos de IR tienen un elemento común el cual es el factbook. El mismo puede ser presentados en diversas formas. Kallina señala que pueden ser presentados de manera interactiva, en formato PDF o inclusive en tablas interactivas en las páginas web de sus departamentos. Anteriormente, se presentaban solo como libro de datos. Esta investigación, decidió presentar el factbook en una forma interactiva (2018).

4.1.2. La herramienta BI

Como parte del benchmarking de las universidades, se recolectó la información sobre qué plataforma de BI cada universidad utiliza en sus páginas web. En el gráfico a continuación se muestran los resultados:

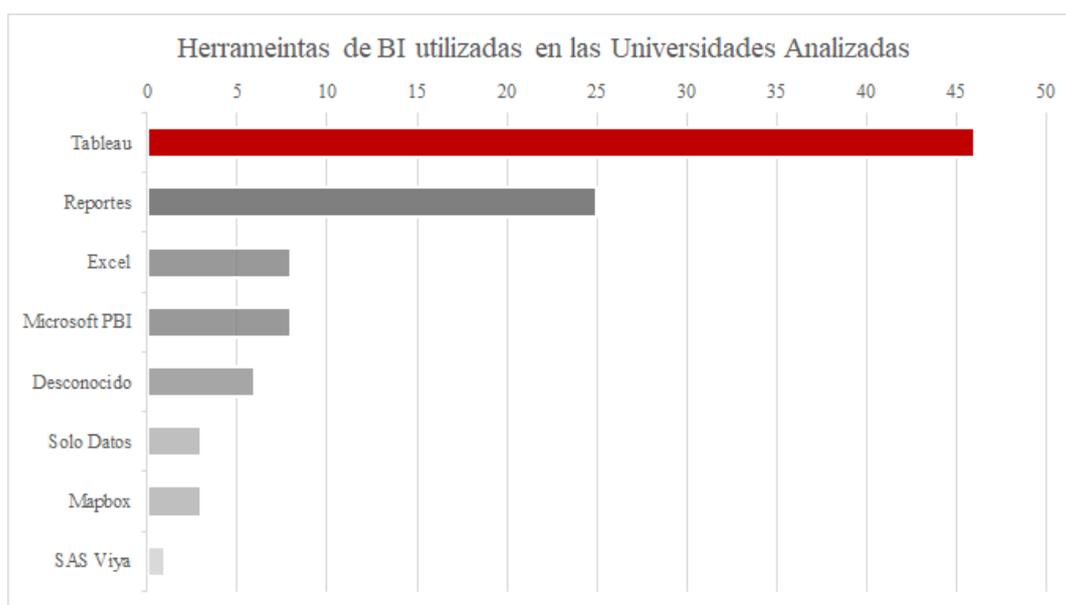


Figura 3: Herramientas que ocupan las universidades en la página web de IR.

La figura 3 muestra que Tableau es la herramienta de BI más utilizada entre las 100 universidades analizadas, el 47% de las instituciones lo emplean en su página web de IR. Seguido de reportes, con el 25%, los mismos que son informes que muestran los datos de las instituciones, más no son interactivos. Se desconoce la herramienta generadora de estos reportes. Con un 8% le sigue los reportes de Excel y la herramienta de Microsoft Power BI, también con un 8%. No se identificó la herramienta que emplean el 8% de las

universidades. Finalmente, el 4% de las universidades restantes ocupan entre Mapbox SAS Viya o se muestran solo datos institucionales.

En cuanto a los resultados referentes al análisis de Gartner, de manera preliminar los resultados de dicho análisis evidenciaron que los tres líderes en el mercado hasta febrero del presente año son Microsoft Power BI, Tableau y Qlik como se puede observar en el Anexo H. Por lo mismo, el análisis se enfocó en estas tres plataformas de análisis e inteligencia empresarial. Mediante la información expuesta se realizó la comparación de las diferentes herramientas considerando los resultados principales del cuadrante mágico de Gartner en base a las capacidades mencionadas en la sección de metodología.

Los resultados referentes al precio se presentan a continuación, cabe resaltar que la información expuesta se extrajo de las páginas web oficiales de los proveedores de las herramientas BI.

- **Microsoft Power BI**

Hay tres planes para Microsoft Power BI, el de escritorio, Pro y Premium

Tabla 8: Precio de los productos de Microsoft Power BI .

Plan	Productos y funcionalidades	Precio
MS Power BI Escritorio	MS Power BI de escritorio. *gratis para usuarios individuales	Gratis 1 usuario
Pro	Incluye MS Power BI en el móvil, publicar y compartir reportes, 1 GB de límite de tamaño del modelo, seguridad y cifrado de datos y el almacenamiento máximo es de 10 GB/usuario y posibilidad de que 8 bases de datos se refresquen cada día automáticamente. Si se tiene la licencia de MS 365 E5 Microsoft Power BI es gratuito.	\$9.99/usuario/mes
Premium por usuario	Incluye todas las funcionalidades de plan Pro, además de un aumento de tamaño de un modelo a 100 GB, 48 bases de datos actualizadas cada día automáticamente y un máximo de 100 TB de almacenamiento total. Incluye informes paginados conocidos como reportes que pueden encajar en una página inclusive si se tienen muchas visualizaciones. Incorpora análisis de inteligencia artificial y visualización de los cambios en un archivo.	\$20/usuario/mes
Premium por capacidad	Incluye todas las funcionalidades de MS Power BI Pro, adicionalmente no se requiere adquirir licencias de usuario para consumidores de contenido, se aumenta el límite máximo del modelo a 400 GB y el almacenamiento máximo se aumenta a 100 TB. Se emplea un servidor por lo que deja de ser dependiente del hardware y la escalabilidad de los recursos es automática.	\$4995/capacidad/mes

Fuente: Elaboración propia basado en Microsoft Power BI (2021)

- **Tableau**

Tableau ofrece tres planes de licencias mostrados en la siguiente tabla:

Tabla 9: Precio de los productos de Tableau.

Plan	Productos y funcionalidades	Precio
Creador	Incluye Tableau Desktop, Tableau Prep Builder. Y una licencia de creador de Tableau Server o Tableau Online.	\$70/ usuario/mes (se factura anualmente)
Explorador	Una licencia de explorador de Tableau Server.	\$35/ usuario /mes (se factura anualmente)
	Una licencia de explorador de Tableau Server si es implementado en la nube es decir que los recursos son administrados por un tercero y no se requiere administrar el hardware.	\$42/usuario/ mes (se factura anualmente)
Espectadores	Una licencia de espectador de Tableau Server. * Este plan requiere de al menos 100 espectadores para que se pueda adquirir este plan.	\$12/usuario/mes (se factura anualmente)
	Una licencia de espectador de Tableau Server sí es implementado en la nube es decir que los recursos son administrados por un tercero y no requiere administrar el hardware. * Este plan requiere de al menos 100 espectadores para que se pueda adquirir este plan.	\$15/usuario/mes (se factura anualmente)

Fuente: Elaboración propia basado en Tableau (2021)

- **Qlik Sense**

Qlik Sense presenta dos ediciones:

Tabla 10: Precio de los productos de Qlik Sense.

Plan	Productos y funcionalidades	Precio
Qlik Sense Business	Incluye analítica moderna y analítica aumentada y colaboración grupal, y funcionalidades específicas de estas características. Los espacios compartidos con controles de usuario están restringido a 5 espacios, 250 GB de almacenamiento en la nube, hasta 50 bases de datos actualizadas cada día y un límite de 1.25 GB de memoria de tamaño de aplicación.	\$30/usuario/ mes (se factura anualmente)
Qlik Sense Enterprise Saas	Este plan incluye todas las funcionalidades de Qlik Sense Business, pero los espacios compartidos son ilimitados, 500 GB de almacenamiento en la nube, al menos 100 bases de datos actualizadas cada día y un límite de 2.5 GB de memoria de tamaño de aplicación. *requiere contactar a ventas para detalle de costos.	\$40/\$70/ usuario/ mes (se factura anualmente)

Fuente: Elaboración propia basado en Qlik y G2

Mediante el análisis y resultados del cuadrante de Gartner, fue posible evaluar a la herramienta desde sus fortalezas como la visualización, la interactividad con los usuarios, facilidad de uso y manejo de datos en el software. Así como también compararla con el resto de los competidores en el mercado de herramientas de BI. Una de las ventajas que brinda Tableau, a pesar de ser más costoso que su competidor principal Microsoft Power BI, y que hace a esta herramienta distinguirse de las demás es la experiencia de

usuario de análisis, otorgándole facilidad de uso al usuario y permitiéndole realizar análisis sin mayor complejidad, de acuerdo con el informe presentado por Gartner (2021).

4.1.3. Factbook

Como parte de los resultados del benchmarking de las 100 universidades analizadas, se logró identificar la estructura que sigue este elemento común de los departamentos de IR: el factbook, así como también la información que contiene este. Como se ha mencionado, el factbook es uno de los elementos básicos de un departamento de IR y esto se comprobó al analizar las páginas web de estos departamentos. Además, mediante este análisis se concluyó que en efecto estos factbooks cambian principalmente en forma, más no en fondo. Esto permitió establecer la estructura y las áreas en las que se dividen estos factbooks. En la tabla a continuación se muestran los hallazgos acerca de la información que debería contener el factbook y por ende la propuesta inicial del mismo:

Institutional Research USFQ	Quick Facts	Estudiantes	Registrados
	Factbook		Personal Académico y Universitario
		Universidad y Otros Datos Institucionales	
			Personal de Apoyo Académico
			Administrativos
		Investigación	
		Programas y Carreras	
		KPIs Exigidos por el Gobierno Ecuatoriano	

Figura 4: Estructura inicial del factbook

Este factbook fue diseñado en formato digital a manera de tableros interactivos construido en la herramienta de BI seleccionada que es Tableau donde la primera sección de este es la de quick facts la cual contiene información rápida e interesante acerca de la universidad. La siguiente sección es la de estudiantes, la cual a su vez se divide en información acerca de los estudiantes registrados y en información de estudiantes involucrados en el proceso de admisión. La segunda sección es acerca del personal, el cual se divide en personal académico, personal de apoyo académico y

personal administrativo. Por último, se incluye la última sección que se trata de la universidad y otros datos institucionales. En esta sección se proponen recopilar la información acerca de la investigación realizada, los programas y carreras que se ofertan y los KPIs exigidos por el gobierno, los cuales se tratarán en la siguiente sección. Los dashboards creados que conformarán el factbook se resumen en la siguiente sección.

4.1.4. Dashboards KPIs

Para crear los dashboards, primero se generó un listado con los posibles indicadores en base al resultado del benchmarking. Con dicha información se planearon once dashboards tentativos divididos en cuatro categorías del factbook: Admisiones, Estudiantes, Personal USFQ, Publicaciones. Se utilizó la metodología de creación de dashboards efectivos de Jaggernouth previamente mencionada. Para lo cual primero se fueron llenando las plantillas de requerimiento de dashboards que se pueden apreciar en Anexo D y Anexo E. Posteriormente se creó un bosquejo dentro de las dimensiones previamente escogidas. En este caso se utilizaron las medidas establecidas por Tableau de un escritorio de 1300 px de largo y 950 px de ancho. De igual manera se tomó en cuenta los colores institucionales. Luego se creó la plantilla de estructura de construcción del dashboard en Canva y las visualizaciones que este va a poseer en Tableau. Finalmente se colocaron todas las vistas previamente definidas previamente en el dashboard de Tableau y se genera una historia en el cual se van colocando los otros dashboards que tengan relación con el tema. Se crearon en total de 11 dashboards los cuales se clasificaron en 2 de admisiones, 4 de registrados, 4 del personal USFQ y 1 de publicaciones. Los cuales se los colocó en un Manual de Dashboards y los Donde el listado de las vistas y dashboards se pueden ver en el Anexo J.

4.2. Analítica de Datos

4.2.1. Descripción de los conjuntos de datos y su preprocesamiento

Para limpieza y preprocesamiento de los datos se obtuvieron las variables con las que finalmente se trabajaron para el desarrollo del presente modelo. Tomando en consideración las variables empleadas en otros modelos de análisis de deserción, mencionados previamente en la revisión de literatura, así como también la restricción al acceso de la información dentro la USFQ, las variables seleccionadas para el desarrollo de este modelo se describen en la tabla 11. Los datos se refieren a toda aquella información previa al ingreso de un estudiante nuevo regular de pregrado modalidad presencial.

Tabla 11: Resumen de todas las variables

No.	Variable	Descripción
1	Genero	Género del Estudiante
2	Colegio	Colegio de la carrera del Estudiante
3	Carrera	Carrera en el que está registrado el Estudiante
4	Provincia	Provincia de residencia del Estudiante
5	Nota SHCH	Nota de graduación del Colegio
6	Nota PAAM	Nota del examen de admisión de matemática
7	Nota PAAR	Nota del examen de admisión de razonamiento lógico
8	Nota PAAV	Nota del examen de admisión de aptitud verbal
9	Nota PAAT	Nota del examen de admisión total
10	Transferencia	Si es un Estudiante transferido o no
11	Periodo	Si inicio en el periodo 10 o en el periodo 20

El conjunto de datos estuvo compuesto por 3490 observaciones y 12 variables, 1 variable de respuesta categórica binaria, 5 numéricas y 6 categóricas. Dentro del mismo se encontraron valores faltantes los mismos que fueron remplazados por la mediana y en otros casos por la moda puesto que estos, de acuerdo con Shmueli, a pesar de existir métodos más sofisticados, como regresión lineal para el remplazo de valores nulos, estos son una correcta elección pues no se ven afectados por los valores atípicos (2021, p.64).

4.2.2. La variable de respuesta

La variable de respuesta o la variable por predecir es si el estudiante desertó o no. Es decir, la variable de respuesta tiene 2 clases de desertor y no desertor. Los cuales se representaron en una variable binaria que indica si el estudiante es un desertor se colocó un valor de 0 y si es un no desertor de 1, donde el 81.52% pertenecen a la clase 1 y el 18.48% restante a la clase 0.

4.2.3. Problema desbalanceado

Al analizar la variable respuesta categórica creada se encontró que los datos se encuentran en los siguientes porcentajes: la variable de estudiantes activos posee un 81.52% mientras que la variable de estudiantes retirados posee un 18.48%. Estos porcentajes nos indican que estamos trabajando con un set de datos desequilibrado o de desbalanceado. Según Guo et al. (2008) Los algoritmos con sets de datos desbalanceados tienden a ser abrumados por la clase mayoritaria e ignorar a la clase minoritaria. Este desequilibrio causa un rendimiento subóptimo de clasificación, ya que la mayoría de los algoritmos se comportan mal cuando los conjuntos de datos están muy desequilibrados. Esto pasa porque no poseen suficiente información para entrenar al algoritmo sobre la variable minoritaria y el algoritmo le da más peso a la mayoritaria.

4.2.4. Backward Elimination

Para evaluar el rendimiento de las funciones es el valor P dentro de un modelo de Mínimos cuadrados ordinarios, el cual se utiliza para realizar regresiones lineales. En el modelo se estableció un alfa de 0.05 ya que este valor es muy común en este algoritmo de backward elimination. Para poder ingresar las variables de la tabla 11. Donde primero se convirtieron a todas las variables categóricas en variables ficticias. Dando como resultado al correr el algoritmo las variables de la tabla 12.

Tabla 12: Variables seleccionadas por el backward elimination.

No.	Variable	Descripción
1	Nota SHCH	Nota de graduación del Colegio
2	Nota PAAM	Nota del examen de admisión de matemática
3	Provincia_Pichincha	Estudiantes residentes en la Provincia de Pichincha
4	Transferencia_Si	Estudiantes Transferidos
5	Genero_M	Estudiante con Género Masculino
6	Carrera_agronomía	Estudiantes que estudiarán la carrera de agronomía
7	Carrera_animación digital	Estudiantes que estudiarán la carrera de Animación digital
8	Carrera_artes musicales ejecución mc	Estudiantes que estudiarán la carrera de artes musicales ejecución MC
9	Carrera_biólogía	Estudiantes que estudiarán la carrera de Biología
10	Carrera_física	Estudiantes que estudiarán la carrera de Física
11	Carrera_hospitalidad y hotelería pr	Estudiantes que estudiarán la carrera de Hospitalidad y hotelería Pr
12	Carrera_matemática	Estudiantes que estudiarán la carrera de Matemática
13	Carrera_medicina veterinaria	Estudiantes que estudiarán la carrera de Medicina Veterinaria

4.2.5. El modelo

Debido a que se trata de lenguaje supervisado y una variable de respuesta binaria, se seleccionó el algoritmo de regresión logística, como se mencionó en la revisión literaria. Además de ser ampliamente utilizado en el ámbito de educación, este algoritmo permitirá identificar las variables que influyen sobre la deserción de un estudiante. Dentro de los parámetros de la función de regresión logística, existe una opción que permite solucionar el problema del desbalanceo de clases, sin necesidad de realizar un método de muestreo. De acuerdo con la documentación específicamente, al configurar (`class_weight = "balanced"`) pesará automáticamente las clases de manera inversamente proporcional a su frecuencia:

$$w_j = \frac{n}{kn_j}$$

Donde w_j es el peso de la clase, j , n es el número de observaciones, n_j es el número de observaciones en la clase y k es el número total de clases. Lo cual ayuda a mejorar la predicción de la variable desbalanceada.

4.2.6. Resultados del modelo

En base a las métricas de evaluación que se presentaron en metodología, se evaluaron los resultados del modelo de la regresión logística. El modelo se evaluó en base a dos escenarios. Primero la ejecución del algoritmo sin el código de balanceo (`class_weight = "balanced"`) y el segundo con el uso de este. En el cual se encontraron dos escenarios diferentes.

En el primer escenario evaluando al algoritmo sin el uso del código de balanceo se encontraron los siguientes resultados:

En primera instancia se creó la matriz de confusión, como se puede ver en la figura 5. En la cual podemos apreciar que nuestro algoritmo calcula de forma efectiva a la variable respuesta de que nos indica si la estudiante continua (1) pues tiene mucha más información. Sin embargo, cuando el algoritmo intenta clasificar al estudiante retirado (0) se presenta un problema ya que no calcula ningún estudiante del total de los ciento trece estudiantes retirados y los está clasificando como estudiantes que continúan (1) esto ya que tiene un alto número de falsos negativos (FN).

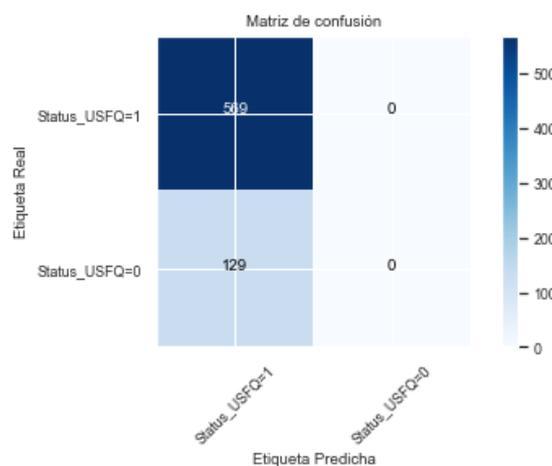


Figura 5: Matriz de confusión algoritmo sin código de balanceo.

De igual manera se puede apreciar en la tabla 13 otras métricas de evaluación las cuales nos ayudan a comprender el comportamiento de la variable respuesta desde diferentes aspectos específicos. Evaluando la sensibilidad podemos observar que no se

está clasificando de manera correcta a la clase estudiante retirado (0), ya que posee un valor de 0.00 lo cual indica que no está prediciendo nada de esta variable respuesta. Sin embargo, la sensibilidad de la clase estudiante continua (1) tiene un valor de 1 lo cual es muy alto lo cual significa que se la está clasificando de manera correcta. Lo cual se puede inferir de igual manera de los resultados de score 1. Evaluando el Accuracy del modelo nos da un valor alto ya que el algoritmo está prediciendo de forma efectiva a la clase, la estudiante continua (1).

Tabla 13: Métricas algoritmo sin código de balanceo.

Variable/Métrica	Sensitivity	Precision	F1 Score	Accuracy
0	0.00	0.00	0.00	0.82
1	1	0.82	0.90	

Mientras que en el segundo escenario con el código de balanceo se encontraron los siguientes resultados:

En primera instancia de igual manera primero se creó la matriz de confusión, como se puede ver en la figura. En este escenario podemos apreciar que se incrementa el Verdaderos negativos (TN) lo cual es lo cual es el objetivo del desarrollo del algoritmo. Sin embargo, se está aumentando el número de forma significativa el número de los falsos positivos (FP). Lo cual significa que se está clasificando a los estudiantes con un peso mayor a la predicción de deserción sin embargo no lo está haciendo de manera correcta.

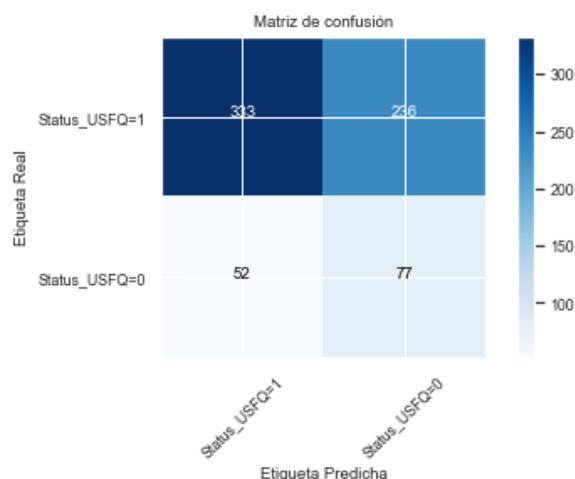


Figura 5: Matriz de confusión algoritmo con código de balanceo.

Con respecto a la tabla 14 se puede apreciar que la sensibilidad de la clase estudiante retirado (0) se incrementa a un valor de 0.60, lo cual presenta una mejora en relación con la comparación anterior de 0.00, lo cual significa que está prediciendo de mejor manera a la clase estudiante retirado (0). Sin embargo, la sensibilidad la clase estudiante continua (1) decrece ya que está calculando un número mayor de predicciones Falsos positivos (FP). Lo cual se puede ver de igual manera en el F1 Score, el cual mejoró de un 0.0 a un 0.35 clase estudiante retirado (0) y un decremento del 0.9 a un 0.7 de la clase estudiante continua (1). Evaluando el Accuracy anterior con la actual se puede apreciar que este disminuyó de un valor de 0.82 a 0.59 lo cual significa que, aunque se incrementó el cálculo de los Verdaderos negativos también se incrementó la cantidad de los Falso positivo (FP).

Tabla 14: Métricas algoritmo con código de balanceo.

Variable/Métrica	Sensitivity	Precision	F1 Score	Accuracy
0	0.60	0.25	0.35	0.59
1	0.59	0.86	0.70	

4.2.7. Coeficientes

En cuanto a los coeficientes, a diferencia de un modelo lineal, los coeficientes de regresión logística no tienen una interpretación directa (Fernandez, 2020). El autor menciona que hay dos formas principales de leer los coeficientes: a) analizar la razón de probabilidades y b) convertir la razón de probabilidades en un porcentaje (2020). Las dos formas de interpretar los coeficientes se muestran en la tabla 15.

Tabla 15: Coeficientes del modelo y porcentajes de la razón de probabilidades de regresión logística

Variable	Coeficiente	Porcentaje
Nota_SHCH	1.712827	5.544613
Nota_PAAM	1.081720	2.949750
Carrera_hospitalidad y hotelería	0.819442	2.269232
Provincia_Pichincha	0.315528	1.370982
Genero_M	-0.193897	0.823743
Transferencia_si	-0.447149	0.639448
Carrera_medicina veterinaria	-0.462957	0.629420
Carrera_animación digital	-0.606632	0. 545184

Carrera_biolología	-0.649729	0.522187
Carrera_matemática	-0.658925	0.517407
Carrera_ artes musicales ejecución mc	-0.751396	0.471708
Carrera_física	-0.769627	0.463186
Carrera_agronomía	-0.895735	0.408308

En el primer caso se concluye por ejemplo que si un estudiante proviene de la provincia de Pichincha o es candidato estudiar la carrera de hospitalidad y hotelería, aumenta las probabilidades de pertenecer a la clase 1, es decir de no desertar, pues los coeficientes son positivos. Si es hombre, donde dicha variable tiene un coeficiente negativo, quiere decir que disminuye la probabilidad de pertenecer a la clase 1 o de no desertar. Es decir, que las variables que presentan coeficientes negativos son aquellas características de un posible desertor. En términos de porcentaje como se puede visualizar en la tabla 15 que si el estudiante es transferido aumenta en tan solo un 0.63% la probabilidad de no desertar cuando el resto de las variables permanecen constantes (Fernandez, 2020).

5. Conclusiones y Recomendaciones

5.1. Institutional Research

Se logró el objetivo de conocer más acerca del departamento de Institutional Research en las universidades de Norteamérica, donde el 89% de ellas eran de Estados Unidos y el 11% restante de Canadá, ambos países pertenecientes a la misma Asociación de Institutional Research (Association of Institutional Research – AIR). De esta investigación se concluye que, si bien existen diferencias de estructura, funciones y roles de estos departamentos dentro de las universidades investigadas, se identificó mediante la literatura y el benchmarking los factores comunes que caracterizan al departamento para asegurar el objetivo principal de brindar una educación de buena calidad. Si bien es cierto que existen diferentes caminos para llegar a dicho objetivo, los elementos identificados y empleados en esta investigación sobre el departamento de IR deberían conducir a alcanzar este objetivo.

- Los elementos seleccionados han sido identificados como aquellos que representan el nivel 1 de un departamento de IR, identificando al nivel 1 como el más básico de desarrollo de este. El elemento principal fue la construcción del factbook interactivo donde se muestran las métricas de las áreas de enfoque que fueron los estudiantes y profesores (incluyendo publicaciones). Así como también el proyecto del modelo de analítica de datos donde se analizaron los factores que influyen sobre una posible deserción de un estudiante. Por lo tanto y mediante la implementación de esta propuesta, se espera que la USFQ pueda presentar los rasgos básicos de un departamento de IR, es decir de nivel 1 como se mencionó en resultados. Actualmente se desconoce el nivel madurez del departamento de IR en la USFQ, pero esta propuesta debería estar dirigida a aumentar dicha madurez. Para esto se recomienda utilizar el modelo de evaluación de AIR (Remitirse a Anexo I) del nivel de madurez del departamento de IR para controlar este indicador a lo largo del tiempo y validar la efectividad de la propuesta implementada. Cabe recalcar que para la construcción del factbook también se tomó en consideración los indicadores propuestos en el Modelo de Evaluación Externa de Universidades y Escuelas Politécnicas del 2019.
- En el cuadrante de Gartner, Tableau se encuentra como el segundo líder después de Microsoft Power BI. Pero, a pesar de esto Tableau, se presenta como la mejor opción frente a los competidores analizados porque, aunque las visualizaciones e imágenes en el mercado actual es un factor bastante mercantilizado, Tableau logra diferenciarse en este aspecto hasta la actualidad. También es una plataforma muy atractiva e inclusive sus clientes la han catalogado como agradable de usar. Así como también la experiencia de usuario de análisis, según Gartner a pesar de estar

constantemente agregando nuevas capacidades siempre mantiene una experiencia elegante permitiendo al usuario realizar análisis sin problema. Además, la herramienta de BI empleada en el 47% de las universidades analizadas permitieron establecer un criterio más para la selección de Tableau como la herramienta de visualización del factbook interactivo, pues la selección de una herramienta es muy dependiente de la empresa o negocio a la que se aplique. Conocer sobre que están utilizando las instituciones de educación superior hoy en día fortaleció los criterios de evaluación para esta decisión.

- En cuanto al ámbito de los indicadores se identificó que no todos son indicadores clave de desempeño de las instituciones. Para esto también se recomienda analizar el contexto de la universidad, considerando a todos los involucrados en la toma de decisiones de la USFQ y a qué objetivo se busca llegar como universidad para así definir qué indicadores permitirán llegar a dichos objetivos.
- La extracción de los datos es un factor importante dentro la construcción del factbook. Por lo tanto, se recomienda que no existan tantas barreras entre el departamento de mejora continua y el departamento que manejan los mismos.

5.2. Analítica de Datos

- En este trabajo, se muestran los resultados preliminares para predecir la deserción de un estudiante de primer año mediante el uso de información previa a su ingreso a la universidad partiendo de un conjunto de datos mediano de registros demográficos y académicos de los estudiantes.
- La predicción de la eventual deserción de estudiantes arrojó resultados prometedores a partir de un conjunto de datos desbalanceados de más de 3.400 registros con uso del algoritmo de regresión logística.

- Al tener una variable respuesta categórica binaria, se corre el riesgo de que los datos se encuentren desbalanceados. Este fue el caso del presente proyecto, pues en la variable de respuesta el 81.52% de los datos se encontraban en la clase mayoritaria, es decir estudiantes que se quedan en la universidad. Mientras que el 18.48% restante estaban contenidos en la clase minoritaria, es decir estudiantes que desertaron. Los sets de datos desbalanceados tienden a ser abrumados por la clase mayoritaria e ignorar a la clase minoritaria para el análisis del modelo.
- Regresión logística es un algoritmo muy bueno para la predicción de variables categóricas binarias, sin embargo, cuando un set de datos es desbalanceado puede llegar a dar mucho más peso a la variable mayoritaria en su predicción y no generar buenos resultados. La contextualización de las variables es de vital importancia para el preprocesamiento de datos, debido a que esto nos da la idea de por dónde empezar y que se relaciona con qué.
- En el primer escenario, es decir sin la implementación de balanceo de clases, se está presentando un problema de una alta concentración de predicciones de falsos negativos (FN), lo cual significa que un estudiante retirado de la universidad (0) fue identificado incorrectamente como estudiante que continua en la universidad, lo cual no es correcto para desarrollar el modelo, ya que nos interesa predecir de mejor manera la deserción de los estudiantes.
- En el segundo escenario, con la implementación del código de balanceo, se pudo apreciar que el cálculo de los verdaderos negativos (TN) se incrementaron lo cual es positivo para el objetivo del modelo, sin embargo, cabe señalar que se incrementó de forma considerable el número de falsos positivos (FP) del modelo, implicando de forma negativa en el desempeño del modelo.

- El análisis de la deserción de los estudiantes mostró que de donde proviene el estudiante, donde estudio, las notas del examen de admisión de ingreso de matemática, las de graduación del colegio y el mes en el que se matricula y que rinde el examen son las principales variables que podrían explicar en una posible deserción.
- Se recomienda realizar la implementación del algoritmo de ensamble Gradient Boosting Trees como una forma de mejorar la predicción del modelo. Así como también analizar las implicaciones de haber considerado a estudiantes inactivos como retirados. También se podría conducir un estudio por colegio para determinar si la carrera influye sobre la deserción de un estudiante, puesto que en el actual modelo el colegio académico no representó ser información útil.

6. Limitaciones

- A pesar de que se tuvo toda la disposición del departamento de Mejora Continua para proporcionar los datos para el estudio. La extracción de algunos de ellos dependía de terceros y complicando así el acceso a algunas bases de datos.
- Cierta información propuesta para el modelo no se está levantado actualmente, o era información clasificada por lo que no pudo ser provista para el estudio.
- A pesar de existe información sobre los objetivos, funciones, áreas de aplicación y proyectos del IR, no existe información acerca de la implementación de este; ya que, la información que utiliza es información sensible de las universidades.
- Dentro del modelo de analítica una limitación es que no se considera como los estudiantes inactivos se comportan a lo largo del tiempo, por lo que se sugiere que uno de los pasos siguientes, para mejorar el poder predictivo del modelo, es analizar este grupo de estudiantes

7. Siguietes pasos

- Evaluar el nivel de madurez del IR que se tiene actualmente en la USFQ.
- Complementar la información del factbook considerando las demás áreas que conforman la USFQ, como finanzas, planta física y otras actividades de soporte.
- Estandarizar el manejo de datos dentro de la universidad para que esos indicadores se vayan actualizando de manera regular y sin necesidad de preprocesamiento.
- Construir una sección en la página web de la USFQ para el departamento de mejora continua donde se muestren los elementos implementados del IR.
- Para la replicabilidad del proyecto presentado, es necesario incluir una estandarización de pasos a seguir para limpieza de datos, puesto que no se trata de una conexión directa con las bases de datos.
- Socializar esta investigación sobre el departamento de IR con universidades de la región, para establecer conversaciones sobre las prácticas que ya se están realizando para la mejora de la calidad de la educación y que probablemente no se lo esté haciendo bajo el nombre de IR. Esto con el fin de encontrar más puntos de información y que posiblemente se convierta en una asociación.

Referencias

- Alban, M., & Mauricio, D. (2019). Predicting University Dropout through Data Mining: A Systematic Literature. *Indian Journal of Science and Technology*, 13.
- Aldowah, H., Al-Samarraie, H., & Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics and Informatics*, 37, 13–14.
<https://doi.org/10.1016/j.tele.2019.01.007>
- Amara, Y. (2008). Business Intelligence Software Evaluation_Testing the SSAV Model.
- Berens, J., Oster, S., Schneider, K., & Burghoff, J. (2018). Early Detection of Students at Risk -.40

- Anguita, D., Ghelardoni, L., Ghio, A., Oneto, L., & Ridella, S. (2012). The 'K' in K-fold Cross Validation. *ESANN 2012*, 1–5.
<http://www.i6doc.com/en/livre/?GCOI=28001100967420>.
- Botha J. (2018) The Impact of Global Forces in Higher Education on the Development of Institutional Research. In: Webber K. (eds) Building Capacity in Institutional Research and Decision Support in Higher Education. Knowledge Studies in Higher Education, vol 4. Springer, Cham. https://doi.org/10.1007/978-3-319-71162-1_2
- Consejo de Aseguramiento de la Calidad de la Educación Superior (CACES) (2019). Modelo de Evaluación Externas de Universidades y Escuelas Politécnicas. Recuperado de:
<file:///C:/Users/Javier/OneDrive/Programacion%20Neurologistica/modelo%20evaluacion%20uep%20caces%202019%20publicado.pdf>
- Coughlin, M. A., & Howard, R. (2011) The Association for Institutional Research: The First 50 Years. 160.
- De Lisle, J. (2014). Institutional research as organisational intelligence: Using evidence to inform continuous quality improvement. *Call for Peer Reviewers*, 31.
- Delaney, A. (1997). The Role of Institutional Research in Higher Education: Enabling Researchers to Meet New Challenges. *Research in Higher Education*, 38(1), <http://www.jstor.org/stable/40196232>
- Fernandes, A. A. T., Figueiredo Filho, D. B., Rocha, E. C. D., & Nascimento, W. D. S. (2020). Read this paper if you want to learn logistic regression. *Revista de Sociologia e Política*, 28(74).
- Scott, R. (2011). Benchmarking: A literature review. *Academic Excellence Centre for Learning and Development, Edith Cowan University*.
- G2. (2021). *Qlik Sense Pricing*. <https://www.g2.com/products/qlik-sense/pricing>
- Gartner. (2021). *Definition of Magic Quadrant - Gartner Information Technology Glossary*. <https://www.gartner.com/en/information-technology/glossary/magic-quadrant#:~:text=Gartner%20Magic%20Quadrants%20offer%20visual,Gartner's%20standard%20criteria%20and%20methodology>.
- Gartner, Inc. (2021, febrero) Gartner Magic Quadrant for Insight Engines Anthony Mullen, Stephen Emmott
- Gholamy, A., Kreinovich, V., & Kosheleva, O. (s/f). *Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation*. 6.
- Gounder, M.S., Iyer, V.V., & Mazyad, A.A. (2016). *A survey on business intelligence tools for university dashboard development. 2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC)*, 1-7.

- Guo, X., Yin, Y., Dong, C., Yang, G., & Zhou, G. (2008). On the Class Imbalance Problem. *Fourth International Conference on Natural Computation*, 192–201. <https://doi.org/10.1109/ICNC.2008.871>.
- Jaggernouth, C. (15 de noviembre de 2019). *Zen Master: Design Secrets for a Non-Designer* [Sesión de conferencia]. Tableau Conference, Charlotte, Carolina del Norte, <https://www.youtube.com/watch?v=bq0jeF9bv20>
- Jo, S. (1995). Machine Learning and Statistics: A matter of perspective. *Machine Learning and Statistics: A matter of perspective*, 1–9. https://www.researchgate.net/publication/2791124_Machine_Learning_and_Statistics_A_matter_of_perspective
- Kallina, W. (2018, 20 diciembre). *The Institutional Factbook*. AIR. <https://www.airweb.org/resources/publications/eair-newsletter/special-features/2018/12/20/the-institutional-factbook>
- Karnan, M., & Kalyani, P. (2010). Attribute reduction using backward elimination algorithm. *2010 IEEE International Conference on Computational Intelligence and Computing Research*, 1–4. <https://doi.org/10.1109/iccic.2010.5705893>
- Madsen, D. Ø., Slåtten, K., & Johanson, D. (2017). The emergence and evolution of benchmarking: a management fashion perspective. *Benchmarking*. Emerald Group Publishing Ltd. <https://doi.org/10.1108/BIJ-05-2016-0077>
- Mduma, N., Kalegele, K., & Machuve, D. (2019). Machine learning approach for reducing students dropout rates. *International Journal of Advanced Computer Research*, 9(42), 156–169. <https://doi.org/10.19101/IJACR.2018.839045>
- Nel, H. (2016). The Role of Institutional Research in Support of Strategic Planning. En SciSTIP, J. Botha, & N. J. Muller (Eds.), *Institutional Research in South African Higher Education—Intersecting Contexts and Practices* (1a ed., pp. 99–116). SUN MeDIA. <https://doi.org/10.18820/9781928357186/06>
- UNESCO. (2017, octubre). *Informe de Seguimiento de la Educación en el Mundo*. <https://unesdoc.unesco.org/ark:/48223/pf0000261016>
- Raschka, S. (2018). Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning. *University of Wisconsin*, 12–22. <https://arxiv.org/pdf/1811.12808.pdf>
- Saavedra, Mauricio., Pita-Carranza, María. & Opazo, Pablo. (2015). Institutional research in Latin America. En Karen Webber y Angel Calderon (Eds.), *Institutional research and planning in higher education. Global contexts and themes* (pp. 128-138). Londres: Routledge.
- Sandoval-Palis, I., Naranjo, D., Vidal, J., & Gilar-Corbi, R. (2020). Early Dropout Prediction Model: A Case Study of University Leveling Course Students. *Sustainability*, 12(22), 9314. <https://doi.org/10.3390/su12229314>

- Shmueli, G., Bruce, P. C., Gedeck, P., & Patel, N. R. (2021). *Data mining for business analytics*. 671.
- Suryadi, K. (2007). Framework of Measuring Key Performance Indicators for Decision Support in Higher Education Institution. 7.
- Volkwein, J. (2008). The Foundations and Evolution of Institutional Research. *New Directions for Higher Education*. 2008. 5 - 20. 10.1002/he.289.
- Webber, K., & Calderon, A. (2015). Institutional Research, planning, and decision support in higher education today. *ResearchGate*, 3–4, <https://www.researchgate.net/publication/283103366>

Anexos:

Anexo A: Criterios para la identificación de métricas dentro de las universidades.

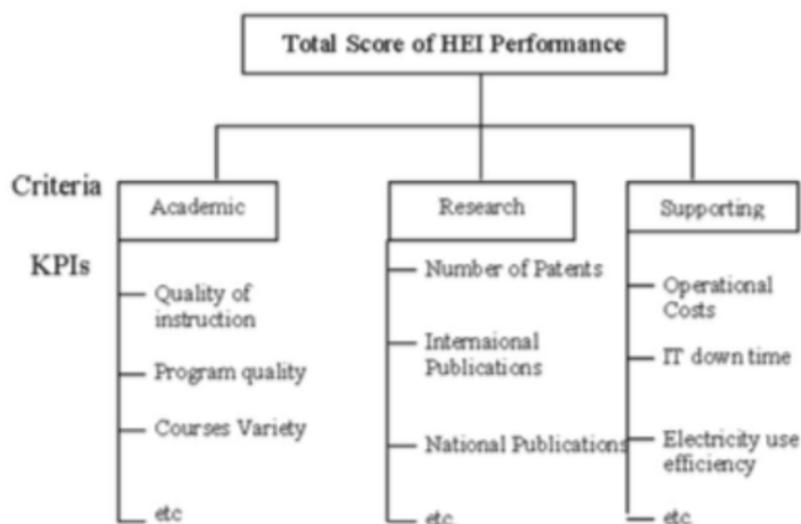
Nombre de la Asociación	Abreviatura	Establecido
Association for Institutional Research	AIR	1966
European Association for Institutional Research	EAIR	1979
Australasian Association for Institutional Research	AAIR	1988
Southern African Association for Institutional Research	SAAIR	1994
Canadian Institutional Research and Planning Association	CIRPA	1994
Southeast Asian Association for Institutional Research	SEAAIR	2001
China Association for Institutional Research	China AIR	2003
Middle East and North Africa Association for Institutional Research	MENA-AIR	2007
Higher Educational Institutional Research Network (*)	HEIR	2008

(*) British and Ireland Association for Institutional Research (established in late 1990s) evolved into the UK & Ireland

In Latin-America an IR association does not yet exist, but talks have been initiated in 2013 in Argentina and Ecuador to establish an association.

Fuente: Botha J. (2018)

Anexo B: Criterios para la identificación de métricas dentro de las universidades.



Fuente: Suryadi, K. (2007)

Anexo C: Organización de métodos de minería de datos de acuerdo con la naturaleza de los datos

	<u>Supervisados</u>		<u>No supervisados</u>
	<u>Respuesta Continua</u>	<u>Respuesta Categórica</u>	<u>Ninguna Respuesta</u>
<u>Predictores Continuos</u>	Linear regression k-Nearest neighbors Neural nets Ensembles	k-Nearest neighbors Logistics regression Neural nets Discriminant analysis Ensembles	Principal components Collaborative filtering Cluster analysis
<u>Predictores Categóricos</u>	Linear regression Regression trees Neural nets Ensembles	Naive Bayes Classification trees Logistic regression Neural nets Ensembles	Association rules Collaborative filtering

Fuente: Shmueli (2020)

Anexo D: Plantillas de requerimiento de dashboard

REQUERIMIENTOS DE LOS DASHBOARDS

- 1
Información general de los Dashboards

Nombre: _____

Objetivo: _____

Fecha Estimada: _____
- 2
Audiencia
Escoger todas las que apliquen

Ejecutivo

Gerencia

Análisis

Público
- 3
Modo de visualización
Escoger todas las que apliquen

Laptop

Tablet

Celular
- 4
Uso
Escoger todas las que apliquen

Web

PDF

Power Point

Imagen
- 5
Notas y filtros de los Dashboards

Enliste cualquier filtro que aplique y deba ser mostrado en el dashboard

 - Filtros de Dashboards:
 - Colores de la compañía: #EE9186 #DCAB27 #1D1D1D #ED3225

Fuente: Jaggernouth, 2019

Anexo E: Plantillas de requerimiento de dashboards por vista

REQUERIMIENTOS DE LOS DASHBOARDS





Nombre de la vista

Demanda de estudiantes por año



Descripción
Proveer de una descripción de alto nivel de la vista (Chart)

Demanda de estudiantes por año



Los campos de datos se utilizan en la vista
Enumere los campos necesarios para crear la vista. No incluya campos que se utilicen estrictamente como filtros

Dimensión Primaria: NA
Dimensión adicional: NA
Medición Primaria: NA
Medición Adicional: NA



Notas y Cálculos
Proporcione detalles para cualquier campo escuchado anteriormente que requiera un cálculo.

Estudiantes: Contar distinto a ID de Estudiantes
Profit Ratio : Profit/Sales



Formato
Describe cualquier formato requerido. Ejemplo Colorea las 10 categorías principales de azul y las categorías restantes de gris.

No se necesita un formato adicional en la vista



Filtros
Enumere cualquier filtro que deba aplicarse a la vista.

No se necesitan filtros adicionales en la vista



Base de datos
Proporcionar detalles sobre las fuentes de datos

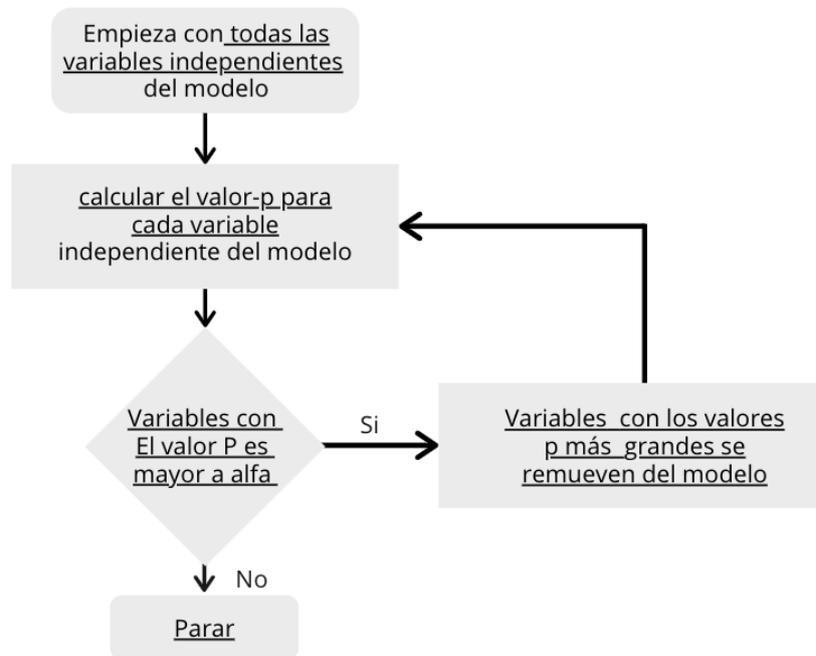
Estudiantes 1 - Archivo de Excel



Notas Adicionales
Proporcione detalles o sugerencias adicionales. Ejemplo: me gustaría ver esta vista como un gráfico apilado

Prefiero verlo como una tabla de texto

Fuente: Jaggernouth, 2019

Anexo F: Proceso de Backward Elimination.

Fuente: Karnam, M. (2010)

Anexo G: Estructura del libro de datos y los indicadores que lo conforman el mismo

Sección	Subsección	#	Indicadores
Estudiantes	Admisiones	1	Número de Aplicantes USFQ por Periodo Académico y Año Calendario
		1.1	Número de Aplicantes USFQ por Año Calendario
		1.2	Número de Aplicantes USFQ por Año Académico
		1.3	Aplicantes USFQ por Programa de Posgrado por Periodo Académico
		1.4	Porcentaje de Aplicantes por Países
		1.5	Porcentaje de Aplicantes por Colegio por Provincia
		1.6	Porcentaje de Aplicantes por Carrera Provincia/Estados
		2	Aplicantes de cada programa Académico por Género y periodo Académico
		3	Aplicantes USFQ por Colegio por Periodo Académico
		4	Aplicantes por Programa Académico por Género y periodo Académico
		5	Aplicantes USFQ por Carrera por Periodo Académico
		6	Estudiantes Matriculados por Primera vez por Periodo Académico
		6.4	Porcentaje de aplicantes por Provincia/Estados
		7	Estudiantes Matriculados por Primera vez de cada programa Académico por Género y Periodo Académico
		8	Estudiantes Matriculados por Primera vez por Colegio y Periodo Académico
		9	Número de Matriculados por Programa Académico por Género y periodo Académico
		10	Tasa de admisión por Periodo académico
	Estudiantes Registrados	1	Número de registrados en la Universidad
		2	Número de registrados por programa académico
		2.1	Número de registrados por programa académico y género
		3	Número de registrados por colegio
		3.1	Número de registrados por colegio y género
		4	Número de registrados por programa
		4.1	Número de registrados por programa y género
		5	Número de registrados por Países
		5.1	Número de registrados por Países y género
		6	Número de registrados por Provincias
		6.1	Número de registrados por Provincias y género
Personal en General	Personal en General	1	Clasificación del personal en la USFQ en las 3 categorías
		1.1	Clasificación del Personal Académico, Personal Universitario, en la USFQ en las 3 categorías por género

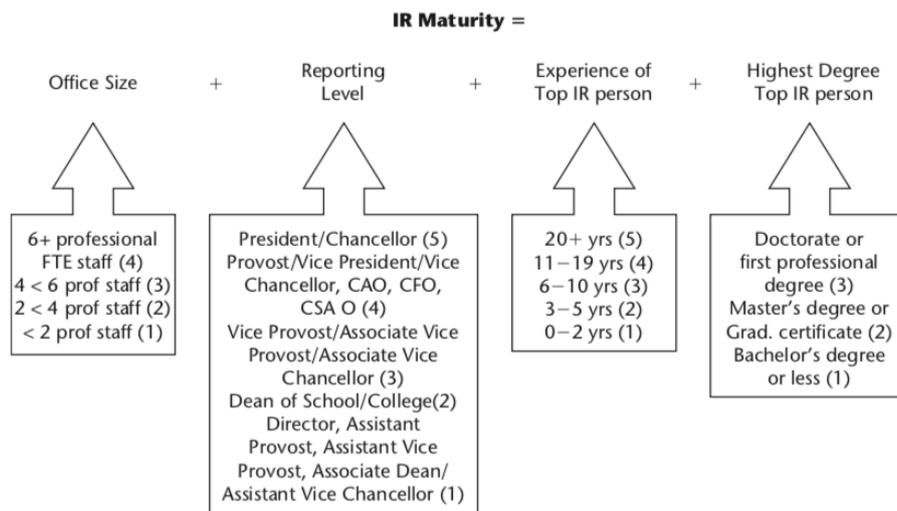
Personal Académico	Personal Académico	2	Número de personal académico por Colegio
		2.1	Clasificación del personal académico (Titular y No titular) por Género
		2.2	Clasificación del personal académico (Titular y No titular) por Nacionalidad & Colegio
		2.3	Clasificación del personal académico (Titular y No titular) por Nivel Académico (PhD, Máster)
		2.4	Clasificación del personal académico (Titular y No titular) por Tiempo de dedicación (T. Completo & Medio T.)
		2.5	Clasificación del personal académico (Titular y No titular) por Tiempo de dedicación (T. Completo & Medio T.) por género
		2.6	Clasificación Personal académico por Nacionalidad
	2.7	Número de cursos que da cada profesor	
	Titular	3	Clasificación Personal académico No Titular por Nacionalidad
	No Titular	3.1	Clasificación Personal académico Titular por nivel (1,2,3) & Género
	4	Clasificación Personal académico No Titular por Nacionalidad	
Personal Universitario	Personal Universitario	5	Clasificación de Nacionalidad del Personal Universitario
		5.1	Clasificación del Personal Universitario (Personal de Apoyo Académico & Administrativos)
	Personal de Apoyo Académico	6	Clasificación del Personal de Apoyo Académico por Colegio
		6.1	Clasificación del Personal de Apoyo Académico por Colegio & Género
		6.2	Clasificación del Apoyo Académico por Nivel Académico (Phd, Máster)
	Administrativos	7	Número del personal Administrativo por Departamento
		7.1	Clasificación del personal Administrativo por Departamento
		7.2	Clasificación del personal Administrativo por Género
Universidad y Otros Datos Institucionales	Investigación	8	Número de publicaciones por Profesor
		8.1	Número de publicaciones por programa académico, por colegio
		8.2	Número de publicaciones por género & nacionalidad
		8.3	Número de publicaciones p por nacionalidad
	Programas, carreras, cursos	9	Número de cursos ofertados por programa académico
		9.1	Número de cursos ofertados por carrera
		9.2	Número de carreras ofertadas, por colegio
		9.3	Número de subespecializaciones

Anexo H: Cuadrante Mágico de Gartner de Herramienta BI 2021



Fuente: Gartner Inc. (2021)

Anexo I: Modelo para el cálculo del nivel de madurez del departamento de IR propuesto por la AIR



Fuente: Volkwein, J. (2008)

Anexo J: Estructura del libro de datos y los indicadores que lo conforman el mismo

Vista	Objetivo	Fórmulas y cálculos
Dashboards de Admisiones		
Dashboards de Admisiones		
BAN de métricas de admisiones del periodo actual	Información resumen generada a través de un BAN (Big Angry Numbers) de varias métricas de admisiones en función de estudiantes postulados, admitidos y matriculados para el periodo actual.	Matriculación: Conteo de Estudiantes matriculados del periodo actual Aceptación: Conteo de Estudiantes Aceptados del periodo actual Postulantes: Conteo de Estudiantes Postulados del periodo actual
Gráfico y tabla de Métricas de Admisiones	Se creo un gráfico y una tabla de varias métricas de admisiones en función de estudiantes postulados, admitidos y matriculados.	Matriculación: Conteo de Estudiantes matriculados Aceptación: Conteo de Estudiantes Aceptados Postulantes: Conteo de Estudiantes Postulados Tasa de Rendimiento : Conteo de Estudiantes Matriculados/ Conteo de Estudiantes Aceptados Tasa de admisión : Conteo de Estudiantes Aceptados/ Conteo de Estudiantes Postulados
Información Geográfica- Admisiones		
Número de aplicantes por colegio y carrera	Se generó un diagrama de barras para poder apreciar el número de estudiantes postulados por colegio y carrera.	Postulantes: Conteo de Estudiantes Postulados dentro de un periodo.
Porcentaje de aplicantes por provincia	Se generó un gráfico en el cual se puede apreciar el número y porcentaje de aplicantes por provincia. Los cuales se los puede modificar con los filtros previamente mencionados.	% Aplicantes: Conteo de Estudiantes Postulados por provincia / Suma de Estudiantes Postulados
Tabla de resumen de la ubicación geográfica de los aplicantes	Se calculó una tabla en la cual se pueda apreciar el número y porcentaje de aplicantes por provincia. Los cuales se los puede modificar con los filtros previamente mencionados.	% Aplicantes: Conteo de Estudiantes Postulados por provincia / Suma de Estudiantes Postulados Número de Aplicantes: Conteo de Estudiantes Postulados por provincia / Suma de Estudiantes Postulados
Dashboards registrados		
Estudiantes Registrados		
BAN de métricas de estudiantes registrados del periodo actual	Información resumen generada a través de un BAN de número de estudiantes registrados de pregrado, posgrado y universidad para el periodo actual.	Número estudiantes de pregrado : Conteo de estudiantes de pregrado del periodo actual Número estudiantes de posgrado : Conteo de estudiantes de posgrado del periodo actual Número estudiantes de la universidad: Conteo de estudiantes de toda la universidad del periodo actual
Número de estudiantes registrados en la universidad en el tiempo	Vista del número de estudiantes a lo largo de los periodos académicos por programa académico.	Número de Estudiantes: Conteo de distintos de código del estudiante
Número de estudiantes registrados en la universidad por género	Vista de los estudiantes por programa académico y su distribución por género.	Número de estudiantes: Conteo de código de estudiante
Número de Estudiantes por	Tabla resumen del número de estudiantes por colegio y por programa	Número de estudiantes: Conteo de distintos de código de estudiante Total de

Colegio - tabla resumen	académico. Se muestra la información por año calendario y periodo académico.	estudiantes: Total del conteo de código de estudiante
Estudiantes Registrados		
Candidatos a títulos de pregrado y posgrado	Vista del número de estudiantes candidatos a un título de pregrado o posgrado a lo largo de los periodos académicos.	Número de Estudiantes: Conteo de distintos de código del estudiante
Tabla resumen de los candidatos a graduarse de pregrado y posgrado	Información de resumen del número de estudiantes candidatos a un título de pregrado o posgrado a lo largo de los periodos académicos y por género.	Número de Estudiantes: Conteo de distintos de código del estudiante Total de estudiantes: Total del conteo de distintos de código del estudiante
Estudiantes Registrados- Pregrado	Vistas del número de estudiantes por colegio y carrera y el porcentaje de los géneros masculino y femenino en cada gráfica.	Número de estudiantes: Conteo de distintos de código del estudiante de pregrado
Tabla resumen del número de estudiantes de pregrado por colegio y carrera	Información resumen del número de estudiantes de pregrado por colegio y carrera y el porcentaje de los géneros masculino y femenino en estos campos.	Número de estudiantes: Conteo de distintos de código del estudiante de pregrado Total de estudiantes: Total del conteo de distintos de código del estudiante de pregrado
Promedio de créditos registrados por colegio y carrera	Vista del promedio de créditos en los que están registrados los estudiantes de pregrado por colegio y carrera .	Promedio de créditos registrados: Promedio de Cred. Registrados Promedio: Promedio total del promedio de cred. registrado
Estudiantes Registrados- Posgrado		
Número de estudiantes de posgrado por colegio y carrera	Vistas del número de estudiantes por colegio y carrera y el porcentaje de los géneros masculino y femenino en cada gráfica.	Número de estudiantes: Conteo de distintos de código del estudiante de posgrado.
Tabla resumen del número de estudiantes de posgrado por colegio y carrera	Información resumen del número de estudiantes por colegio y carrera y el porcentaje de los géneros masculino y femenino en estos campos.	Número de estudiantes: Conteo de distintos de código del estudiante de posgrado Total de estudiantes: Total del conteo de distintos de código del estudiante de posgrado
Créditos registrados por colegio y carrera (máximo, promedio y mínimo)	Vista del máximo, promedio y mínimo de créditos en los que están registrados los estudiantes de posgrado por colegio y carrera.	Máximo de créditos registrados: Máximo de Cred. Registrados Promedio de créditos registrados: Promedio de Cred. Registrados Mínimo de créditos registrados: Mínimo de Cred. Registrados
Dashboards Personal USFQ		
Personal USFQ		
Clasificación del tipo de personal USFQ	Creación de BANs con el número total de personas por cada tipo de personal de la USFQ (académico , administrativo , apoyo Académico) con una rosquilla alrededor que indica el porcentaje de hombres y mujeres del mismo.	Conteo de tipo de personal : Conteo de tipo de personal % Sexo Masculino: Conteo de personal de Sexo Masculino por Tipo de personal / Total del Tipo de personal. % Sexo Femenino: Conteo de personal de Sexo Femenino por Tipo de personal / Total del Tipo de personal.
Tabla y Gráfico de Nivel & Nivel Académico de profesores Titulares	Creación de una tabla y un gráfico del nivel asignado por la universidad y el Nivel Académico de los profesores Titulares. De igual manera en la tabla se puede apreciar el total por cada uno de los niveles.	Conteo de tipo de personal : Conteo de tipo de personal % Sexo Masculino: Conteo de personal de Sexo Masculino por Tipo de personal / Total del Tipo de personal. % Sexo Femenino: Conteo de personal de

		Sexo Femenino por Tipo de personal / Total del Tipo de personal.
Tabla y Gráfico de la clasificación del tipo de personal por tiempo de dedicación del personal de la USFQ	Creación de una tabla y un gráfico en el cual se puede comparar la clasificación del tipo de personal por tiempo de dedicación de la USFQ. De igual manera en la tabla se puede apreciar el total por cada sexo uno de los niveles.	Conteo de tipo de personal : Conteo de tipo de personal por tiempo de dedicación
Información Geográfica - Personal Académico		
Gráfico y tabla de la nacionalidad del personal académico por nacionalidad	Se creó un gráfico geográfico y una tabla en el cual se puede apreciar nacionalidad del personal académico por nacionalidad y por género.	Número de personal académico por país : Conteo de personal académico por país.
Personal Académico		
BANs de las clasificación y total del personal académico por sexo	Se crearon tres BANs dos de la clasificación del personal académico (Titular y no titular) y uno del sub total, con una rosquilla alrededor que indica el porcentaje de hombres y mujeres del mismo.	Conteo de tipo del personal académico : Conteo de cada tipo de personal académico. Conteo del total del personal académico: Conteo del total del personal académico. % Sexo Masculino: Conteo de personal académico de Sexo Masculino por tipo de personal / Total del Tipo de personal académico. % Sexo Femenino: Conteo de personal de Sexo Femenino por Tipo de personal / Total del Tipo de personal.
Tabla y gráfico del título de nivel académico por nivel otorgado por la USFQ, colegio y carrera del personal académico	Se creó una tabla y un gráfico de la clasificación del personal académico titular por nivel otorgado por la USFQ , colegio y carrera.	Conteo de tipo del personal académico : Conteo de cada tipo de personal académico por Sexo, Colegio , Nivel y Nivel Académico,
Información Geográfica - Personal Académico		
Mapa de obtención de título por nivel Académico , nivel USFQ y carrera del personal académico del personal académico.	Se creó un gráfico geográfico y una tabla en el cual se puede apreciar el país de obtención de títulos del personal académico.	Número de títulos obtenidos por el personal académico por país: Conteo de los títulos obtenidos por el personal académico por país.
Tabla del título de nivel académico por nivel otorgado por la USFQ, colegio y carrera del personal académico	Se creó una tabla y un gráfico de la clasificación del personal académico titular por nivel otorgado por la USFQ , colegio y carrera.	Conteo de tipo del personal académico : Conteo de cada tipo de personal académico por Sexo, Colegio , Nivel y Nivel Académico,
Dashboards Publicaciones		
Publicaciones		
Números de publicaciones por Clasificación de publicación de la USFQ	Se crearon varios BANs con números de publicaciones por clasificación de publicación de la USFQ (Scopus, LatinIndex, Otros)	Número de publicaciones de Scopus: Conteo de número de publicaciones de Scopus. Número de publicaciones de LatinIndex: Conteo de número de publicaciones de LatinIndex Número de publicaciones de Otros: Conteo de número de publicaciones de Otros

Número de personal que publico por año de publicación sexo	Creación de BANs con el número de publicaciones por año con una rosquilla alrededor que indica el porcentaje de hombres y mujeres que realizaron publicaciones por cada año. De igual manera se generó in gráfico circular con el porcentaje de personal inactivo que publico durante ese año.	Creación de BANs con el número de publicaciones por año con una rosquilla alrededor que indica el porcentaje de hombres y mujeres que realizaron publicaciones por cada año. De igual manera se genero in gráfico circular con el porcentaje de personal inactivo que publico durante ese año
Clasificación de publicaciones por año	El gráfico nos muestra el número de de publicaciones por año y por la clasificación de publicación de la USFQ (Scopus, LatinIndex, Otros).	Número de publicaciones de Scopus: Conteo de número de publicaciones de Scopus por año Número de publicaciones de LatinIndex: Conteo de número de publicaciones de LatinIndex por año. Número de publicaciones de Otros: Conteo de número de publicaciones de Otros por año
Números de publicaciones por Clasificación de publicación de la USFQ	La tabla creada nos muestra el números de publicaciones por el tipo de personal, tipo de personal académico , clasificación de publicación de la USFQ (Scopus, LatinIndex, Otros) y sexo a través de los años.	Número de publicaciones : Conteo de número de publicaciones por el tipo de personal, tipo de personal académico , clasificación de publicación de la USFQ y sexo a través de los años.