

MINEDUC

Deserción IP/CFT

Integrantes

Cristóbal Pérez
Gianina Salomó
Leonardo Meneses
Werner González

Profesores:

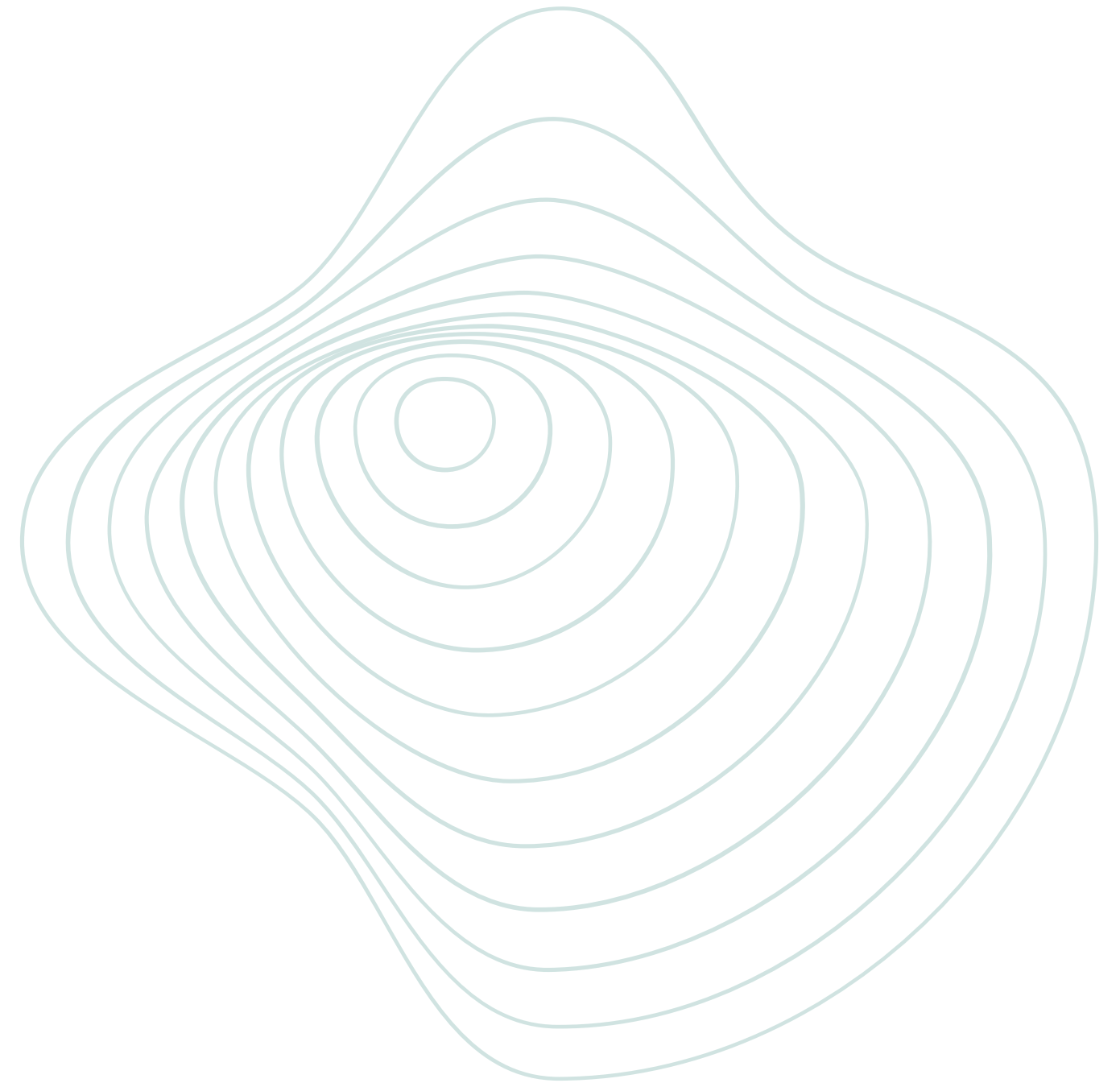
Constanza Contreras
Francisco Foster

Fecha

21 de Septiembre, 2022

Tabla de Contenidos

1. Introducción
2. Situación actual
3. Proyecto de Data Science
4. Revisión de Literatura
5. Propuesta de solución
6. Conclusiones - Reflexiones finales



Introducción

Organización



MINEDUC, MINISTERIO DE EDUCACIÓN

- Industria: Gobierno Central

MISIÓN MINEDUC

"Asegurar un sistema educativo inclusivo y de calidad que contribuya a la formación integral y permanente de las personas y al desarrollo del país, mediante la formulación e implementación de políticas, normas y regulación, desde la educación parvularia hasta la educación superior."

CONTRAPARTE

- **Danny Avello:** Asesor Técnico de la División de Información y Acceso del MINEDUC

DIVISIÓN DE INFORMACIÓN Y ACCESO

- Desarrollar y mantener el Servicio de Información de Educación Superior
- Coordinar el Sistema de Acceso a las Instituciones de Educación Superior
- Mantener el Registro de las Instituciones de Educación Superior reconocidas oficialmente por el Estado y de aquellas cuyo reconocimiento oficial y personalidad jurídica hubiese sido revocada.

Negocio



NECESIDADES E INTERESES

- Bajar la tasa de deserción de estudiantes de carreras técnico profesional.
- Desarrollar políticas públicas para favorecer el egreso de la educación superior.

RESTRICCIONES

- Falta de tiempo
- Falta de análisis para encontrar variables relevantes que causen deserción para crear una política pública adecuada

OBJETIVO

Lograr que los estudiantes de carreras técnico profesionales permanezcan en sus instituciones (CFT e IP).

Situación actual

Problema

ALTA DESERCIÓN EN CARRERAS DE IP/CFT

- Aproximadamente un 14,5% de los estudiantes entre 2016 y 2021 ha desertado sin volver a la educación superior.

Situación actual

FACTORES QUE INFLUYEN EN EL PROBLEMA

- Poco tiempo invertido en resolver la problemática por parte del MINEDUC.
- No hay un instrumento universal para el ingreso a estas instituciones.
- Estudiantes muy heterogéneos.
- Presupuesto

Proyecto de Data Science

Proyecto



PRESENTACIÓN

Crear un modelo que sea capaz de predecir la probabilidad de deserción de un estudiante de carrera técnico profesional junto con identificar las variables relevantes de su causa.

OBJETIVO GENERAL

Encontrar asociación (posteriormente causales) entre variables relevantes para poder armar un buen modelo que prediga la deserción.

Alcance, riesgos y resultados esperados

ALCANCE

Centrarse en aquellos estudiantes que no vuelven al sistema de educación superior.

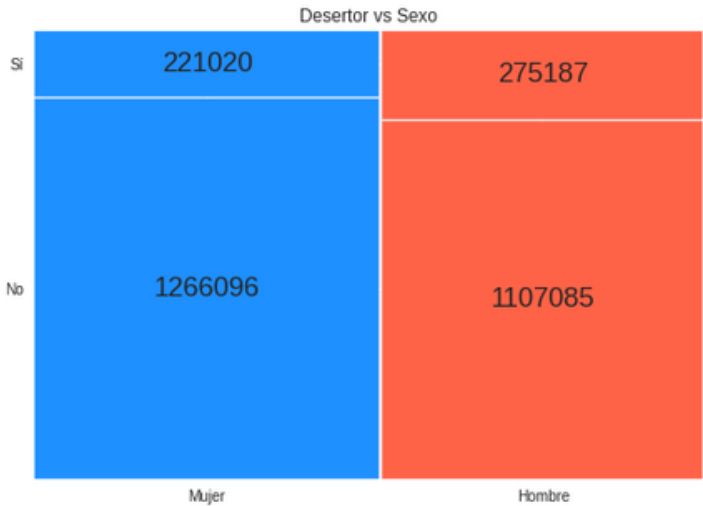
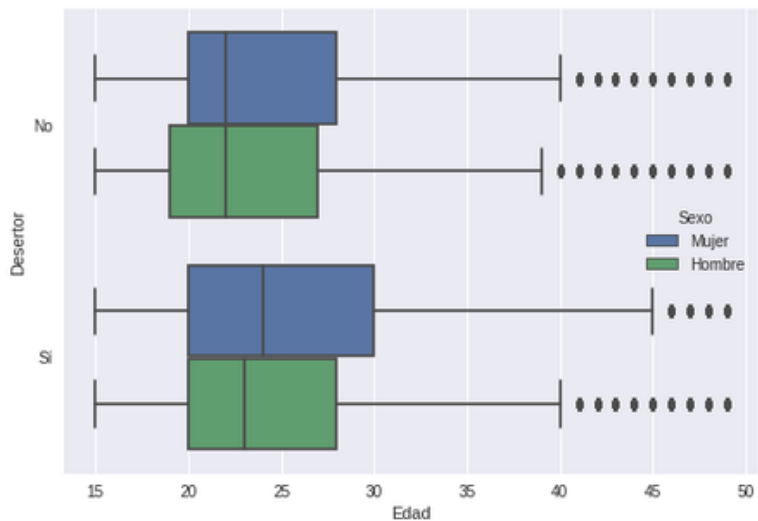
ENTREGABLE

Datos analizados en archivos .CSV, junto con presentar los resultados de manera visual con gráficos, árboles de decisión, etc.

periodo	codigo_unico	mrun	gen_alu	fec_nac_alu	tipo_inst_1	region_sede	area_conocimiento	region_colegio	tipo_colegio	notas_colegio	dep_colegio	deserto
2020	I111S3C48J1V1	9405558.0	1	199511	Institutos Profesionales	Biobio	Tecnología	7	310.0	5,8	3.0	0.0
2021	I116S6C196J1V1	3024665.0	2	199802	Institutos Profesionales	Maule	Educación	7	310.0	5,7	3.0	1.0

RESULTADOS ESPERADOS TENTATIVOS

Encontrar variables más relevantes para la permanencia de un estudiante y cuáles sirven para determinar la deserción.



Alcance, riesgos y resultados esperados

POTENCIALES RIESGOS

- Existencia de muchos datos nulos dentro del total de instancias.
- Acceso a información sociodemográfica potencialmente relevante.
- Calidad de los datos en bases de datos adyacentes.
- Tiempos de procesamiento de algoritmos.
- Generalización de resultados.
- Definición de deserción no entregue resultados favorables.
- Interpretación de resultados en la toma de decisiones.



Desafíos de aprendizaje

DEFINICIÓN DEL DESERTOR

Encontrar en conjunto con la contraparte, una definición apropiada de lo que es un desertor para el estudio.

CONSTRUCCIÓN DEL VECTOR OBJETIVO

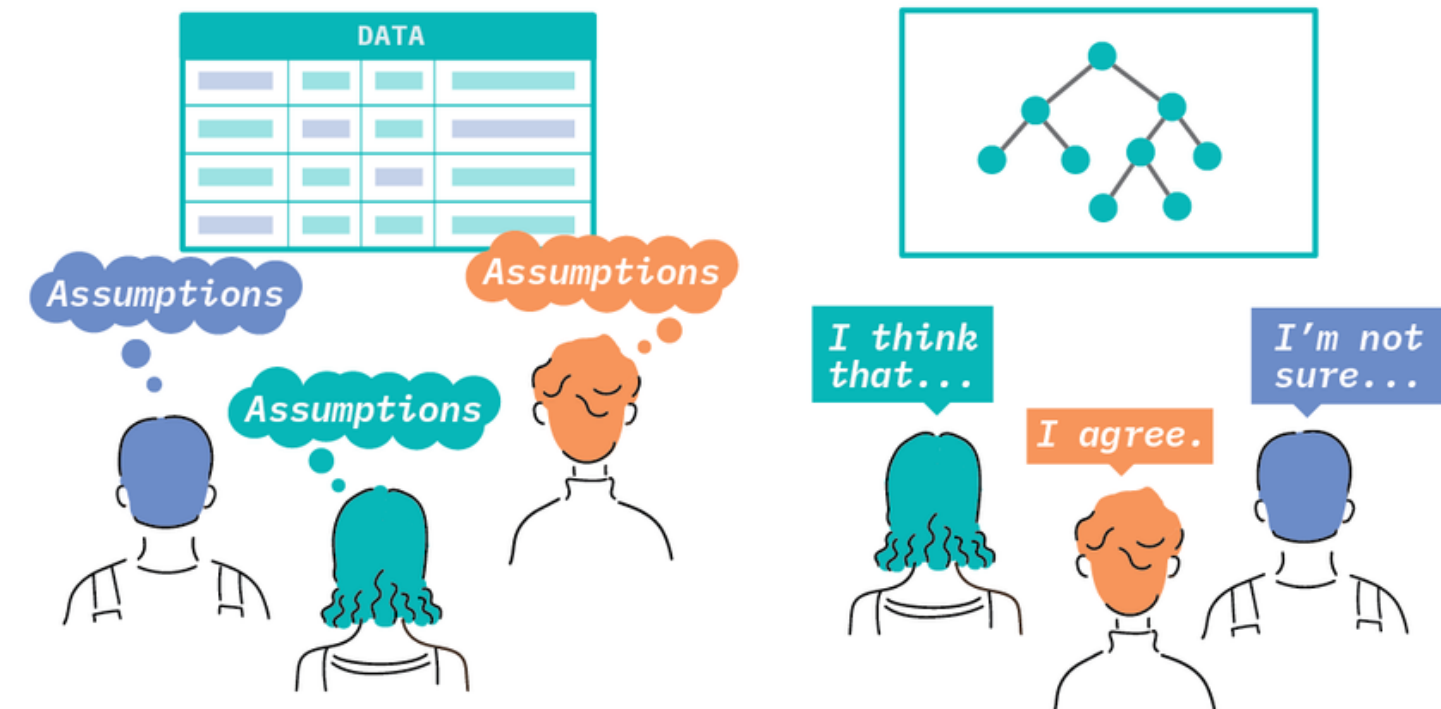
En base a la data presente, encontrar la regla para etiquetar las instancias entre desertor y no desertor.

PROCESAR LOS DATOS

Existe una gran cantidad de datos a procesar, limpiar y analizar, en conjunto con cruzar los datos con variables demográficas para poseer mayor información descriptiva.

ASOCIACIONES CAUSALES

Encontrar asociaciones causales de las deserción.



Datos

DESCRIPCIÓN

- Ventana de tiempo: 2016 a 2021.
- Frecuencia: anual.
- Cantidad de observaciones:
 - Matriculados: 4.131.609 observaciones.
 - Titulados: 748.517 observaciones.
- Cantidad de atributos:
 - Matriculados: 52 features.
 - Titulados: 41 features.

**Variable a construir, correspondiente a la etiqueta de si la persona deserta de la carrera o no.*

VARIABLES DE INTERÉS

Periodo	Género
RUN	Edad
Código Único	Notas
Código de Institución y Sede	Región de Procedencia
Código de Carrera, Versión y Jornada	Tipo de Educación Previa
Tipo de Institución	Área del Conocimiento
Fecha de Obtención de Título	Deserción*

Revisión de literatura

Estado del arte

A NIVEL DE NEGOCIO

- Investigaciones realizadas por **FONIDE¹** y **SIES²**
- **Deserción IP / CFT:**
 - Variante entre 30% y 55% entre 2007 y 2010
 - Leve tendencia de disminución desde 2013 (35%) a 2017 (31%)

1 *Determinantes de deserción en la Educación Superior Chilena, con énfasis en Efecto de becas y créditos (2013)*

2 *Deserción de primer año y Reingreso a la Educación Superior en Chile - Análisis Cohorte 2015 (2019); Informe retención de 1er año de pregrado cohortes 2013-2017 (2018)*

Mayor Deserción	Mayor Permanencia
Instituciones no acreditadas	Instituciones acreditadas
Sin beneficios	Con Beneficios
Sobre 19 años 1er año	Hasta 19 años 1er año
Colegio PP*	Colegios Mun / Sub*
Hombres	Mujeres
Colegios CH*	Colegios TP*
IP / CFT	Universidades
Primer año de carrera	Desde el 2do año de carrera

*Considerando matrículas de carreras técnicas en educación superior

Estado del arte

MODELOS DE DATA SCIENCE / DATOS

- En mayor parte sobre deserción Universitaria
- Diferentes formas de considerar deserción
- Clases desbalanceadas
- Variables regresoras/predictoras agrupadas en:
 - Variables sociodemográficas
 - Variables académicas
 - Características institución educación escolar
 - Características institución educación superior
 - Beneficios económicos

Las mejores variables y mejor modelo dependen de la definición del vector objetivo, pero en general regresión logística y árboles tienen buen desempeño, junto con los atributos socio demográficos y la presencia de beneficios económicos.

Modelos supervisados	Cantidad de referencias encontradas
Regresión Logística*	10
Árboles de decisión	6
Redes Neuronales	5
SVM	4

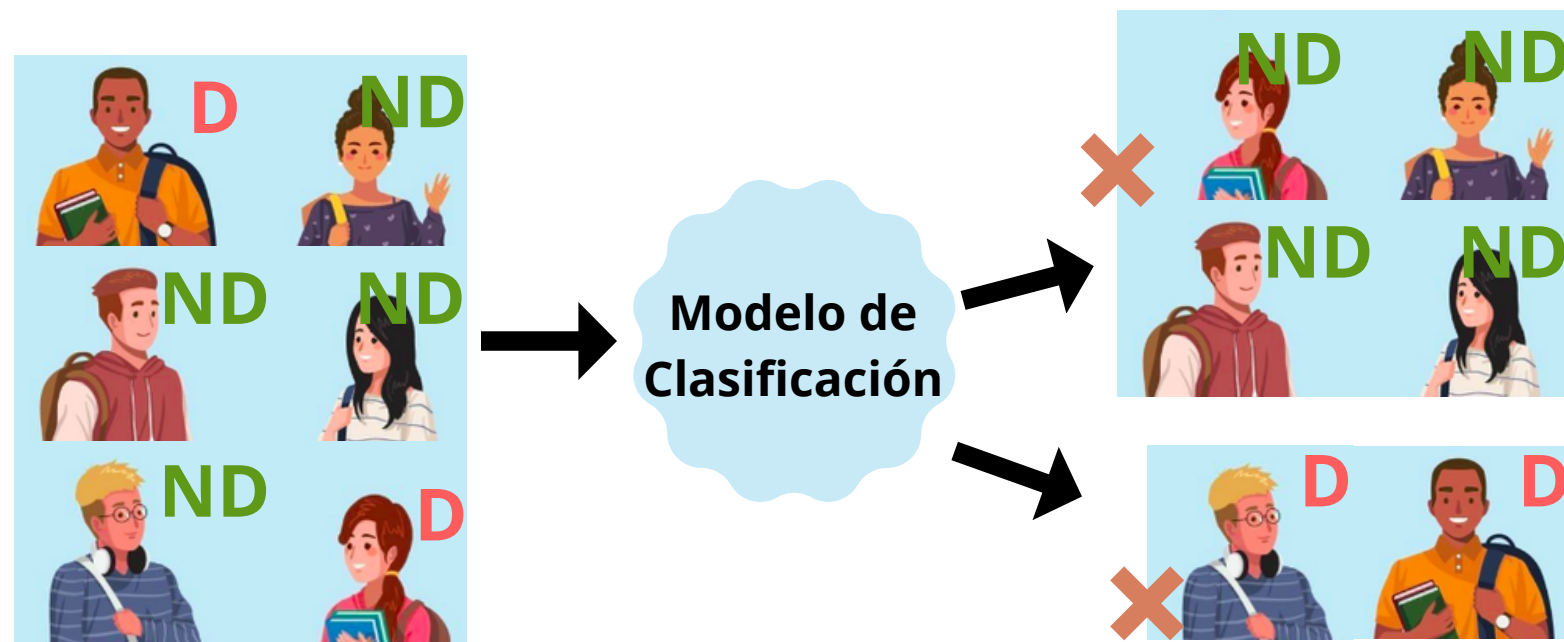
**Incluyendo modelos econométricos logit / probit*

Propuesta de solución

Estrategia

TIPO DE PROBLEMA

Corresponde a un problema *supervisado* de *clasificación*, con desbalance de clases.



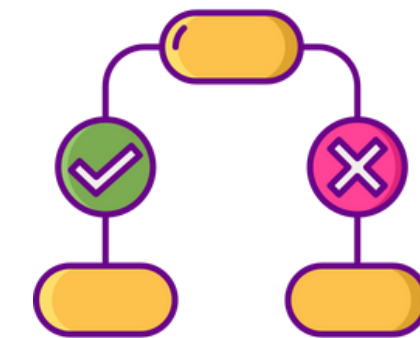
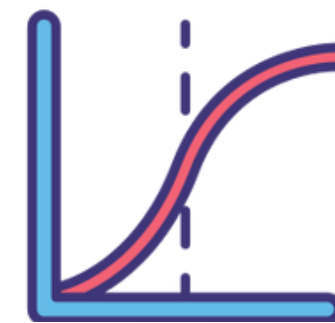
TÉCNICAS / ENFOQUES

Determinar asociación: Variables vs Deserción

- *Regresión Logística*
- *Árboles de Decisión*

Sin balanceo de clases (Inicialmente al menos)

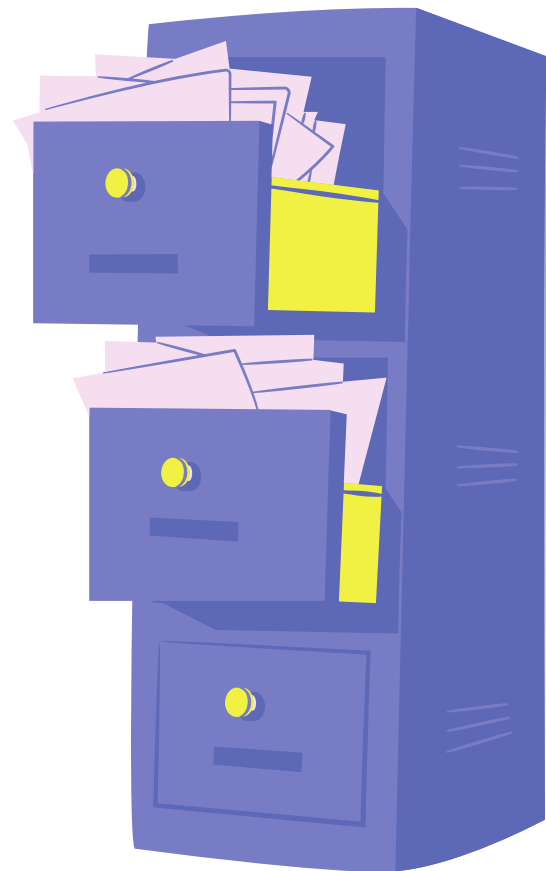
- Se ha observado un buen desempeño sin su uso.



Plan de trabajo

ORGANIZACIÓN ACTUAL

- WhatsApp.
- Github.
- Objetivos definidos en reuniones.
- Participación voluntaria.



ORGANIZACIÓN DESEADA

- Carta Gantt.
- Guiarnos por la metodología CRISP DM y fechas de presentación.
- Asignar responsables a cada tarea.



Plan de trabajo

Titulo del proyecto:	Deserción en Institutos Profesionales y Centros de Formación Tecnica															
Tareas	Semanas															
	3	4	5	Receso	6	7	8	9	10	11	Receso	12	13	14	15	Exámenes
Primer ciclo CRISP-DM																
Comprensión del problema																
Explorar datos																
Definición de desertor																
Presentación 1																
Preparación de los datos																
Modelado																
Evaluación																
Presentación 2																
Segundo ciclo CRISP-DM																
Comprensión del problema																
Preparación de los datos																
Presentación 3																
Modelado																
Evaluación																
Presentación 4																
Tercer ciclo CRISP-DM																
Comprensión del problema																
Preparación de los datos																
Modelado																
Evaluación																
Despliegue																
Presentación Final																

Conclusiones

Reflexiones finales

- Reuniones periódicas con la contraparte ayuda al avance consistente del proyecto
- Se logró acotar el problema (definir el tipo de deserción) para el análisis de los datos
- Es importante comunicarse de forma efectiva con la contraparte para definir objetivos
- Es posible pedir más recursos si se estiman convenientes. Por ej: pedir datos socioeconómicos de los estudiantes
- Se debe recurrir a otras fuentes de datos para complementar con datos iniciales

MINEDUC

Deserción IP/CFT

Integrantes

Cristóbal Pérez
Gianina Salomó
Leonardo Meneses
Werner González

Profesores:

Constanza Contreras
Francisco Foster

Fecha

21 de Septiembre, 2022