

# Tabla de contenido

<b>1.</b>	<b>INTRODUCCIÓN</b>	<b>1</b>
1.1.	SARS-CoV-2	1
1.2.	Proyecto ANID	1
1.2.1.	Concurso y adjudicación de fondos	1
1.2.2.	Áreas del Proyecto	2
1.3.	Servicio de Salud Metropolitano Sur Oriente	3
1.3.1.	Información general	3
1.3.2.	Unidad de Salud Digital	3
1.3.3.	Situación actual y oportunidad	3
1.4.	Objetivos	4
1.4.1.	Objetivo general	4
1.4.2.	Objetivos específicos	4
1.5.	Hipótesis de investigación	4
1.6.	Metodología	5
1.7.	Resultados esperados	6
<b>2.</b>	<b>MARCO TEÓRICO</b>	<b>7</b>
2.1.	Feature Engineering	7
2.1.1.	Imputación de datos	7
2.1.2.	Manejo de valores atípicos	8
2.1.3.	Binning	9
2.1.4.	One-Hot encoding	9
2.2.	Dimensionalidad	9
2.3.	Análisis de Clústeres	12
2.3.1.	Objetivo general	12
2.3.2.	Principales familias de algoritmos	12
2.3.3.	Métricas de desempeño	13
2.4.	Aprendizaje supervisado	14
2.4.1.	Objetivo general	14
2.4.2.	Tipos de error	14
2.4.3.	Métricas de desempeño	15
2.5.	Trabajos relacionados	15
2.5.1.	Respuesta inmunológica ante el Sars-CoV-2	15
2.5.2.	Dinámicas Sintomáticas	16
2.5.3.	Factores de riesgo	19
2.5.4.	Variables sociales en Chile	19
<b>3.</b>	<b>DESARROLLO DEL TRABAJO</b>	<b>22</b>
3.1.	Comprensión del dominio	22
3.2.	Elección y creación del set de datos	22
3.3.	Preprocesamiento y limpieza	27
3.4.	Transformación de los datos	27

<b>4.</b>	<b>RESULTADOS</b>	<b>40</b>
4.1.	Clustering	40
4.1.1.	Cantidad de clústers y desempeño	40
4.1.2.	Trayectorias sintomáticas y caracterización	42
4.1.3.	Trayectorias con desenlaces críticos	50
4.1.	Modelo de predicción	53
4.1.	Muestra de validación - Clustering	58
<b>5.</b>	<b>DISCUSIÓN</b>	<b>64</b>
5.1.	Análisis del clustering y validación	64
5.2.	Análisis de los modelos predictivos	71
<b>6.</b>	<b>CONCLUSIONES</b>	<b>73</b>
6.1.	Algoritmos, técnicas y métricas implementadas	73
6.2.	Objetivos y resultados	74
6.3.	Hipótesis	75
6.4.	Base de datos y limitaciones	76
6.5.	Trabajo futuro	77
<b>7.</b>	<b>BIBLIOGRAFÍA</b>	<b>78</b>
<b>8.</b>	<b>ANEXOS</b>	<b>83</b>