

Tabla de Contenido

1	Planteamiento del problema	1
1.1	Antecedentes generales	1
1.1.1	Contexto del trabajo	1
1.1.2	COVID-19: el virus, la enfermedad y la pandemia	2
1.1.3	COVID-19 en Chile	2
1.1.4	Actores relevantes	6
1.2	Problema y justificación	9
1.2.1	Área donde se desarrollará el trabajo	9
1.2.2	Problema identificado y su relevancia	9
1.2.3	Justificación del problema	10
1.2.4	Valor generado a través de la solución	11
2	Objetivos	13
2.1	Objetivo general	13
2.2	Objetivos específicos	13
2.3	Alcances	14
2.3.1	Coronavirus, una historia en desarrollo	14
2.3.2	Modelos predictivos, sin capacidad prescriptiva	14
2.3.3	Integración con los sistemas de salud del SSMSO	14
3	Marco conceptual	15
3.1	Machine Learning	15
3.2	Entrenamiento y validación de modelos de Machine Learning	15
3.3	Técnicas para manejo de datos desbalanceados	17
3.3.1	Under-Sample	18
3.3.2	Over-Sample	18
3.3.3	SMOTE	18
3.4	Árboles de decisión (CART)	18
3.5	Algoritmos de ensamble	21
3.5.1	Bagging	21
3.5.2	Boosting	21
3.6	Optimal Classification Tree	21
3.7	Métricas de desempeño de los algoritmos	22
3.7.1	Matriz de confusión	22
3.7.2	Curva ROC y AUC	24
4	Metodología	26
4.1	Fase de comprensión del negocio	27
4.1.1	Entendiendo el problema u oportunidad	27

4.1.2	Los datos	30
4.2	Fase de recolección y comprensión de los datos	34
4.2.1	Base de datos	34
4.2.2	Atributos de la base	34
4.2.3	Modificación de atributos	36
4.2.4	Creación de atributos	37
4.2.5	Variable a predecir	39
4.2.6	Análisis descriptivo de los datos	41
4.3	Fase de análisis y selección de datos	47
4.3.1	Población de estudio pacientes ingresados al hospital	47
4.3.2	Población de estudio pacientes en seguimiento domiciliario	53
4.3.3	Perspectiva general del estudio	58
4.4	Fase de modelado	59
4.4.1	Machine Learning para predecir riesgo de fallecer en pacientes hospitalizados por COVID-19	59
4.4.2	Machine Learning para predecir la severidad de la enfermedad en pacientes COVID-19	63
4.5	Fase de evaluación y despliegue	66
4.5.1	Evaluación con nuevos datos	66
4.5.2	Diseño de la aplicación	67
5	Conclusiones	68
5.1	Resumen del trabajo realizado	68
5.2	Principales conclusiones	69
5.3	Trabajo futuro	69
5.3.1	Integrar los modelos con los sistemas del SSMSO	69
5.3.2	Vacunas	70
5.3.3	Modelo tiempo en estadía	70
	Bibliografía	72
	Apéndice A Algoritmos utilizados	73
A.1	Bagging	73
A.2	Random Forest	74
A.3	Boosting	75
A.4	Extreme Gradient Boosting	76
	Apéndice B Proceso seguimiento plataforma COVID19	77
B.1	Ingreso información del paciente	77
B.2	Detalle de atención del paciente	80
B.3	Visualización pacientes en seguimiento	81
B.4	Informe de seguimientos	82
B.5	Criterios de clasificación de riesgo	83
	Apéndice C	84
C.1	Glosario	84
C.2	Definiciones síntomas	85
C.3	Matriz de correlación	87
C.4	Aplicación web	88
	Apéndice D Análisis pacientes vacunados	89