

Índice

1. Introducción.....	7
1.1. Contexto del Problema.....	7
1.2. Objetivos.....	9
1.2.1. Objetivo General.....	9
1.2.2. Objetivos Específicos	9
1.3. Contribuciones de la Tesis	9
1.4. Organización de la Tesis.....	10
1.5. El problema de Estimación de calidad del aire	10
1.5.1. El problema de estimación de calidad del aire en la ciudad de Talca	11
1.5.2. Modelación de calidad del aire con algoritmos de aprendizaje automático.....	12
1.6. Algoritmos de Aprendizaje Automático	14
1.6.1. Modelación con Algoritmo TTOSOM.....	14
1.6.2. Perceptrón Multicapa (MLP).....	15
1.6.3. Random Forest (RF).....	16
1.7. Metodología.....	18
1.7.1. Base de Datos Original.....	19
1.7.2. Normalización de Datos	26
1.7.3. Manipulación de Datos Faltantes	26
1.7.4. Eliminación de valores atípicos	27
1.7.5. Medidas de precisión para modelos de regresión	28
1.7.6. Modelación con Algoritmo TTOSOM.....	29
1.7.7. Modelación con Algoritmo Perceptrón Multicapa (MLP)	31
1.7.8. Modelación con Algoritmo Random Forest (RF)	32
1.7.9. Validación del Modelo	32
1.8. Resultados Experimentales.....	33
1.8.1. Resultados de regresión.....	33
1.8.2. Predicción de Episodios de Emergencia.....	35

1.9.	Conclusiones	38
2.	Artículo: Predicción de calidad del aire con el algoritmo Random Forest en la ciudad de Talca, Chile.	40
2.1.	Introducción	40
2.1.1.	Contribución del paper	42
2.1.2.	Organización del paper	42
2.2.	El problema de estimación de calidad del aire	42
2.3.	Materiales y métodos	44
2.3.1.	Conjunto de Datos	44
2.3.2.	Procesamiento del Conjunto de Datos	47
2.3.3.	Random Forest (RF)	47
2.3.4.	Medidas de precisión para modelos de regresión	48
2.3.5.	Validación de Modelos	49
2.4.	Resultados	49
2.4.1.	Predicción de Episodios de Emergencia	50
2.5.	Conclusiones	51
2.6.	Agradecimientos	52
2.7.	Referencias Bibliográficas	53

Índice de Tablas

Tabla 1:	Medidas de emergencia ambiental según norma primaria de calidad	8
Tabla 2:	Sensores de medición de parámetros contaminantes y meteorológicos de las estaciones La Florida, U.C. Maule y Universidad de Talca (Ministerio de Medio Ambiente 2009-2015).	19
Tabla 3:	Correlación entre la concentración 24 horas de $MP_{2,5}$ con las variables temperatura promedio 24 horas, velocidad del viento 24 horas y concentración 24 horas de $MP_{2,5}$ del día anterior en las estaciones Cefam La Florida, U.C. Maule y Universidad de Talca.	21
Tabla 4:	Enfoques, ventajas y desventajas en missing data (Bougoudis, Demertzis and Iliadis 2015)	26
Tabla 5:	Medidas de Precisión.	28
Tabla 6:	Grilla de Parámetros del modelo TTOSOM	29

Tabla 7: Algoritmo de Interpolación de neuronas de TTOSOM.....	30
Tabla 8: Parámetros que optimizan el resultado del algoritmo TTO-SOM	31
Tabla 9: Grilla de Parámetros de MLP	32
Tabla 10: Parámetros que optimizan el resultado del algoritmo MLP	32
Tabla 11: Número de registros y porcentajes para validación de los modelos	33
Tabla 12: Resultados de regresión del algoritmo TTOSOM	33
Tabla 13: Resultados de regresión del algoritmo MLP	34
Tabla 14: Resultados de regresión del algoritmo RF.....	34
Tabla 15: Tabla de contingencia de la estación Estación La Florida entre 01/07/2015 al 15/08/2015 .	36
Tabla 16: Tabla de contingencia de la estación Universidad Católica del Maule entre 01/07/2015 al 14/07/2015.....	37
Tabla 17: Tabla de contingencia de la estación Universidad de Talca entre 01/07/2015 al 14/08/2015	38
Tabla 18: Correlación entre la concentración 24 horas de $MP_{2,5}$ con las variables temperatura promedio 24 horas, velocidad del viento 24 horas y concentración 24 horas de $MP_{2,5}$ del día anterior en las estaciones Cesfam La Florida, U.C. Maule y Universidad de Talca.	46
Tabla 19: Resultados de regresión del algoritmo RF.....	49
Tabla 20: Tabla de Contingencia de niveles de calidad del aire.....	51

Índice de Figuras

Figura 1: Estaciones de monitoreo de calidad del aire en la ciudad de Talca.....	8
Figura 2: Cómo aprenden una distribución en forma de triángulo (Astudillo and Oommen 2011).....	15
Figura 3: Ejemplo de un perceptrón multicapa con dos capas ocultas (Gardner and Dorling 1998)	16
Figura 4: Arquitectura general de Random Forest (Verikas, Gelzinis and Bacauskiene 2011)	17
Figura 5: Metodología Experimental	18
Figura 6: Promedio diarios de $MP_{2,5}$ entre los meses de Abril de 2014 y Agosto de 2015	21
Figura 7: Comportamiento de las variables meteorológicas y de contaminación con respecto a la variable a estimar, entre los meses de Abril de 2014 y Agosto de 2015.	22
Figura 8: Promedio Diario de $MP_{2,5}$ entre los meses de Abril y Agosto de los años 2014 y 2015	24
Figura 9: Registros horarios de $MP_{2,5}$ entre los meses de Abril y Agosto de los años 2014 y 2015 ...	25

Figura 10: Rendimientos de modelación en los datos de validación para el promedio diario móvil de $MP_{2,5}$	35
Figura 11: Estaciones de monitoreo de calidad del aire en la ciudad de Talca.....	41
Figura 12: Promedio Diario de $MP_{2,5}$ entre los meses de Abril y Agosto entre los años 2014 y 2017 ..	45
Figura 13: Promedios horarios de registros de $MP_{2,5}$ entre los meses de Abril de 2014 y Agosto de 2017.....	46
Figura 14: Arquitectura general de Random Forest (Verikas, Gelzinis and Bacauskiene 2011).	48
Figura 15: Rendimientos de modelación en los datos de validación para el promedio diario móvil de $MP_{2,5}$	50