



Universidad del Azuay

Facultad de Ciencias de la Administración

Escuela de Ingeniería de Sistemas y Telemática

**RECOMENDADOR PARA ACTIVIDADES AL
AIRE LIBRE SEGÚN CONTAMINACIÓN
AMBIENTAL: REVISIÓN SISTEMÁTICA Y
MÉTODO DE CREACIÓN**

Trabajo de titulación previo a la obtención del grado en:

Ingeniero de Sistemas y Telemática

Autores:

Pablo Santiago Loja Arévalo

John Gerardo Calle Buestan

Director:

Marcos Orellana Cordero

Cuenca – Ecuador

2022

DEDICATORIA

Dedico este trabajo de titulación a mi madre, quien, con su esfuerzo, perseverancia y constancia, me ha ayudado a superar y afrontar las dificultades presentadas a lo largo de mi vida, siendo mi ejemplo y modelo a seguir.

A mis abuelos y tías quienes han sido pilares durante toda mi vida, ya que son sinónimos de unión y fortaleza. A mi hermano quien ha sido mi compañero brindándome apoyo en los momentos más difíciles.

A mi pareja, que siempre me ha motivado y que, gracias a ella, me ha enseñado otra forma de entender la vida.

Finalmente, a mis familiares, compañeros y amigos ya que han formado parte de esta experiencia de vida.

DEDICATORIA

Este trabajo se lo dedico a las personas que estuvieron presente durante el transcurso de mi etapa universitaria, en especial a mi madre, mi ejemplo a seguir, quien a pesar de la distancia siempre estuvo ahí, apoyándome, dándome ánimos para seguir adelante, y poder culminar con mis estudios.

De igual manera a mi padre que me apoyó económica y moralmente, ya que sin su ayuda nada de esto hubiera sido posible.

A mi hermana quien es como mi segunda madre por cuidarme y darme fuerzas en los momentos más difíciles de mi vida. A mi sobrina quien es como una hermana menor por estar siempre presente.

Y una dedicación especial a mi abuelita que ya no se encuentra con nosotros, pero sé que estaría orgullosa por haber alcanzado esta meta.

Finalmente, a mis compañeros y amigos por darme la confianza y el apoyo incondicional

AGRADECIMIENTO

Agradezco a nuestro director del presente trabajo de titulación, Marcos Orellana, quien, con su conocimiento, guía y consejos, supo orientarme a lo largo de toda la carrera.

Agradezco de igual manera, a todo el grupo de investigación LIDI, por su ayuda y colaboración en la realización de este proyecto ya que sin ellos nada de esto hubiera sido posible.

Agradezco a mi familia por su apoyo, paciencia y por brindarme las fuerzas para cumplir esta etapa de mi vida, a mi amigo y compañero de tesis, Pablo, quien supo ser un apoyo durante gran parte de la carrera. Agradezco también a todos los profesores que formaron parte de mi instrucción académica, por compartir sus conocimientos y por la paciencia otorgada.

Agradezco a mi compañero de tesis John Calle por su apoyo y amistad durante el transcurso de la carrera y también fuera de la misma. También hago llegar mi agradecimiento a la Universidad del Azuay, y a todos mis amigos y compañeros que formaron parte de esta increíble etapa.

ÍNDICE

DEDICATORIA	I
DEDICATORIA	II
AGRADECIMIENTO	III
ÍNDICE	IV
RESUMEN	VII
ABSTRACT.....	VIII
CAPÍTULO 1	1
1. INTRODUCCIÓN	1
CAPÍTULO 2.....	2
2. TRABAJOS RELACIONADOS	2
CAPÍTULO 3.....	3
3. REVISIÓN SISTEMÁTICA DE LITERATURA.....	3
3.1 Planteamiento del SLR	3
3.1.1 Campo a estudiar.....	3
3.1.2 Necesidad de la revisión sistemática.....	3
3.1.3 Preguntas de investigación.....	4
3.1.4 Elección de la cadena de búsqueda	4
3.1.5 Selección de bibliotecas e indexadores digitales	5
3.1.6 Criterios de exclusión e inclusión	6
3.1.7 Evaluación de calidad	6
3.1.8 Estrategia de extracción de datos	7
3.2 Ejecución de revisión.....	9
3.2.1 Selección de estudios primarios.....	9
3.2.2 Extracción de datos	9
3.3 Reporte de resultados y difusión.....	11
3.3.1 Variables de clasificación usadas en la calidad del aire	11
3.3.2 Enfoques de los sistemas de recomendación aplicados según el dominio.....	12
3.3.3 Diversos enfoques de los sistemas de recomendación usados en los diversos tipos de clasificación	13
3.3.4 Enfoques de los sistemas de recomendación usados en las diversas áreas de conocimiento.....	14
3.3.5 Fuentes de conocimiento usados en los diferentes enfoques de los sistemas de recomendación	15
CAPÍTULO 4.....	17
4. MATERIALES Y MÉTODO	17
4.1 Materiales.....	17
4.1.1 Datos	17
4.1.2 Software Utilizado	19
4.2 Método	20
4.2.1 CRISP-DM.....	20
4.2.1 Comprensión del problema	21
4.2.2 Entendimiento de los datos	21
4.2.3 Preparación de los datos.....	22
4.2.4 Modelado de datos	22
4.2.5 Evaluación.....	22
4.2.6 Despliegue.....	22

4.2.7 SPEM.....	23
4.3 Clustering.....	23
4.3.1 K-Means.....	23
4.3.2 Funcionamiento.....	24
CAPÍTULO 5.....	26
5. METODOLOGÍA.....	26
5.1 Recopilación de datos.....	26
5.1.1 Contaminantes atmosféricos.....	27
5.1.2 Variables Meteorológicas.....	27
5.2 Procesamiento previo de datos.....	27
5.3 Análisis de datos.....	28
5.3.1 Selección de datos.....	29
5.3.2 Transformación de datos.....	31
5.3.3 Filtrado de datos.....	31
5.3.4 Normalización.....	32
5.3.5 Aplicación del algoritmo de clustering.....	33
5.3.6 Evaluación.....	33
CAPÍTULO 6.....	38
6. RESULTADOS.....	38
CAPÍTULO 7.....	43
7. ESTUDIO DE CASO.....	43
7.1 Diseño y planificación.....	43
7.1.1 Objetivos y preguntas de investigación.....	43
7.2 Preparación y recolección de datos.....	44
7.2.1 Selección del contexto.....	44
7.2.2 Materiales.....	44
7.2.3 Recolección de datos.....	45
7.2.4 Ejecución.....	46
7.3 Análisis e interpretación de los datos recogidos.....	46
7.4 Resultados obtenidos.....	46
7.5 Amenazas a la validez.....	47
7.5.1 Validez de conclusión.....	47
7.5.2 Validez interna.....	47
7.5.3 Validez de constructo.....	47
7.5.2 Validez externa.....	48
CONCLUSIONES.....	49
REFERENCIAS.....	51
ANEXOS.....	55

Índice de tablas

Tabla 1.....	5
Cadena de búsqueda automática.....	5
Tabla 2. Bibliotecas e indexadores seleccionados.....	5
Tabla 3. Evaluación de calidad.....	7
Tabla 4. Formulario de extracción de datos.....	7
Tabla 5. Matriz de resultados.....	10
Tabla 6 Contaminantes atmosféricos.....	27
Tabla 7 Variables meteorológicas.....	27

Tabla 8 Índice de calidad del aire	35
Tabla 9 Calidad el aire según cada contaminante.....	35
Tabla 10 Categorización de TEMPAIRE_AV y HR_AV	36
Tabla 11 Clúster óptimo.....	42
Tabla 12 Goal-Question Metric	43
Tabla 13 Cuestionario para medir las variables con TAM	45
Tabla 14 Estadística de los resultados obtenidos en el caso de estudio.....	46
Tabla 15 Estadística descriptiva para las variables basadas en la percepción del usuario.	46

Índice de figuras

Figura 1. Relevancia de los artículos por pregunta de investigación (RQ)	11
Figura 2. Variables de clasificación en la calidad del aire.....	12
Figura 3. Enfoques de los sistemas de recomendación usados en el dominio de la recomendación	13
Figura 4. Diversos enfoques de los sistemas de recomendación usados en los diversos tipos de clasificación	14
Figura 5. Enfoques de los sistemas de recomendación usados en las diversas áreas de conocimiento.....	15
Figura 6. Fuentes de conocimiento usadas en los diferentes enfoques de los sistemas de recomendación	16
Figura 7. Ciclo de vida CRISP-DM.....	21
Figura 8 Metodología del proyecto.....	26
Figura 9 Metodología del análisis de datos.....	29
Figura 10 Gráfico de codo	34
Figura 11 Gráfico de radar con los centroides de cada clúster	39
Figura 12 KDE comparando cada clúster para determinar el tiempo óptimo.....	40

Índice de anexos

Anexo 1 Lista de artículos primarios seleccionados.....	55
Anexo 2 Clústeres resultantes.....	56
Anexo 3 Modelado de la solución en RapidMiner	57
Anexo 4 Resultados de la evaluación a cada paper con los criterios EC01 y EC02.....	58
Anexo 5 Resultados de la evaluación a cada paper con el criterio EC03	60
Anexo 6 Resultados de la evaluación a cada paper con los criterios EC04 y EC05.....	62
Anexo 7 Resultados de la evaluación a cada paper con los criterios EC06, EC07, EC08 y EC09	64
Anexo 8 Resultados de la evaluación a cada paper con los criterios EC10, EC11 y EC12	66
Anexo 9 Resultados de la evaluación a cada paper con los criterios EC13, EC14 y EC15	68

RESUMEN:

En la actualidad, debido al auge de la industrialización, se libera una gran cantidad de toxinas en el aire, representando un riesgo al ambiente y a la salud. La existencia de grandes volúmenes de datos generados por los dispositivos que monitorean estos contaminantes permite la construcción de aplicaciones, técnicas y metodologías que pueden relacionar estas variables entre sí, y con ello alcanzar una predicción de estado del aire. Por otra parte, los sistemas recomendadores están presentes en innumerables métodos brindando una ayuda sustancial en la toma de decisiones, y un impulso a la mejora en la calidad de un servicio. El presente estudio se centra en realizar una revisión sistemática de la literatura ahondando en estos temas y posteriormente desarrollar la metodología de un sistema recomendador, el cual, según los contaminantes presentes en el ambiente permita recomendar el mejor momento para realizar actividades al aire libre.

Palabras clave: Sistemas recomendadores, salud, calidad del aire, minería de datos, contaminantes atmosféricos, variables meteorológicas.

ABSTRACT:

Nowadays, due to the increase of industrialization, a large amount of toxins is released into the air, presenting a risk to the environment and health. The existence of large volumes of data generated by the devices that monitor these pollutants allows the construction of applications, techniques and methodologies that can relate these variables to each other, and thus reach a prediction on the state of the air. On the other hand, recommender systems are present in countless methods, providing substantial help in decision-making, and an impulse in improve the quality of a service. The present study focuses on conducting a systematic review of the literature delving into these issues and subsequently developing the methodology of a recommender system, which, according to the pollutants present in the environment, is able to provide suggestions on the best time to perform outdoor activities.

Keywords: Recommender system, health, air quality, air pollutants, data mining, air pollutants, weather variables.

Translated by

Pablo Loja Arévalo

