

Facultad Latinoamericana de Ciencias Sociales, FLACSO Ecuador

Departamento de Asuntos Públicos

Convocatoria 2021-2023

Tesis para obtener el título de Maestría en Estudios Urbanos con mención en Geografía
y Procesos Territoriales

Machine Learning aplicada para evaluar la satisfacción con la vida en el Centro
Histórico de Quito

Jefferson Alfredo Revelo Gutiérrez

Asesor: Pablo Cabrera Barona

Lectores: Pablo Ordoñez y Pablo Velasco

Quito, marzo de 2024

Dedicatoria

A mi amada madre, Nelly, cuya inquebrantable fe y apoyo incondicional han sido mi faro en este viaje académico. A mi querida familia: Yelitza, Diego y Anahí, Lily y Mario por su amor constante y comprensión en cada etapa de este camino.

A mi padre Walter y a mis abuelitos Jorge y Luz, cuyo legado de perseverancia y sabiduría me ha guiado a lo largo de los años.

Al docente investigador, Pablo Cabrera, cuyo valioso aporte en la FLACSO Ecuador ha sido fundamental en el desarrollo de esta tesis. Su mentoría y orientación han sido un faro de conocimiento en mi travesía académica.

Al Ingeniero Richard Aguirre, por su apoyo inquebrantable y por sentar las sólidas bases de mi trayectoria profesional y académica.

Esta tesis es el resultado de la dedicación y el amor de todas estas personas excepcionales: Kay, Lenin, Vane, Karlita, Dan, Mafer, Anita. A cada uno de ustedes, les dedico este logro con gratitud y cariño.

Índice de contenidos

Resumen.....	9
Capítulo 1. Introducción	11
1.1 Antecedentes del problema y conceptos generales	11
1.2 Justificación	12
1.3 Pregunta de investigación	14
1.4 Objetivos.....	14
1.4.1 Objetivos específicos:	14
Capítulo 2. Revisión analítica de literatura.....	15
2.1 Satisfacción con la Vida	15
2.2 Variables explicativas	16
2.2.1 Educación alcanzada y salud actual	16
2.2.2 Salud actual	17
2.2.3 Vivienda	17
2.2.4 Trabajo	18
2.2.5 Relación comunitaria	18
2.2.6 Calidad ambiental urbana.....	19
2.3 El territorio y la satisfacción con la vida	20
2.4 Inteligencia Artificial.....	21
2.4.1 Machine Learning	22
2.4.2 Algoritmos de machine learning y deep learning para evaluar sentimientos humanos	23
2.4.3 Geo IA.....	24
Capítulo 3. Metodología	27
3.1 Flujograma Metodológico.....	27
3.2 Descripción del área de estudio	27
3.3 Fuente de datos - Encuesta Multipropósito 2016 (EMP).....	29
3.4 Elección de variables	29
3.5 Análisis exploratorio de datos.....	31

3.5.1	Análisis de autocorrelación espacial	31
3.5.2	Niveles de significancia	32
3.5.3	Principio LISA	32
3.6	Aprendizaje no supervisado.....	33
3.6.1	Regionalización espacial.....	34
3.7	Aprendizaje supervisado.....	38
3.7.1	Redes Neuronales.....	39
3.7.2	Árbol de decisión	44
3.7.3	Máquinas de vector soporte (SVM)	45
3.8	Representación espacial	47
3.8.1	IDW.....	47
Capítulo 4.	Resultados	49
4.1	Estadísticas descriptivas sociodemográficas EMP	49
4.1.1	Género y edad	49
4.1.2	Nivel educativo	50
4.1.3	Provincia o país de origen	50
4.2	Estadísticas descriptivas de las variables empleadas	51
4.3	Autocorrelación espacial de la Satisfacción con la vida en el CHQ.	51
4.3.1	Puntos calientes – Puntos fríos.....	55
4.3.2	Interpretación del análisis de autocorrelación espacial	56
4.4	Resultados de análisis de regionalización de satisfacción con la vida y variables predictoras	57
4.4.1	k-means	57
4.4.2	SKATER	64
4.4.3	Hierarchical clustering	69
4.5	Resultados de regresión con aprendizaje supervisado	73
4.5.1	Neural Network.....	73
4.5.2	Resultados del Árbol de decisión.....	81
4.5.3	Resultados Support Vector Machine.....	85

Capítulo 5. Discusión y conclusiones	87
5.1 Discusión	87
5.1.1 Escalabilidad	87
5.1.2 Análisis exploratorios.....	88
5.1.3 Aprendizaje no supervisado	88
5.1.4 Aprendizaje supervisado	90
5.1.5 Comparación de Resultados de la investigación frente a otros estudios cuantitativos sobre satisfacción con la calidad de vida.	91
5.1.6 Implicación de resultados para la planificación urbana y las políticas públicas. 93	
5.1.7 Discusión sobre limitaciones del estudio, y sobre ideas futuras de investigación sobre el tema tratado en la Tesis	94
5.2 Conclusiones.....	95
5.2.1 Objetivos, principales resultados y conclusiones.....	98
Referencias.....	103

Lista de ilustraciones

Figuras

Figura 2.1. Diferencia entre IA – ML – DL	22
Figura 3.1. Operacionalización de la metodología.....	27
Figura 3.2 Gráfica método del codo.....	35
Figura 3.3 Procedimiento K- Means	35
Figura 3.4 Procedimiento SKATER.....	37
Figura 3.5. Red neuronal básica (RNA).....	41
Figura 3.6. Esquema de funcionamiento de redes Neuronales.....	42
Figura 3.7. Estructura árbol de decisión.....	44
Figura 3.8. Representación gráfica SVM, función principal, hiperplano y errores.....	46
Figura 3.9. Procesos del método SVM, función principal, hiperplano y errores.	46
Figura 3.10. Representación visual IDW	48
Figura 3.11. Escala de semaforización IDW	48
Figura 4.1. Grupos de edad Encuesta Multipropósito	49
Figura 4.2. Nivel educativo alcanzado	50
Figura 4.3. Esquema de <i>Moran's I</i> (Valores altos y Bajos)	53
Figura 4.4 Gráfica de contigüidad tipo Reina Satisfacción con la vida CHQ	57
Figura 4.5. Gráfica de codo para diferentes pesos y k.	58
Figura 4.6. Gráfica de codo para <i>k-means</i> (todas las variables).....	63
Figura 4.7. Gráficas de codo para skater (1 y 8 variables)	66
Figura 4.8. Árbol de expansión mínima SKATER	66
Figura 4.9. Gráficas de codo para Skater (1 y 8 variables)	70
Figura 4.10. Árbol de decisión.....	83

Mapas

Mapa 3.1. Mapa de ubicación del área de estudio.....	28
Mapa 4.1. Clústeres de <i>LISA</i>	52
Mapa 4.2. Significancia de <i>LISA</i>	53
Mapa 4.3. Autocorrelación espacial <i>Moran's I</i> - <i>LISA</i>	54
Mapa 4.4. Puntos calientes – puntos fríos de la satisfacción con la vida en el CHQ	55
Mapa 4.5. Clústeres <i>k – means</i> Satisfacción con la vida en el CHQ.....	62
Mapa 4.6. Clusters <i>k – means</i> Satisfacción con la vida (todas las variables) en el CHQ.....	64
Mapa 4.7. Clústers SKATER Satisfacción con la vida en el CHQ	68
Mapa 4.8. Clústers Skater (8 variables) Satisfacción con la vida en el CHQ	69

Mapa 4.9. Clústers jerárquico Satisfacción con la vida en el CHQ	72
Mapa 4.10. Clusters jerárquico (8 variables) Satisfacción con la vida en el CHQ	73
Mapa 4.11. Satisfacción con la vida modelo lineal redes neuronales	79
Mapa 4.12. Satisfacción con la vida modelo logístico sigmoide redes neuronales.....	80
Mapa 4.13. Satisfacción con la vida modelo tangente inversa redes neuronales	81
Mapa 4.14. Satisfacción con la vida modelo tangente inversa redes neuronales	84
Mapa 4.15. Satisfacción con la vida modelo support vector machine	86

Tablas

Tabla 3.1. Matriz comparativa de la reducción de variables.....	30
Tabla 3.2. Variables empleadas para el análisis de ML	30
Tabla 3.3. Valores de interés a diferentes niveles de significancia	32
Tabla 3.4. Términos similares entre modelos estadísticos y redes neuronales	40
Tabla 3.5. Parámetros estadísticos resultantes	40
Tabla 3.6. Funciones de salida comunes para RNA.....	43
Tabla 4.1. Provincia o país de origen.....	50
Tabla 4.2. Estadísticas descriptivas variables	51
Tabla 4.3. Optimización del modelo K-Means – datos para la gráfica del codo.....	58
Tabla 4.4. Parámetros y modelos de clasificación k-means.....	59
Tabla 4.5. Modelo <i>k-means</i> (todas las variables) – datos para la gráfica del codo	63
Tabla 4.6. Modelo Skater (1 y 8 variables) – datos para la gráfica del codo	65
Tabla 4.7. Modelo <i>clúster</i> jerárquico (1 y 8 variables) – datos para la gráfica del codo	70
Tabla 4.8. Métricas de evaluación rendimiento estadístico de los modelos de redes neuronales	74
Tabla 4.9 Resultados del diseño de redes neuronales	75
Tabla 4.10. Importancia relativa modelos redes neuronales	77
Tabla 4.11. Métricas de evaluación rendimiento estadístico del modelo árbol de decisión.....	81
Tabla 4.12. Importancia relativa modelo árbol de decisión	82
Tabla 4.13. Métricas de evaluación estadística del modelo support vector machine	85
Tabla 5.1. Matriz resumen de objetivos, resultados y conclusiones	98

Declaración de cesión de derechos de publicación de la tesis

Yo, Jefferson Alfredo Revelo Gutiérrez, autor de la tesis titulada “*Machine Learning* aplicada para evaluar la satisfacción con la vida en el Centro Histórico de Quito”, declaro que la obra es de mi exclusiva autoría, que la he elaborado para obtener el título de maestría en Estudios Urbanos con mención en Geografía y Procesos Territoriales por la Facultad Latinoamericana de Ciencias Sociales, FLACSO Ecuador.

Cedo a la FLACSO Ecuador los derechos exclusivos de reproducción, comunicación pública distribución y divulgación, bajo la licencia Creative Commons 3.0 Ecuador (CC BY-NC-ND 3.0 EC), para que esta universidad publique en su repositorio, siempre y cuando el objetivo no sea obtener beneficio económico

Quito, marzo de 2024



Jefferson Alfredo Revelo Gutiérrez

Resumen

La tesis se centra en la aplicación de *machine learning* para evaluar la satisfacción con la vida en el Centro Histórico de Quito. El objetivo principal es utilizar técnicas de ML para determinar qué factores demográficos, sociales y económicos influyen significativamente en la satisfacción con la vida de los habitantes de esta zona. A diferencia de los enfoques convencionales en ciencias psicológicas, la tesis busca complejizar el análisis mediante métodos cuantitativos y de analítica socioespaciales. La investigación se basa en datos recopilados a partir de la Encuesta multipropósito del Instituto de la Ciudad de Quito 2016. Para analizar la información, se emplearon modelos de autocorrelación espacial, LISA y la identificación de puntos calientes y puntos fríos. Además, se implementaron tres modelos de aprendizaje no supervisado (*k-means*, *SKATER*, cluster jerárquico) y tres modelos de aprendizaje supervisado (Redes neuronales, Árbol de decisiones, Máquinas de vector soporte).

La comparación, análisis y discusión de estos modelos revelaron que los más eficaces de forma estadística y espacial fueron los modelos *k-means* y árboles de decisiones. Estos modelos destacaron por su capacidad para identificar patrones y relaciones en los datos, proporcionando una visión más clara de los factores que impactan la satisfacción con la vida en el Centro Histórico de Quito. Los resultados se visualizaron mediante mapas temáticos multiescala, representando la distribución espacial de las variables subjetivas analizadas. Estos mapas no solo ofrecen una representación visual de la información recopilada, sino que también proporcionan datos valiosos para la toma de decisiones y orientación de la planificación urbana. La inclusión de modelos cuantitativos y socioespaciales en la evaluación de la satisfacción con la vida en esta área urbana contribuye a una comprensión más integral y basada en datos, ofreciendo herramientas útiles para mejorar la toma de decisiones y optimizar la planificación urbana en el Centro Histórico de Quito.

Es importante destacar que la evaluación de la satisfacción con la vida, dada su naturaleza, ha sido tradicionalmente abordada desde un enfoque cualitativo, al igual que otras variables vinculadas a la percepción del bienestar subjetivo. Frente a este escenario, y con el propósito de aportar a la investigación, esta tesis se ha centrado en la aplicación de estrategias de complementariedad, específicamente mediante la integración de enfoques cuantitativos y espaciales. Esta elección se ha realizado de manera coherente con las bases de datos y la información disponible.

Al seguir esta premisa, se garantiza que los resultados obtenidos se conviertan en herramientas novedosas, aplicables en el ámbito de la planificación urbana y la formulación de políticas públicas. Este enfoque innovador no solo enriquece el análisis de la satisfacción con la vida,

sino que también contribuye a la creación de recursos valiosos para la toma de decisiones en contextos urbanos.

Es evidente que la inclusión de elementos cuantitativos y espaciales no solo fortalece la validez de los hallazgos, sino que también amplía el impacto potencial de los resultados. En consecuencia, se destaca la importancia de los análisis de la satisfacción con la vida como indicadores sólidos de la realidad experimentada por los habitantes de los centros históricos, tanto a nivel local como regional. Este enfoque holístico no solo proporciona una comprensión más completa de la situación, sino que también abre nuevas posibilidades para abordar desafíos y mejorar la calidad de vida en ciudades urbanas.

Palabras clave: *Machine learning*, Satisfacción con la vida, Centro Histórico de Quito, redes neuronales, regionalización espacial.

Capítulo 1. Introducción

1.1 Antecedentes del problema y conceptos generales

En 1978 la ciudad de Quito fue declarada Patrimonio de la Humanidad por la UNESCO, la razón primordial por la cual la fue merecedora de este reconocimiento, se debió a que la ciudad posee uno de los tesoros arquitectónicos y artísticos mejor conservados y más vivos del planeta, como es el caso del Centro Histórico de Quito (CHQ). Este Centro Histórico ha sido considerado como el más importante de la región Latinoamericana y se ha convertido, en un símbolo de la ciudad y el país por su relevancia patrimonial al ser el sitio de encuentro de diversas culturas y expresiones artísticas desde hace 4 siglos. Sin embargo, en las últimas dos décadas, en casi todas las ciudades latinoamericanas ha existido un proceso de deterioro y abandono de este tipo de áreas, que en su momento fueron las principales centralidades de desarrollo económico y social de las ciudades. Para Carrión (2001), el deterioro y abandono de las áreas históricas se lo atribuye a diferentes factores (sociales, económicos, naturales, etc.), que dentro de un contexto de modernización y globalización modificaron de forma negativa la percepción que tienen los habitantes de estas áreas (Carrión 2001).

Los centros históricos son áreas que han aportado al desarrollo social, físico, económico y cultural de las ciudades, especialmente son los contenedores de la identidad de la sociedad que lo habita. Sin embargo, en los últimos años ha sido un verdadero reto de la planificación territorial y dinámicas urbanísticas, que su condición como principal centralidad de la ciudad, se mantenga y prospere (Carrión 2001). Ante esta situación hay que entender la problemática que enfrenta el Centro Histórico de Quito desde diferentes ámbitos, como por ejemplo el de la movilidad, ya que, al encontrarse sobre un valle angosto y topográficamente heterogéneo, hicieron que su diseño vial se implante de forma rectangular y angosta, y se adapte a las necesidades de una población de hace más de 4 siglos. Con el pasar de los años este diseño se mantuvo y convirtieron al CHQ un entorno aislado y desconectado con el sistema de centralidades de la ciudad debido a sus condiciones propias de accesibilidad, ya que su estructura urbana colonial - que se conserva mayormente inalterada- no posee la capacidad de asumir el flujo vehicular de la ciudad, lo que genera problemas de contaminación ambiental y congestión (MIDUVI 2016).

Todas estas problemáticas se han visto traducidas en un constante proceso de abandono del CHQ, ya que, en las últimas tres décadas se ha venido presentando un acelerado proceso de despoblamiento. Las cifras demográficas muestran que para la década de 1990 la población del CHQ era de alrededor de 58.300 personas, en el censo del año 2010 se evidenció que la población había decaído a 40.587 habitantes, y la proyección censal para el año 2022 indica que la población del CHQ alcanzaría un total de 26.727 personas, lo que equivale al 45,84% del total de personas que habitaban el CHQ entre los años 1990 y 1999 (MIDUVI 2016). Ante esta situación es necesario investigar qué factores o razones existen detrás de todas estas problemáticas mencionadas y correlacionarlas de forma innovadora con la percepción de la satisfacción con la vida que tienen los habitantes del CHQ, y que posiblemente se haya deteriorado paralelamente con su entorno de habitabilidad.

La satisfacción con la vida es un término estrechamente relacionado con los conceptos bienestar y felicidad. Alarcón (2006) afirma que la semántica utilizada entre lo que se denomina felicidad, bienestar y satisfacción con la vida, desaparece al momento de proponer herramientas de medición o una escala de valoración de esta serie de conceptos concepto subjetivos. Sin embargo, tanto Alarcón (2006) como Lyubomirsky (2008) coinciden en que los términos abarcan un universo de otras variables y constructos psicológicos que inciden en la percepción con la satisfacción con la vida.

1.2 Justificación

En función de lo planteado, se puede decir que el CHQ, por sus características y grado de complejidad, se ha convertido en un caso de estudio e interés para las gestiones municipales y ciudadanía en general, porque hasta el momento, su incursión al sistema funcional de la ciudad no ha sido clara y ha representado un verdadero reto para la planificación urbana y el ordenamiento territorial. De este modo, los fenómenos que ocurren en el CHQ deben ser tratados como un sistema complejo y dinámico. El presente estudio busca entender el por qué en la ocurrencia de este proceso de abandono de sus residentes, que han hecho del CHQ un espacio aislado dentro del sistema de centralidades, que hace tan solo tres décadas era considerado como el principal eje articulador de la ciudad.

La tendencia que se encuentra actualmente posicionando en el análisis espacial es el *Spatial Big Data*, que demanda de métodos cuantitativos relativamente nuevos e

innovadores que permiten que fenómenos como la satisfacción con la vida sean comprendidos a través de una serie de dimensiones y variables predictoras, a través de la aplicación de algoritmos computacionales conocidos como *Machine Learning (ML)* (Evans et al. 2010).

La inteligencia artificial tiene un amplio recorrido, pero en los últimos 15 años ha tomado más relevancia y aplicabilidad con el desarrollo del aprendizaje automático, las redes neuronales y el aprendizaje profundo, de tal forma que autores como Chen et al. (2021) indican que este rápido desarrollo ha repercutido en otras ciencias, como el caso de las Ciencias Sociales. De forma particular el ML o aprendizaje automático es una técnica computacional que se encuentra dentro de inteligencia artificial en la que convergen y se intersecan varias disciplinas como las matemáticas, estadística, informáticas y de neurociencia y que usan algoritmos para extraer información y conocimiento de los datos masivos generando una infinidad de aplicaciones que alcanzan a ciencias sociales, sociología, economía e inclusive ciencias políticas (Di Franco and Santurro 2021). De esta forma, gracias a las capacidades computacionales la característica principal del ML es que esta serie procesos y algoritmos, tienen la capacidad de aprender y mejorarse por sí mismos a través de diferentes técnicas o modelos, permitiendo analizar grandes cantidades de datos en tan solo unos segundos o minutos encontrando una serie de patrones o tendencias que permiten modelar de forma acertada una serie de fenómenos sociales (Gualda 2022).

Estas aproximaciones mixtas (cualitativas y cuantitativas) hacen que un fenómeno que tradicionalmente ha sido estudiado desde el ámbito de la psicología pueda reinventarse y complementarse con enfoques interdisciplinarios. Las ciencias sociales computacionales con apoyo de análisis espacial pueden generar resultados que son herramientas útiles para los tomadores de decisiones. Adicionalmente, información cuantitativa-espacial puede aportar en la transformación del imaginario social que se tiene actualmente de los centros históricos, cuyo proceso de abandono y decadencia ha llamado la atención de las autoridades gubernamentales y municipales que han demostrado el interés de incluir en las agendas de desarrollo territorial la regeneración y la repotenciación de estas áreas con el objetivo de incentivar su rehabilitabilidad (MIDUVI 2016).

De este modo, resulta importante entender que la satisfacción con la vida es un proceso dinámico que depende de una serie de procesos o fenómenos físicos, ambientales,

culturales y económicos que son parte de la cotidianeidad de las personas. Sin embargo, resulta aún más importante considerar a la satisfacción con la vida como un indicador de la gestión y planificación de una ciudad o territorio. Por ejemplo, países y ciudades desarrollados, son los que tienen mayores índices de felicidad y satisfacción con la vida y de igual forma son aquellos que registran los mayores índices de desarrollo humano en variables relacionadas a la salud, la educación y los ingresos económicos (Gordón, Murillo, y Hernández 2018). De esta forma, analizar la satisfacción con la vida desde una perspectiva integral, revelaría los aciertos y falencias que tiene un determinado territorio y permitiría plantear una serie de ejes y planes de acción con objetivos específicos y metas alcanzables para mejorar la condición de vida de sus habitantes.

1.3 Pregunta de investigación

Con base a lo expuesto anteriormente, la presente Tesis plantea la siguiente pregunta de investigación.

- ¿Cómo se puede explicar la satisfacción con la vida de los habitantes del Centro Histórico de Quito a través de un enfoque de *Machine Learning* (ML)?

1.4 Objetivos

Respecto a la pregunta de investigación planteada el objetivo general de la Tesis es el siguiente:

- A través de ML, determinar qué factores demográficos, sociales y económicos son los que influyen más en la satisfacción con la vida de los habitantes del Centro Histórico de Quito

1.4.1 Objetivos específicos:

- Comparar diferentes modelos de ML para evaluar la satisfacción con la vida de los habitantes del Centro Histórico de Quito
- Expresar espacialmente los factores más influyentes en la satisfacción con la vida de los habitantes del Centro Histórico de Quito
- Discutir la potencialidad de ML para analizar datos cualitativos, en el marco de las Ciencias Sociales y Estudios de la ciudad.

Capítulo 2. Revisión analítica de literatura

2.1 Satisfacción con la Vida

Para comprender qué es la satisfacción con la vida, primero se debe comprender lo que significa la felicidad. Este concepto ha sido estudiado por la psicología positiva, ya que la felicidad siempre ha sido conceptualizada como una meta de vida para las personas, tanto así que en los últimos años, las investigaciones acerca de la felicidad y su relación con diversas variables demográficas, socioculturales y psicológicas han aumentado sustancialmente desde el año 2000, según datos de la “Base Mundial de Datos sobre Felicidad” y el número de artículos en constante crecimiento publicados en el *Journal of Happiness Studies*. A pesar de que los estudios de la felicidad han experimentado un auge en las últimas dos décadas, el concepto ha sido estudiado desde la filosofía desde hace ya varios años, por Sócrates, Platón, Aristóteles y Epicuro, quienes comprendieron y definieron la felicidad como: “El regocijo, el placer o la satisfacción que se experimenta al alcanzar un bien deseado” (Caycho Rodríguez 2010).

Alarcón (2009) muestra diversas concepciones de la felicidad, como bienestar subjetivo y como experiencia óptima y plantea una definición propia y original que la define como “un estado de satisfacción, más o menos duradera, que experimenta subjetivamente el individuo en posesión de un bien deseado”, y basado en la filosofía griega concibe la felicidad como un sinónimo de satisfacción con la vida (Barragán 2013).

Tkach and Lyubomirsky (2006) indican que el bienestar y la felicidad se hallan significativamente relacionados y ponen en manifiesto desde un enfoque explicativo, que los rasgos de personalidad, las relaciones sociales, el estado de ánimo, la vida con sentido y el nivel de satisfacción general anticipan en gran medida el desarrollo de la felicidad.

Para (Diener 2000) el término “felicidad” se utiliza de manera indiscriminada con el de bienestar subjetivo, entendido como un juicio cognitivo global resultado de la combinación entre la satisfacción de la vida y el balance generado por la frecuencia de afecto positivo y negativo

La felicidad conceptualmente ha dejado de ser un ideal de la sociología, filosofía y psicología, para convertirse en un constructo multidimensional en las diferentes etapas de la vida. Visto

desde esta perspectiva, las tres definiciones antes mencionadas de Alarcón (2009), Tkach and Lyubomirsky (2006) y Diener (2000), que son aparentemente diferentes, en realidad comparten grandes similitudes entre sí. Esta similitud consiste en que, al momento de comparar las herramientas de medición operacionales o categorías que proponen estos tres autores, son prácticamente equivalentes y por ende las diferencias semánticas de lo que se denomina bienestar, felicidad y satisfacción con la vida, desaparecen.

De ahí que la terminología en inglés *psychological well-being* (bienestar psicológico), *well-being* (bienestar), *subjective well-being* (bienestar subjetivo) y *satisfaction with life* (satisfacción con la vida) se utilicen como términos semánticos iguales o sinónimos manteniendo su equivalencia conceptual para la generación de resultados en investigaciones científicas (Alarcón 2006).

2.2 Variables explicativas

Una vez definida a la variable dependiente que es la satisfacción con la calidad de vida, es conveniente sustentar de forma teórica la elección de cada una de las variables independientes que serán incluidas en los análisis posteriores de esta Tesis.

En este apartado es necesario mencionar al trabajo realizado por Rodríguez (2023) el cual presenta un modelo que emplea variables de los distintos dominios con la vida como, por ejemplo: salud actual, la vivienda y el trabajo. Dichos resultados obtienen un valor de coeficiente de determinación de un R^2 hasta un 50%. En dicho trabajo también se señala que existe una importante evidencia empírica que pretende entender la influencia o causalidad entre una gran cantidad de variables subjetivas y objetivas con la Satisfacción con la Vida. El autor citado indica que estas variables independientes son significativas para explicar el comportamiento de la satisfacción con la vida y toma como referencia literatura que respalda su discusión ya que llegan en sus trabajos discuten sobre como las diversas esferas o dimensiones que influyen en la satisfacción con la vida presentan variaciones en su importancia, dependiendo de las condiciones y experiencias objetivas y subjetivas que los individuos experimentan en sus entornos (M. Rojas 2012; Bialowolski y Weziak-Bialowolska 2021).

2.2.1 Educación alcanzada y salud actual

Bustamante-Ubilla et al. (2020) indican que los habitantes urbanos al presentar un mayor nivel de educación y un mayor nivel de acceso a la salud presentan un mejor nivel de satisfacción con la vida, a diferencias de los habitantes del área rural. Adicionalmente se menciona que, el nivel de instrucción alcanzado es un determinante significativo que podría explicar un mayor nivel de satisfacción en la calidad de vida para los habitantes adultos mayores del área urbana.

2.2.2 Salud actual

Generalmente la medición de la salud se ha realizado a través de indicadores o criterios objetivos y físicos, que justamente son las pruebas de laboratorio, imagen, entre otras. Si bien resulta importante cuantificar el estado de salud de esta forma, también lo es la percepción que tienen las personas sobre su estado de salud. Por esta razón se han implementado ciertos criterios subjetivos basados en la percepción individual del estado de salud y de este modo tener una evaluación más inclusiva e integral (Moreta-Herrera et al. 2018) que permita entender la satisfacción con la calidad de vida.

2.2.3 Vivienda

La satisfacción con la vida se puede entender desde dos diferentes ámbitos de estudio. Uno hace referencia a las condiciones consideradas necesarias para una buena vida, y otro a la práctica adecuada de la dignidad y el buen vivir. Cuando se aborda desde el primer ámbito se puede hacer referencia a la calidad o percepción del estado de la vivienda, que es un indicador que puede ser objetivo y subjetivo en función del tipo de estudio que se está realizando. En este caso, hablar de vivienda adecuada en Ecuador y el CHQ implica que cumpla con las siguientes premisas (Secretaría técnica Ecuador crece sin desnutrición infantil, s. f.)

- Asequibilidad: hace referencia a la vivienda y el costo de esta, y consiste en que su costo no debe poner en peligro los otros derechos humanos u otras necesidades básicas.
- Habitabilidad: la vivienda debe proporcionar a sus ocupantes la seguridad y el espacio suficiente para vivir y dar protección contra elementos naturales como calor, frío, lluvia y otras amenazas de salud y riesgos naturales.
- Ubicación: tanto en zonas rurales como urbanas (como el caso del CHQ) la vivienda adecuada se debe encontrar en sitios donde se permita el fácil acceso a los servicios de salud, educación, transporte y empleo. Adicionalmente la vivienda debe ubicarse en sitios alejados de lugares peligrosos y de fuentes de contaminación.
- Entorno cultural: la vivienda debe ser un espacio de expresión de identidad y diversidad cultural de las personas quienes forman parte de ese entorno. De forma particular el CHQ es un símbolo cultural patrimonial que revela la importancia de las interacciones sociales de sus integrantes con el territorio, en el cual desarrollan sus diferentes actitudes, valores, normas y por ende su percepción con la satisfacción con la vida.
- Accesibilidad: Todas las personas deben tener acceso a una vivienda adecuada priorizando la atención a grupos desfavorecidos o víctimas de desastres naturales (grupos considerados en estado de vulnerabilidad)

2.2.4 Trabajo

En el año 2017 la Organización para la Cooperación el Desarrollo Económico (OCDE) determinaron que existe una correlación positiva significativa entre la satisfacción con la vida (felicidad) y la satisfacción laboral (trabajo). Navarro Astor (2008) menciona que, para definir la satisfacción en el trabajo se pueden emplear dos propuestas teóricas. La primera indica que es “una actitud generalizada hacia el trabajo” y la segunda “como un sentimiento o respuesta afectiva hacia las condiciones de trabajo”. Sin embargo, otros autores como Locke & Dunnet (1976) presentaron una definición clásica de este concepto indicando que la satisfacción laboral es “un estado emocional positivo o de placer producto de la percepción del trabajo o de las experiencias obtenidas en su desarrollo”.

Históricamente se ha creído que el desarrollo económico, o la remuneración económica alta, es un indicador del bienestar subjetivo o satisfacción con la vida de la persona. Sin embargo, estudios como los de Alarcón (2006) han refutado este “mito” e indican que la satisfacción con una persona no solo está relacionada con factores propios del ambiente laboral, sino más bien son parte de la integridad de la persona y son compartidos con los aspectos cognitivos y socioemocionales de las diferentes etapas de la vida de una persona.

Bajo este esquema, estudios como los de Gamero Maldonado (2013) y Navarro Astor (2008) señalan que existe una relación positiva alta entre la satisfacción la vida y la satisfacción laboral en sus correspondientes investigaciones, esto significa que el incremento o disminución en los niveles de una variable se relaciona directamente en los niveles de otra (directamente proporcionales). Sin embargo, a pesar de que una variable se correlacione con otra, esto no implica que la satisfacción o percepción laboral explique en gran o pequeña medida el nivel de satisfacción con la vida que posea una determinada persona, y ante esta situación la presente Tesis pretende incluir y poner a prueba a esta variable como parte de las potenciales variables predictoras de los diferentes modelos de análisis.

2.2.5 Relación comunitaria

Las relaciones comunitarias se refieren a las distintas formas que emplean las personas o habitantes de un espacio para asegurar y mantener una relación mutuamente beneficiosa en la comunidad que reside. Autores como Martínez et al. (2007) indican que la percepción de esta variable tiene implicaciones para el funcionamiento psicológico y social, en este sentido se menciona que las personas más satisfechas con la vida tienen una mayor percepción de apoyo social y una mejor relación comunitaria, mientras que una baja percepción de satisfacción con la vida se encuentra relacionado con diversas formas de aislamiento y rechazo social.

De forma lógica, el apoyo que se percibe de una relación comunitaria es uno de los recursos que disponen las personas para hallar seguridad y generar una zona de confort ya que en este ambiente se puede desarrollar el bienestar en todos los aspectos de la vida. Las relaciones comunitarias son una herramienta de ajuste al desarrollo personal ya permiten a las personas relacionarse con su entorno y específicamente se hace referencia al conjunto de aportaciones de tipo material, emocional o de compañía que la persona percibe o recibe de los diferentes miembros que conforman su red social. Cabe destacar que las relaciones comunitarias cubren un amplio rango de relaciones sociales que se encuentran más allá de la red primaria que puede ser constituida por amigos cercanos y familiares, en este sentido las relaciones comunitarias pretenden generar una red o círculo de personas que no se conocen necesariamente (lazos débiles) pero se reconocen a sí mismos como parte de una misma comunidad. (Gracia y Herrero 2006).

La percepción de una buena relación comunitaria refleja el sentir de pertenencia además de los sentimientos de integración y de ser parte de una comunidad, esto principalmente porque se genera un sentimiento de ser un miembro participativo y activo que construye comunidad. Este proceso al tratarse de carácter voluntario vincula y estrecha las relaciones sociales de un territorio en específico favoreciendo el apoyo social y por ende la percepción con la calidad de vida.

2.2.6 Calidad ambiental urbana

El significado y alcance del concepto de satisfacción con la calidad ambiental urbana se orienta a la comprensión y exploración de aquellos aspectos que conforman el hábitat físico donde las personas desarrollan sus actividades cotidianas básicas como: alimentarse, trabajar, reposar, etc. (A. Rojas 2011). En este sentido existen diferentes aspectos que pueden conformar la calidad ambiental urbana que son:

- Aspectos naturales – físicos: hace referencia a todas aquellas variables de carácter físico – natural (clima, temperatura, áreas de protección ambiental, topografía, geomorfología, calidad de agua - suelo – aire, factores de riesgo, entre otros) que por sus efectos inciden en el bienestar y la seguridad de sus habitantes y que por ende se relacionan con la calidad ambiental de una ciudad.
- Aspectos urbano – arquitectónicos: son aquellas variables relacionadas con la configuración físico espacial de la ciudad, en el que se incluyen parámetros de análisis como el mobiliario urbano, paisaje, morfología urbanística (funcionalidad y movilidad). Este aspecto toma relevancia al momento de hablar del CHQ ya que es un espacio de conservación cultural y arquitectónico por ende su análisis es fundamental porque su

ausencia, deterioro o deficiencia pueden incidir negativamente en la calidad y percepción ambiental de un territorio específico.

- Aspectos socioculturales: en este aspecto se incluyen variables como: expresiones socioculturales, organización y participación comunitaria, preservación de valores, tradiciones históricas y todos aquellos elementos que permitan comprender la identidad organizativa de un territorio (A. Rojas 2011).

Todas estos aspectos y condiciones del ambiente, tanto físicos, urbanos -arquitectónicos y socioculturales, inciden en la posibilidad de satisfacción de las necesidades básicas y expectativas que tengan una persona respecto al territorio que habita, por lo tanto, también incide directamente en la satisfacción con la calidad de vida (López 2007).

2.3 El territorio y la satisfacción con la vida

En los últimos años en el ámbito de los estudios urbanos y las ciencias sociales, se ha adoptado el concepto de *well-being* para referirse al “bienestar” individual y subjetivo, en cual abarca varios aspectos de la vida personal y social de los habitantes residentes de un espacio geográfico concreto. En este sentido ambos intereses investigativos se conjugan para entender el espacio geográfico, a escala local y su relación con el bienestar subjetivo de cada uno de sus habitantes generando de este modo una serie de indicadores territoriales que de forma eficiente brindan una perspectiva metodológica y de resultados con la que se puede “cuantificar/medir” la satisfacción con la calidad de vida entendiéndose así que, el análisis de esta clase de fenómenos pueden traspasar la barrera del análisis estadístico descriptivo individual o grupal y que más bien pueden generarse modelos inferenciales que a su vez pueden ser representados de manera espacial.

Históricamente ha existido un interés por parte de las instituciones locales o estatales en entender la calidad de vida de sus ciudadanos, ya que se encuentra estrechamente relacionado a un diagnóstico del desarrollo y consolidación de una población en un territorio específico, para lo cual se han desarrollado una serie de índices como: el Índice de Calidad de Vida, Índice de Desarrollo Humano (PNUD), Índice de Desarrollo Socioeconómico, entre otros.

Para la generación de estos índices se utiliza una serie indicadores de carácter netamente cuantitativo explícito per se (tasas, ingresos económicos, acceso a servicios, etc.). Sin embargo, existe una diversa cantidad factores que pueden influir en la calidad de vida de un territorio y por ende con la percepción de la satisfacción con la vida, que justamente podría corresponder a variables netamente subjetivas en función a la realidad que atraviesa cada persona y su ubicación espaciotemporal.

Este carácter de subjetividad ha ocasionado que en muchas ocasiones estas variables sean descartadas de manera frecuente como variable de interés para la medición del bienestar y la calidad de vida, por lo cual se puede decir que este espectro de factores influyentes que no han sido utilizados para esta clase de análisis han ocasionado la presencia de vacío en el análisis de la satisfacción con la vida y por ende se ha desaprovechado la potencialidad de esta información para la generación de modelos de análisis territorial (socioespacial) y su posterior aporte a la generación de política pública.

Por tal razón estudios como los de Smith (1973) Gómez et al. (2001) y Lucero et al. (2007) destacan que la perspectiva espacial puede ser útil e influyente a la hora de complementar la valoración integrada de la medición del bienestar subjetivo. Y los conceptos de territorio / espacio geográfico y satisfacción con la calidad de vida son conceptos que aparecen cada vez más afines entre sí y requieren ser enlazados a los índices e indicadores que tradicionalmente se manejan. Tanto así que autores como Leva (2005) y Pacione (2003) mencionan que, a una mayor calidad territorial de una localidad, mayor es el bienestar o calidad de vida de los residentes en ella.

El territorio se construye por diferentes relaciones sociales, económicas, ambientales, entre otras y es necesario mencionar que también está relacionado fuertemente a la temporalidad y acciones que ofrecieron en su momento distintas posibilidades o condiciones particulares. Sin embargo, el territorio puede ser tratado como un factor de la satisfacción con la vida, pero no se trata de considerarlo un determinante sino más bien un condicionante que oferta posibilidades tanto ambientales (físico – naturales) como sociales, y que ejerce un papel activo dentro de la estructura social (Lucero et al. 2007).

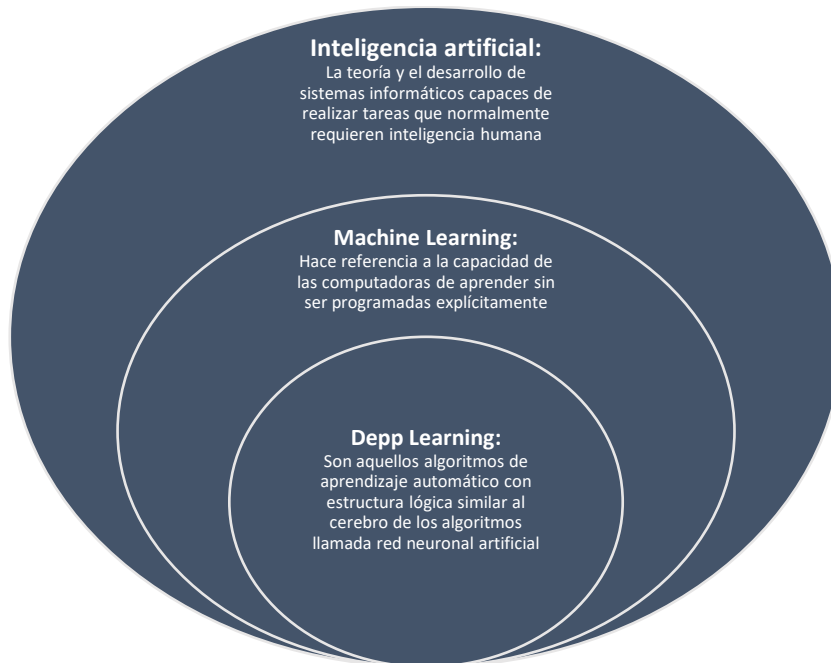
2.4 Inteligencia Artificial

Existen diferentes definiciones de lo que significa la inteligencia artificial (IA), en palabras simples se podría entender como la habilidad de los computadores para hacer actividades que normalmente requieren de la inteligencia humana. En este sentido, una definición más formal podría ser que la IA es la capacidad computacional que tienen las máquinas para usar algoritmos, aprender de los datos y utilizar lo aprendido en la toma de decisiones tal como lo harían los seres humanos (Rouhiainen 2018). Sin embargo, la IA a diferencia de las personas tiene la capacidad de analizar grandes cantidades de información al mismo tiempo y no se ve afectada por las limitaciones humanas, como descansar o reprocesar un conjunto de datos de forma analógica varias veces.

La IA es un macro concepto que en tan solo dos décadas se ha popularizado a nivel global, a este campo se le asocia conceptos como *big data*, inteligencia artificial, *machine learning* y *deep learning*. A pesar de que todas estas palabras comparten semejanzas en sus objetivos no

implica que conceptualmente signifiquen lo mismo. Por esto es necesario entender sus principales diferencias conceptuales para entender sus relaciones y potencialidades específicas, para lo cual se utiliza la siguiente figura:

Figura 2.1. Diferencia entre IA – ML – DL



Fuente: Adaptado por el autor con base a información de Wolfewicz (2023)

De esta forma se puede mencionar que el *deep learning*, es un subconjunto especializado de *machine learning* que, a su vez es un subconjunto de la inteligencia artificial.

2.4.1 Machine Learning

El *machine learning* o aprendizaje de máquinas es una disciplina científica o un subconjunto de técnicas pertenecientes al ámbito de la inteligencia artificial. Su objetivo es el de enseñar a las computadoras que aprendan de un gran conjunto de datos y que tengan la habilidad de encontrar patrones y relaciones que no son identificables a simple vista (A. González 2019).

Esta clase de algoritmos tienen la capacidad de identificar patrones complejos dentro de un conjunto de datos que se puede componer de millones de registros (*big data*). En este sentido, esta gran cantidad de información resultaría imposible de analizar por una persona para sacar conclusiones y mucho menos para hacer predicciones. Ante esta situación, se requiere de un respaldo computacional que sea capaz de revisar, procesar, almacenar información, cientos o inclusive miles de veces (TIBCO 2023b).

El *machine learning* se lo podría considerar como el punto intermedio en el que convergen la informática y la estadística porque se utilizan algoritmos para realizar una tarea específica sin

ser programados explícitamente. El *machine learning* se compone de diferentes modelos y utiliza varias técnicas algorítmicas en función del tipo de datos y el resultado que se espera. De este modo el aprendizaje de máquina se clasifica en cuatro tipos que son: supervisado, no supervisado, semisupervisado o de refuerzo (Wolfewicz 2023).

2.4.1.1 Aprendizaje supervisado

Para este tipo de aprendizaje toma relevancia los datos de carácter etiquetado, ya que el aprendizaje supervisado hace uso de una relación conocida entre la entrada y la salida de estos. El algoritmo entiende y aprende de aquellas “respuestas correctas” en los datos de entrenamiento y usa los conocimientos para hacer predicciones cuando se le ingrese un nuevo conjunto de datos. Este tipo de aprendizaje puede clasificarse en dos que son: la regresión y la clasificación (Wolfewicz 2022).

2.4.1.2 Aprendizaje no supervisado

A diferencia del aprendizaje supervisado, el no supervisado no requiere de datos previamente etiquetados. Su uso se enfoca a problemas en los que no se tiene poca o ninguna idea de cómo deberían ser los resultados (análisis exploratorio), por esta razón se percibe como una técnica muy poca intuitiva pero ideal para adquirir conocimiento y encontrar una estructura de los datos (agrupamientos).

2.4.2 Algoritmos de machine learning y deep learning para evaluar sentimientos humanos

El análisis de sentimiento es un procedimiento que implica la evaluación del tono emocional subyacente en un conjunto de expresiones o palabras. Esta técnica se ha vuelto cada vez más relevante en los últimos años, especialmente en el contexto del aprendizaje automático y el aprendizaje profundo (*machine learning* y *deep learning*).

Es importante mencionar esto, porque el desarrollo de la presente tesis se centra en la evaluación del bienestar subjetivo y la satisfacción con la vida y esto connota en un evidente análisis de varios sentimientos. En este sentido, los modelos de *machine learning* y *deep learning* se utilizan para automatizar el proceso de identificación de emociones y actitudes en textos escritos o hablados. Estos modelos son entrenados en grandes conjuntos de datos que contienen ejemplos de texto etiquetados con sus respectivos sentimientos como positivos, negativos o neutros (Parra 2021).

El campo en el que mayoritariamente se ha desarrollado esta clase de análisis, es el de la percepción de usuarios a través de encuestas, opiniones de datos en redes sociales, posicionamiento de marcas, y previsión de ventas. Su importancia se debe a que este tipo de

metodologías permiten categorizar caracteres alfanuméricos y a su vez someterse a un procesamiento del lenguaje natural (NLP por sus siglas en inglés) y a través del análisis del texto y ciencia de datos se puede identificar, extraer y estudiar información subjetiva. Esto a su vez se puede aplicar a datos levantados en tiempo real para determinar automáticamente las emociones y opiniones expresadas (Parra 2021).

El análisis de sentimiento basado en *machine learning* representa una evolución en el análisis de sentimiento tradicional basado en reglas, ya que depende de las capacidades de aprendizaje automático (ML). A diferencia del enfoque basado en reglas, el análisis de sentimiento de ML requiere la alimentación de datos de entrenamiento para que el modelo aprenda a reconocer qué palabras se asocian con diferentes polaridades de sentimiento (Elastic, s. f.).

Una vez que se completa el entrenamiento del análisis de sentimiento basado en machine learning, el proceso se simplifica en la extracción de características y clasificación. Los métodos de análisis de sentimiento de machine learning pueden emplear una variedad de algoritmos de clasificación, como el aprendizaje profundo, regresiones lineales o máquinas de vectores de soporte, para generar resultados precisos y automatizados (Elastic, s. f.).

2.4.3 Geo IA

La inteligencia artificial ha incursionado en todas las ramas de las ciencias de forma innovadora en la adopción y empleo de estas innovadoras como el *machine learning* o inclusive el *deep learning*. Es así como, la geografía e inteligencia artificial se han conjugado para tomar lo mejor de ambos mundos, proponiendo nuevos conceptos como la Geo Inteligencia Artificial (GeoIA) que consiste en una nueva tendencia para las investigaciones y proyectos que emplean datos espaciales para diferentes análisis (Geograma 2021).

En lo que respecta al análisis geoespacial, esta nueva forma de análisis permite responder dos preguntas claves dentro de cualquier investigación: ¿El dónde ocurren las cosas? Y el ¿Por qué suceden dichos fenómenos en el espacio?, por lo cual, esta clase de análisis espacial brinda dos aristas clave en sus resultados, la primera es conocer con antelación lo que va a ocurrir y dónde va a ocurrir y la otra de carácter netamente prescriptivo enfocado hacia las recomendaciones de unas mejores y oportunas toma de decisiones (Geograma 2021).

2.4.3.1 Geo IA y Satisfacción con la vida

Si bien es cierto que el bienestar subjetivo ha sido estudiado desde la psicología positivista a través de encuestas que han permitido su cuantificación y medición, esta clase de investigaciones han compartido un objetivo común que ha sido el de caracterizar o generar perfiles sociodemográficos de un grupo focal, y su subsecuente análisis explicativo ha permitido sentar las bases sobre cuáles son los factores que inciden en mayor o menor magnitud a que una

persona se sienta más o menos feliz. Sin embargo, su análisis a través de modelos estadísticos y computacionales podría considerarse reciente, esto principalmente por la tendencia de análisis que existe actualmente en todas las ramas de la ciencia, por otra parte, su representación y análisis en el campo de las ciencias geográficas ha sido mínimo y casi nulumamente explotado.

Por lo cual y dentro de este contexto, al contar con datos de satisfacción con la vida que se encuentren georreferenciados las posibilidades de analizarlas a través de un enfoque socioespacial y de *machine learning* resulta factible, ya que se toma lo mejor de los tres mundos con el objetivo de entender la variabilidad y causalidad de una mayor o menor felicidad de las personas vistas desde una perspectiva cuantitativa – inferencial y socioespacial.

Así, estudios recientes como los de Mackû et al. (2020) y Abubakar (2019) usan técnicas de *machine learning* y *deep leargning* de forma innovadora para desarrollar y complejizar análisis que emplean variables de índole subjetivo (como la satisfacción con la vida) que generalmente son levantados a través de encuestas con el uso de metodologías propuestas hace casi 4 décadas por Diener et al. (1985).

En el caso de Mackû et al. se ahonda en el análisis causal de medidas objetivas (ingresos económicos, desempleo, educación, índices de salubridad, índices de carácter ambiental, entre otros) y como estas inciden de diferentes formas al enfoque subjetivo de la evaluación de la calidad de vida, este mismo autor desarrolla toda su investigación en países de la Unión Europea y presenta sus principales resultados en una serie de modelos cartográficos para cada uno de los países que conforman este dominio espacial.

Por otra parte, en el caso de Abubakar se aborda un escenario diferente en el que a pesar de no contar con datos de índole georreferenciado o que correspondan a un nivel de agregación geográfico se plantean modelos de índole inferencial en que se pretende indagar aquellos factores que inciden para una mayor o menor satisfacción con la vida en grupos generacionales diferentes.

Sin duda, los estudios previamente citados ejemplifican cómo las técnicas de inteligencia artificial, como el *machine learning (ML)* y el *deep learning (DL)*, están cada vez más presentes en entornos que se ocupan de variables subjetivas, como la satisfacción con la calidad de vida.

Estos análisis, que amplían el alcance de metodologías propuestas hace décadas, ofrecen un nuevo nivel de profundidad en la comprensión de cómo factores objetivos, como ingresos económicos, educación y salud, influyen en la percepción subjetiva de la calidad de vida. Además, estas investigaciones se han adentrado en contextos diversos, lo cual connota un alto grado de adaptabilidad y potencial futuro de desarrollo investigativo.

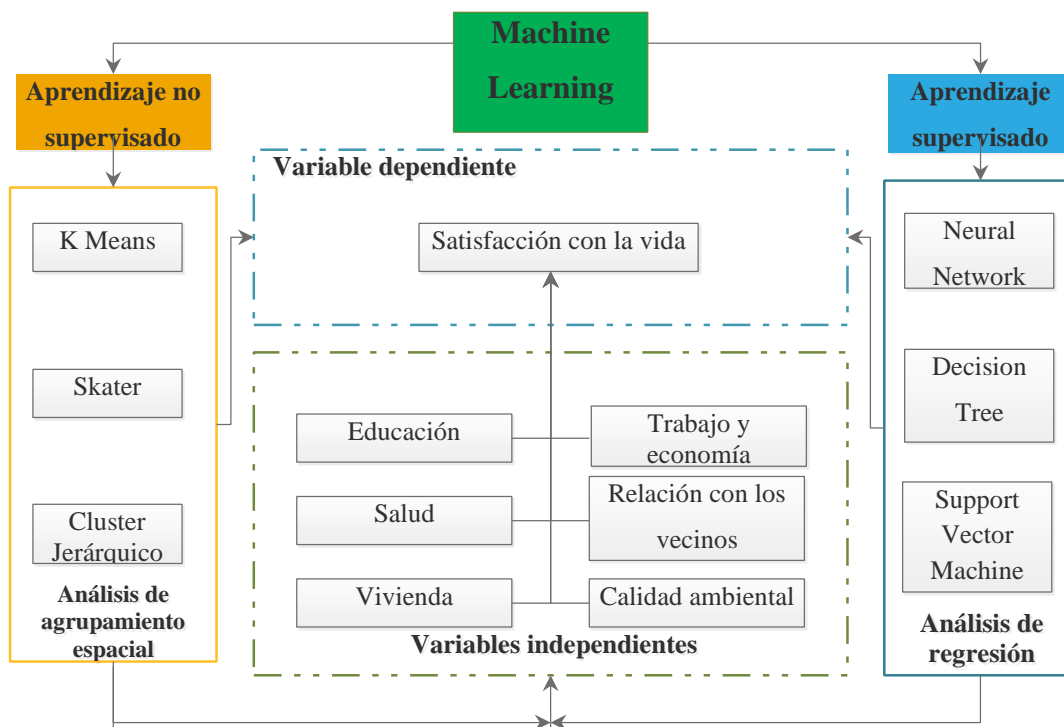
Este enfoque no solo es relevante en términos de investigación académica, sino que también se traduce en aplicaciones prácticas. Estas técnicas se han convertido en herramientas valiosas para el cálculo de índices de calidad de vida y la evaluación de sentimientos en entornos comerciales. La capacidad de analizar y comprender las emociones y las percepciones subjetivas de las personas es fundamental para adaptar estrategias de desarrollo, políticas públicas e inclusive marketing y mejorar la satisfacción del cliente. Razón por la cual, el uso de técnicas de inteligencia artificial en el análisis de variables subjetivas enriquece tanto la investigación como la toma de decisiones en diversos campos, desde la academia hasta el mundo empresarial.

Capítulo 3. Metodología

3.1 Flujograma Metodológico

El esquema lógico de la metodología corresponde a diferentes técnicas de análisis cuantitativo que tienen como variable dependiente a la satisfacción con la vida y como variables independientes a satisfacciones de distintas subdimensiones (educación, salud, vivienda, trabajo y economía, demográficas, ambientales, entre otras). Se plantean dos conjuntos metodológicos a aplicarse para la elaboración de la presente Tesis.

Figura 3.1. Operacionalización de la metodología



Fuente: elaborado por el autor

Como se puede observar, existen dos líneas metodológicas que han sido planteadas y que se encuentran cubiertas por un enfoque metodológico macro que es el *machine learning*, la primera hace referencia al aprendizaje no supervisado con los modelos *k-means*, *skater* y *cluster* jerárquico mientras que el segunda abarca análisis considerados como de aprendizaje supervisado y corresponde a análisis de redes neuronales, árboles de decisiones y máquinas de vector soporte.

3.2 Descripción del área de estudio

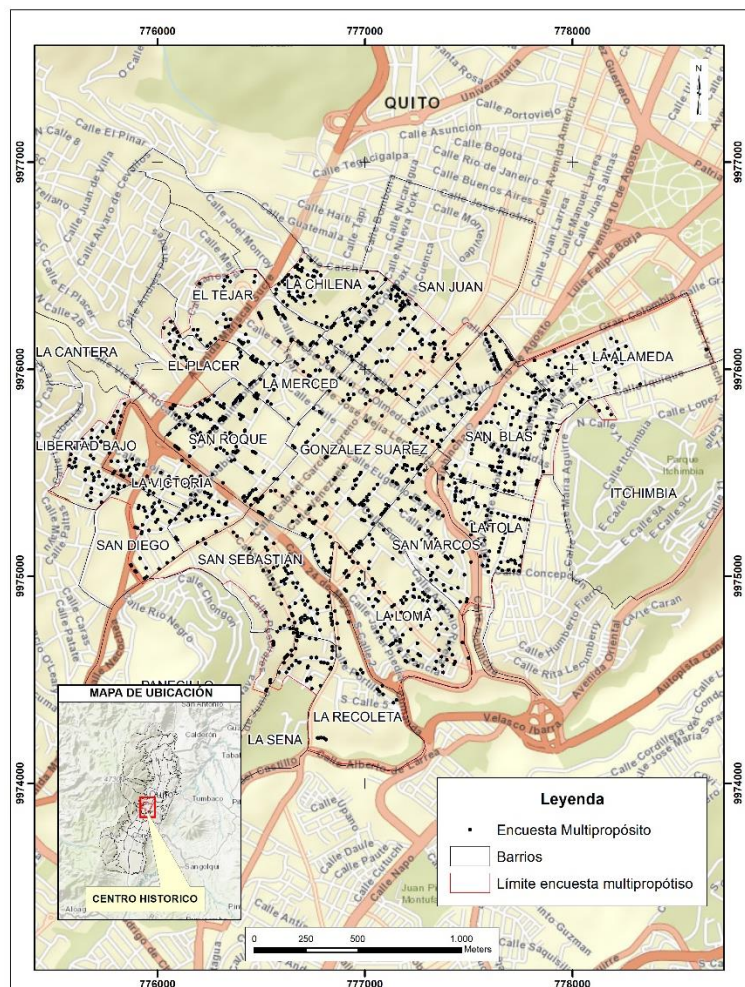
El Centro Histórico de Quito comprende una centralidad urbana de 376 hectáreas de superficie edificada, adicionalmente cuenta con 230 hectáreas de superficie no edificada o de protección natural. Estas 230 ha están constituidas por las elevaciones de El Panecillo, El Itchimbiá,

estribaciones centrales del Volcán Pichincha y un tramo del río Machángara. Por otro lado, en el área edificada se cuenta con una estructura vial definida por una retícula cuadrangular que se adapta a la topografía montañosa existente, en la que se asientan casi cinco mil inmuebles inventariados y alrededor de 130 edificaciones de carácter monumental. El área edificada tiene como núcleo central al barrio Gonzáles Suárez con una superficie de 56 ha y a sus alrededores se encuentran de forma radial los barrios de La Alameda, San Blas, El Tejar, San Roque, La Chilena, El Placer, Aguarico, San Diego, San Sebastián, La Recoleta, La Loma, La Tola, San Marcos y ocupan aproximadamente 320 ha de superficie (MDMQ 2003).

Dentro de las 376 ha de edificación se cuentan con un total de 3.049.292 m² de construcción de los cuales, el 46,76% es utilizado para vivienda residencial, mientras que el 53,24% restante es utilizado para diferentes usos, culturales, comerciales, de entretenimiento, entre otros (MDMQ 2011).

A continuación, se presenta el mapa del área de ubicación de la presente Tesis

Mapa 3.1. Mapa de ubicación del área de estudio



Fuente: elaborado por el autor con base a información de ICQ (2016)

3.3 Fuente de datos - Encuesta Multipropósito 2016 (EMP)

La presente Tesis se encuentra delimitada espacialmente en el CHQ y enmarcada en los resultados de la Encuesta Multipropósito del DMQ del año 2016, realizada por el Instituto de la Ciudad

Esta encuesta tuvo como cobertura geográfica 4 subáreas del Centro Histórico de Quito: Núcleo Central, Eje 24 de Mayo, Zona Envolverte Oriental y Zona Envolverte Occidental. Se realizaron un total de 1967 encuestas a una población objetiva de hogares particulares de estas subzonas, que fueron representadas por el jefe o subjefe de familia. Adicionalmente, la encuesta tenía por objetivo “recabar información sobre el acceso de la población residente ante bienes y servicios públicos, los niveles de satisfacción de dichos servicios, la capacidad de acceso igualitario a oportunidades y el bienestar psicosocial del entorno donde habitan, entre otros aspectos” (ICQ 2016).

La encuesta maneja un nivel de confianza del 95% y un error muestral inferior al 5% para cada una de las subzonas y fue administrada de forma personal en el domicilio de cada informante siguiendo una estrategia de muestreo estratificado por tipos de hogar en caída libre por subzona.

3.4 Elección de variables

La encuesta multipropósito obtuvo información estadística de aspectos económicos y sociales del Centro Histórico de Quito, se compone de 125 indicadores que se repartieron en ejes y dimensiones de análisis. De este modo para el desarrollo metodológico de esta Tesis se utilizaron los 1967 registros (toda la base de datos) para el análisis descriptivo-estadístico de las variables sociodemográficas como: género, edad, nivel educativo, y origen de la persona.

Por otra parte, para los análisis de *machine learning* (inferencial) se utilizaron tan solo 1424 registros (base depurada) de la dimensión bienestar subjetivo, ya que se omitieron aquellos registros cuya respuesta para alguna de sus subdimensiones sean “No sabe” o “No responde”.

Al existir una reducción del 27.56% de las observaciones totales del levantamiento de la encuesta es necesario justificar que el sentido estocástico del muestreo no se haya perdido para lo cual se analizan los márgenes de error para todo el CHQ obtenidos para la 1967 datos y para la base de datos reducida (1424 registros).

Tabla 3.1. Matriz comparativa de la reducción de variables

PRIMER NIVEL: POR DOMINIOS GEOGRAFICOS	1. DATOS y PARAMETROS POBLACIONALES (1)				2.- NIVEL DE ERROR: 1967 registros dada la muestra “n” (con efecto del UNIVERSO N) *		2.- NIVEL DE ERROR: 1424 registros dada la muestra “n” (con efecto del UNIVERSO N) *		Diferencia en el número de observaciones	Diferencia en el margen de error
	N=Universo	P	Q	Z (confiabilidad)	Muestra	Nivel de error (n dado y con efecto del UNIVERSO)	Muestra	Nivel de error (n dado y con efecto del UNIVERSO)		
				95						
TOTAL	15902	50	50	1,96	1.967	2,07%	1.424	2,48%	543	0,41%
Eje 24 de mayo	4211	50	50	1,96	504	4,10%	311	5,35%	193	1,25%
Núcleo Central	919	50	50	1,96	271	5,00%	214	5,87%	57	0,87%
Zona envolvente occidental	4881	50	50	1,96	547	3,95%	428	4,52%	119	0,58%
Zona envolvente orienta	5891	50	50	1,96	645	3,64%	471	4,33%	174	0,69%

Fuente: elaborado por el autor con base a información de ICQ (2016)

Una vez depurada la base de datos de 1967 observaciones a 1424 datos, correspondientes a registros con todos los campos completos, se logra mantener los niveles de confianza. En este sentido, los datos reportados para el CHQ conservan un margen de error del 2.5%, por debajo del límite del 5%, lo cual resulta altamente satisfactorio para propósitos de investigación.

En otras palabras, la reducción en la base de datos es plenamente justificada, ya que no compromete el sentido de aleatoriedad en el muestreo establecido por el ICQ en ese período. Ni tampoco pone en riesgo la validez general del estudio.

En la siguiente tabla se detallan los 8 indicadores utilizados para el *análisis machine learning*, sus etiquetas y escala de valoración con la que se podía calificar al momento de realizar la encuesta:

Tabla 3.2. Variables empleadas para el análisis de ML

Dimensión	Identificador / Etiqueta	Descripción / Subdimensión	Escala de valoración EMP
Bienestar Subjetivo	F4_6	Satisfacción con la vida	0 – Nada satisfecho 10 – Muy Satisfecho No sabe No responde
	F4_1	Educación alcanzada	
	F4_2	Salud actual	
	F4_3	Situación económica	
	F4_4	Vivienda actual	
	F4_5	Trabajo	
	F4_8	Relación con los vecinos	
	F4_13	Calidad Ambiental	

Fuente: elaborado por el autor con base a información de ICQ (2016)

3.5 Análisis exploratorio de datos

El análisis exploratorio de datos desempeña un papel fundamental en las ciencias socioespaciales, una disciplina interdisciplinaria que combina la geografía, la estadística, la demografía y otras áreas para comprender las interacciones entre la sociedad y el entorno geográfico.

Este enfoque analítico se centra en examinar y visualizar conjuntos de datos complejos relacionados con fenómenos sociales y espaciales, con el objetivo de revelar patrones, tendencias y relaciones ocultas.

El análisis exploratorio de datos en las ciencias socioespaciales no estaría completo sin considerar la autocorrelación espacial y el uso de técnicas como el Análisis de Autocorrelación Espacial Local (LISA, por sus siglas en inglés) y G de Getis Ord. La autocorrelación espacial se refiere a la tendencia de las observaciones similares a estar agrupadas en el espacio, lo que significa que las unidades geográficas vecinas tienden a exhibir valores similares en una variable específica.

Al aplicar el análisis de autocorrelación espacial LISA en las ciencias socioespaciales, y junto con análisis complementarios como G de Getis Ord, se pueden identificar áreas geográficas con patrones significativos, lo que contribuye a una mejor comprensión de los procesos sociales y espaciales subyacentes y proporciona información crítica para la toma de decisiones en políticas públicas y planificación territorial.

3.5.1 Análisis de autocorrelación espacial

Este análisis permite medir la autocorrelación espacial basada en las ubicaciones (distribución en el espacio) y los datos de cada una de las entidades simultáneamente. En el caso de esta técnica metodológica, se ejecutará esta metodología para las variables: Satisfacción general con la vida y sus coordenadas geográficas de ubicación. La herramienta es una estadística deductiva, lo que significa que su interpretación estará basada en el contexto de su hipótesis nula y alternativa detalladas a continuación (Mitchell 2005):

H_0 : *No hay autocorrelación espacial*

H_1 : *Existe una autocorrelación espacial*

La estadística de Moran I puede tomar valores entre -1 a 1. Aquellos valores que tienden a -1 indican una dispersión perfecta mientras que los valores que tienen a 1 indican una correlación perfecta y un valor de 0 indica un patrón espacial completamente aleatorio. Para fines prácticos los valores de Moran I pueden ser transformados al estadístico Z y de este modo realizar su posterior interpretación con los diferentes niveles de significancia estadística.

Para este análisis se debe tener en consideración los valores de confianza y significancia más comúnmente usados para fines de investigación. Estos valores pueden variar entre el 1% y 10% en función al nivel de confiabilidad deseado para el estudio. En la siguiente tabla se detallan los valores de interés (valores críticos) que se debe tener en cuenta para la interpretación de resultado.

3.5.2 Niveles de significancia

Tabla 3.3. Valores de interés a diferentes niveles de significancia

Niveles de significancia para diferentes valores Z			
Significancia	10%	5%	1%
P - Valor α	$\alpha = 0,1$	$\alpha = 0,05$	$\alpha = 0,01$
Confianza	90%	95%	99%
Valor Z	± 1.65	± 1.96	± 2.58

Fuente: elaborado por el autor con base a información de Mitchell (2005)

Esta herramienta permite identificar la tendencia general de un fenómeno en el espacio y determinar si el mismo se encuentra aislado, concentrado o difuso (sigue o no una distribución aleatoria en el espacio).

De este modo existen 3 posibles resultados de este análisis y su interpretación estará dada con el análisis del valor P y su puntuación Z asociada que se detallan a continuación:

- El valor P no es estadísticamente significativo: No se puede rechazar H_0 . Es posible que la distribución espacial de los valores sea el resultado de procesos espaciales aleatorios.
- El valor P es estadísticamente significativo y la puntuación Z es positiva: Se puede rechazar H_0 . La distribución espacial de los valores altos y los valores bajos en el conjunto de datos está más agrupada espacialmente de lo que se esperaría si los procesos espaciales subyacentes fueran aleatorios.
- El valor P es estadísticamente significativo y la puntuación Z es negativa: Se puede rechazar H_0 . La distribución espacial de los valores altos y bajos en el conjunto de datos está disperso y refleja un proceso de competitividad en los datos y también indica que los procesos espaciales subyacentes son más dispersos de lo que se esperaría si estos procesos fueran aleatorios (ESRI - Arcgis Pro 2022).

3.5.3 Principio LISA

Una vez que se realiza el análisis de autocorrelación espacial con el estadístico de Moran's I resulta conveniente utilizar y representar de forma visual e informativa estos resultados con el principio de indicador local de asociación espacial (LISA) por sus siglas en inglés. LISA tiene

dos características importantes, la primera es que proporciona una estadística para cada ubicación con una evaluación de importancia y la segunda es que establece una relación proporcional entre la suma de las estadísticas locales y una estadística global correspondiente.

Es importante mencionar que existen muchas estadísticas para la autocorrelación espacial global y por ende existen muchos LISA correspondientes. Sin embargo, para fines prácticos de esta investigación se enfocará al LISA que comparte el *Moran's I*. Esto principalmente porque este análisis indica de forma genérica (a manera de análisis exploratorio de datos) que la autocorrelación espacial de la satisfacción con la vida en el CHQ consiste en una combinación de una medida de similitud de atributos entre un par de observaciones de un conjunto de datos y que además tiene una distribución específica dentro del espacio que puede ser categorizado de la siguiente forma:

- Alto – Alto: Valores altos rodeados de valores altos
- Bajo – Alto: Valores bajos rodeados de valores altos
- Alto – Bajo: Valores altos rodeados de valores bajos
- Bajo – Bajo: Valores bajos rodeados de valores bajos
- No significativos: Valores cuyo P-valor no cumple con la significancia deseada.

Al considerar esta categorización de forma implícita se entiende que los datos pueden llegar a formar clústers. Sin embargo, estrictamente hablando los mapas de significancia y conglomerados vendrían a representar los núcleos (Puntos calientes y puntos fríos) de los clústers (Anselin 2020).

De forma similar al caso del análisis de autocorrelación espacial se utilizará la variable Satisfacción general con la vida y su ubicación espacial correspondiente para la utilización de esta técnica metodológica.

3.6 Aprendizaje no supervisado

Para el caso de la metodología de aprendizaje no supervisado se utilizan dos subconjuntos de variables, en el primer escenario se lo ejecuta únicamente a la variable Satisfacción con la vida (F4_6), y para el segundo caso se emplean las variables: Educación alcanzada (F4_1), Salud Actual (F4_2), Situación económica (F4_3), Vivienda actual (F4_4), Trabajo (F4_5), Relación con los vecinos (F4_8), Calidad Ambiental (F4_13) y también la inclusión de la variable Satisfacción con la vida (F4_6). En este apartado metodológico no se plantea un modelo conceptual de variables dependientes e independientes.

Para el desarrollo de todas las metodologías propuestas en los siguientes acápites del aprendizaje no supervisado se empleó el software de procesamiento GEODA en su versión 1.20 que es de carácter gratuito y de código abierto.

3.6.1 Regionalización espacial

La presente Tesis se centra en el contexto geográfico del Centro Histórico de Quito y la percepción de la satisfacción con la vida. Al contar con datos georreferenciados resulta factible realizar análisis de regionalización espacial, que permitan distinguir regiones agrupadas de regiones no agrupadas, que en este caso particular no resulta tan sencillo ya que el tipo de elemento gráfico espacializado son diferentes puntos distribuidos de forma heterogénea en el área de interés.

Bajo este contexto, se emplearon diferentes técnicas de regionalización espacial que básicamente son “aprendizajes computacionales” que pueden ser de tipo supervisados y no supervisados. Estas técnicas de regionalización espacial, más conocidas como *clustering*, son una de las herramientas más utilizadas en el análisis exploratorio de datos, permiten identificar segmentaciones de datos y entender de forma intuitiva cómo están estructurados los mismos.

3.6.1.1 K-Means

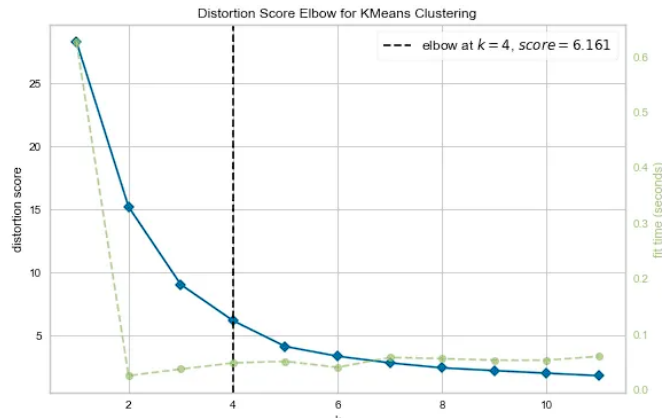
Una de las técnicas computacionales más conocidas es el algoritmo *k-means*, que tiene por objetivo agrupar observaciones semejantes para descubrir patrones que simple vista no se distinguen. Para obtenerlos, el algoritmo busca un número fijo de k – *clusters* en el conjunto de datos.

Para saber el número de (k) *clusters* se puede emplear el método del codo ya que este es un parámetro indispensable para la ejecución del algoritmo. Este valor representa a la cantidad ideal de centroides que se requiere clasificar al conjunto de datos (Sanz 2022).

El procedimiento para elegir al número de *clusters* ideales consiste en ejecutar de forma reiterativa al algoritmo con la variación del parámetro k , el mismo que puede tomar valores que va desde 2 hasta 10 o más centroides, a continuación se construye una gráfica con el uso de un plano cartesiano en el que en el eje X se colocará al número K y en el eje Y se colocará los valores de la distancia media de cada punto hasta su centroide y el tiempo de entrenamiento necesario, que generalmente depende del número de iteraciones (conocido como criterio de parada). El objetivo de esto es entender que a medida que se va aumentando la cantidad de centroides, la distancia media de los puntos al centroide disminuirá cada vez menos.

La grafica resultante es una curva y para determinar el número de *clusters* óptimo se lo realizará mediante la observación del “codo” o punto de inflexión en el cual la pendiente de la curva cambie considerablemente de dirección. A continuación, se presenta una figura en la que se puede observar el punto de inflexión o codo a manera de ejemplo (Sanz 2023).

Figura 3.2 Gráfica método del codo



Fuente: (Sanz 2023)

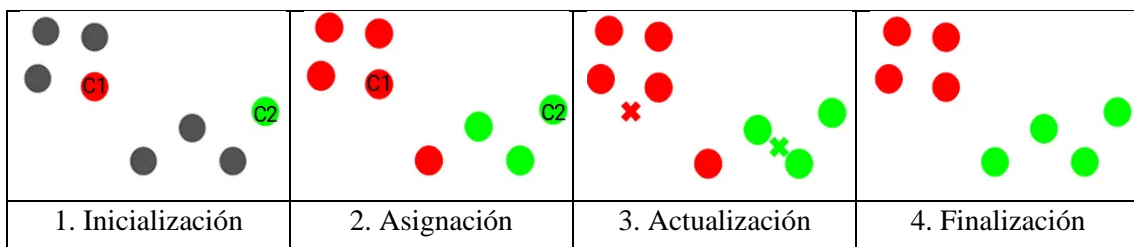
En este ejemplo se puede observar que el “codo” se encuentra entre los datos de $k=4$, $k=5$ o $k=6$, para este punto puede tomar relevancia la experiencia o el criterio del investigador para elegir cual es el más conveniente en función al tipo de datos y análisis.

Una vez que se cuenta con el número de *clusters* o valor de k el algoritmo ejecuta los siguientes pasos (Sanz 2023):

1. Inicialización: se establecen k centroides en el espacio de los datos, de forma aleatoria se escoge la ubicación de un punto.
2. Asignación objetos a los centroides: cada objeto de los datos es asignado a su centroe más cercano.
3. Actualización centroides: se recalcula la posición del centroe de cada grupo tomando como nuevo centroe la distancia y posición del promedio de los objetos pertenecientes a dicho grupo.
4. Finalización: Se repiten los pasos 3 y 4 hasta que se cumple el criterio de parada que generalmente responde al número de iteraciones asignado para la tarea.

En la siguiente figura se puede observar de forma visual el proceso de asignación de *clusters*

Figura 3.3 Procedimiento K- Means



Fuente: (Sanz 2023)

3.6.1.2 Skater

El método *SKATER* por sus siglas en inglés (*Spatial Kluster Analysis by Tree Edge Removal*) consiste en un procedimiento netamente espacial que identifica clústers locales de las variables estudiadas. A diferencia del método *k-means* el método *Skater* opera a través de árboles de decisión eliminando (podando) las relaciones de árboles que no son eficientes (Assunção et al. 2006).

Dentro de los análisis de regionalización y *clustering* a menudo se determina la existencia de un efecto espacial que afecta las características del conglomerado es así como *SKATER* permite el agrupamiento basado en la relación espacial entre la dependencia de ubicaciones, utilizando el enfoque de la teoría de grafos con la técnica del árbol de expansión mínimo.

El algoritmo se compone de 3 subprocesos que consisten en:

1. Creación de un gráfico para la estructura de contigüidad (Ubicaciones X y Y en el espacio): A partir de la ubicación de cada una de las observaciones (o del centroide en el caso que sean polígonos) se construye una red que conecta a cada uno de los nodos (puntos), para la construcción de este mallado se tiene en cuenta una matriz de pesos espaciales que generalmente depende de las distancias más cercanas entre observaciones.
2. El cálculo del árbol de expansión mínimo para ese gráfico: el dominio espacial de este gráfico corresponde a un espacio de búsqueda reducido y determinado por el árbol de expansión mínimo (MST), este gráfico incluye todos los nodos de la red, pero pasa por cada uno solo una vez. Por lo tanto, se simplifica la complejidad del gráfico original (podar al árbol).
Este árbol resultante tiene n nodos y $n-1$ aristas y se denomina “mínimo” porque el algoritmo minimiza la longitud total (o el costo) del árbol. De este modo se entiende que, en su forma más simple el costo consiste en las distancias entre nodos. Sin embargo, se utiliza una medida de costo más general en forma de medida de disimilitud multivariante (distancia euclidiana multivariante) entre cada par de nodos.
3. Clasificación: una vez que se cuenta con el MST, se ejecuta el algoritmo con el ingreso de la matriz de datos (observaciones) y el número de clústers que se requiere para clasificar. De este modo se cortan los bordes de manera que minimice la falta de similitud en los grupos resultantes, obteniendo así clústers que tanto espacialmente como en sus datos presenten valores similares para todos los campos de análisis especificados (Anselin 2017).

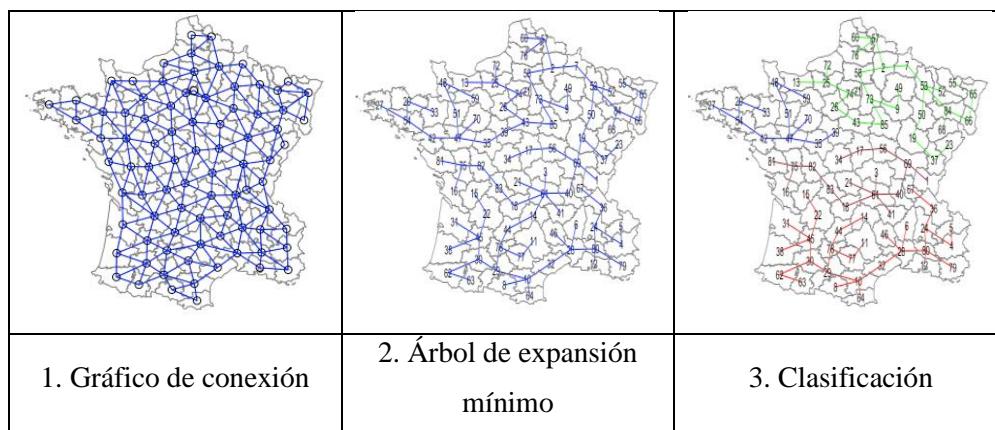
Para determinar el número de *clusters* del algoritmo *SKATER* se evalúa la eficacia de este mediante la interpretación de la pseudo estadística F o Calinski – Harabaz que consiste en una

proporción de la varianza entre *clusters* respecto a la varianza dentro del clúster (este concepto también es conocido como ratio de la varianza) (Duque, Ramos, y Surinach 2007). Dado que esta estadística se basa en la comparación de la relación ponderada entre la suma de los cuadrados (la medida de separación del clúster - dato del numerador) y la suma de los cuadrados dentro del clúster (la medida de cómo se asemejan los puntos dentro de un clúster – dato del denominador) (Alteryx 2022).

Su correcta interpretación radica en que mientras este valor sea mayor (tienda a 1) es mejor ya que idealmente, los *clusters* deben estar bien separados o diferenciados el uno del otro, por esta razón el numerador debe ser un valor alto pero los puntos dentro del clúster deben estar lo más cerca posible el uno del otro, lo que implica que en el denominador deben encontrarse valores más bajos (Alteryx 2022).

En la siguiente figura se puede observar de forma visual el proceso de asignación de *clusters* con el algoritmo SKATER

Figura 3.4 Procedimiento SKATER



Fuente: (Anselin 2017)

3.6.1.3 Hierarchical clustering

Este modelo a agrupamiento a diferencia de *k-means* y *Skater* utiliza un enfoque de agrupación jerárquica, lo que significa que puede ser entendido de forma descendente o ascendente (árbol piramidal o dendrograma).

Un enfoque de arriba hacia abajo se entiende cuando el conjunto de datos se encuentra en un primer momento agrupado en su totalidad y posteriormente el algoritmo encuentra un punto de ruptura para crear dos subgrupos y el proceso continúa hasta que cada una de las observaciones es su propio grupo. Por otro lado, un enfoque de abajo hacia arriba comienza con la asignación de cada observación a su propio grupo y el algoritmo encuentra observaciones más cercanas y

similares para combinarse en un grupo hasta que al final del proceso existe un único grupo que contiene todas las observaciones (Anselin 2018).

La forma más óptima de visualizar esta partición / agrupación es a través de un dendrograma y se utiliza para establecer un nivel de corte para un k especificado. Al ingresar este dato el punto de corte se ve representado a diferentes niveles en el árbol, obteniendo así subgrupos con dimensiones variables (Anselin 2018).

Para definir la partición o agrupación el algoritmo evalúa la distancia entre dos conglomerados y los une o separa un función a los enlaces que minimicen el nivel de disimilitud (mejor relación entre observaciones y menor relación entre *clusters*) (Anselin 2018).

La forma más eficiente de entender el funcionamiento de este algoritmo es la agrupación jerárquica aglomerativa. Sin embargo, también existe una explicación para la agrupación jerárquica divisiva. A continuación, se explica la operacionalización aglomerativo (Lidgi González 2019).

En enfoque aglomerativo se entiende cuando cada observación se trata como un *cluster* y luego se fusionan de forma sucesiva y por pares. A medida que cada iteración se va ejecutando los grupos más similares se fusionan entre sí hasta formar k grupos o un único grupo. Este modelo se compone de 4 pasos que son:

1. Cálculo de la matriz de proximidad
2. Identificación del número de clústers iniciales (n observaciones)
3. Iteración de fusión entre los clústers más cercanos y con mayor nivel de similitud en sus observaciones
4. Definición de parada, hasta obtener un único clúster.

Las particularidades de este método es que es una técnica poderosa que permite construir estructuras de tipo árbol y que puede ir en ambos sentidos, también es necesario mencionar que al igual que el método de *k-means* se debe definir el número de k clústers a partir de un análisis previo, y que puede ser la misma gráfica del codo. (Anselin, 2017)

3.7 Aprendizaje supervisado

El aprendizaje supervisado es un método de análisis de datos que utiliza algoritmos de aprendizaje iterativo dentro de un conjunto de datos. El aprendizaje supervisado utiliza un conjunto de datos de entrenamiento para “enseñar” a los modelos a generar la salida deseada (TIBCO 2023b).

El conjunto de datos se clasifica en los datos de entrada (variables independientes) y resultados correctos (variables dependientes) que permite que el modelo aprenda a través del tiempo. Este tipo de aprendizaje se clasifica en dos tipos que son: la clasificación y la regresión (IBM 2023).

Para fines de esta investigación se hace énfasis en el aprendizaje supervisado de regresión que como su nombre indica se utiliza para comprender la relación entre variables dependientes e independientes. Los algoritmos de regresión en el aprendizaje supervisado necesitan identificar una relación funcional entre los parámetros de entrada y de salida, la precisión de esta relación funcional se evalúa en función a la desviación entre la salida precisa y la salida prevista

Este método se utiliza generalmente con el objetivo de construir un modelo que pueda predecir con precisión la salida en función de las características de entrada y a su vez comprender cuáles son las relaciones subyacentes entre dichas características (IBM 2023).

Para el caso de la metodología de aprendizaje supervisado se plantea un modelo conceptual de variables dependientes e independientes. En este sentido la variable dependiente corresponde a la variable Satisfacción con la vida (F4_6), mientras que las variables independientes son: Educación alcanzada (F4_1), Salud Actual (F4_2), Situación económica (F4_3), Vivienda actual (F4_4), Trabajo (F4_5), Relación con los vecinos (F4_8), Calidad Ambiental (F4_13) y también la inclusión de la variable Satisfacción con la vida (F4_6).

Para el desarrollo de todas las metodologías propuestas en los siguientes acápites del aprendizaje supervisado se empleó el software de procesamiento JASP que es de carácter multiplataforma y está orientado al uso de paquete estadísticos de código abierto cuyas siglas en inglés expresan “*Jeffrey’s Amazing Statistic Program*” en reconocimiento al pionero de la inferencia bayesiana Sir Harold Jeffreys.

En los siguientes acápites se describirá a detalle los algoritmos de aprendizaje supervisado utilizados para esta investigación:

3.7.1 Redes Neuronales

Las redes neuronales artificiales (RNA) son modelos matemáticos y computacionales que imitan la interconectividad del cerebro humano a través de capas de nodos. Su funcionamiento se basa en el comportamiento biológico del cerebro humano ya que simula un sistema adaptable que las computadoras emplean para aprender de sus errores y mejoran continuamente.

Esta clase de algoritmos pueden parecer nuevos métodos de análisis porque últimamente se ha popularizado su uso en diferentes ámbitos de las ciencias y la cotidianidad. Sin embargo, es un campo de análisis de más de 70 años. Esta clase de métodos se remonta a la década de 1950 cuando un científico - psicólogo estadounidense generó un primer prototipo de “máquina”

llamado perceptrón que sería utilizado para tomar decisiones binarias sencillas tal como lo haría la mente humana.

Las redes neuronales artificiales pueden ser considerados como modelos de regresión lineales o no lineales, de este modo comparten ciertas semejanzas con los modelos estadísticos clásicos utilizados para la predicción. Sin embargo, las redes neuronales sostienen que los supuestos de normalidad y homogeneidad de varianzas son recomendables, pero no estrictamente necesarios como lo requerido para el correcto funcionamiento de modelos estadísticos clásicos (Menacho Chiok 2014). En este sentido existe terminología homologable entre ambos tipo modelos de predicción que se detallan a continuación:

Tabla 3.4. Términos similares entre modelos estadísticos y redes neuronales

Modelos estadísticos	Modelos de redes neuronales
Observación	Patrón
Muestra	Datos de entrenamiento
Muestra de validación	Datos de validación y prueba
Variables independientes / explicativas / predictoras	Variables de entrada
Variables dependientes	Variable de salida
Modelo	Arquitectura
Residual	Error
Error aleatorio	Ruido
Estimación	Entrenamiento o aprendizaje
Interacción	Conexión funcional
Coefficientes	Pesos de conexión
Intercepto	Peso umbral o sesgo

Fuente: elaborado por el autor con base en datos de Menacho Chiok (2014)

Empleando este mismo principio resulta conveniente indicar aquellos parámetros estadísticos resultantes de los algoritmos que permiten realizar una comparación y evaluación de la calidad que tiene cada uno los diferentes modelos de predicción:

Tabla 3.5. Parámetros estadísticos resultantes

MSE Error Cuadrático Medio	Mide la cantidad de error que hay entre dos conjuntos de datos, compara un valor predicho y un valor observado. Mide el promedio de los errores elevados al cuadrado
--------------------------------------	---

<p style="text-align: center;">RMSE</p> <p>Raíz del error cuadrático medio</p>	<p>Es la desviación estándar de los valores residuales (errores de predicción).</p> <p>Los valores más bajos de RMSE indican un mejor ajuste. Este valor es una buena medida de la precisión con que el modelo predice la respuesta.</p>
<p style="text-align: center;">MAE / MAD</p> <p>Error absoluto medio del modelo</p>	<p>Es el promedio de la diferencia absoluta entre el valor observado y los valores predichos</p>
<p style="text-align: center;">MAPE</p> <p>El error porcentual absoluto medio del modelo</p>	<p>Es el promedio de la diferencia absoluta entre el valor observado y los valores predichos</p>
<p style="text-align: center;">R²</p> <p>La proporción de la varianza de una variable dependiente que se explica por una o varias variables independientes</p>	<p>Indica la bondad o la aptitud del modelo, a menudo se utiliza con fines descriptivos y muestra que tanto explican las variables independientes la variabilidad en su variable dependiente.</p> <p>Su escala es intuitiva va de 0 a 1, cuando sus valores tienden a 0 indican que el modelo propuesto es completamente imperfecto, mientras que un valor de 1 indica una predicción perfecta.</p>

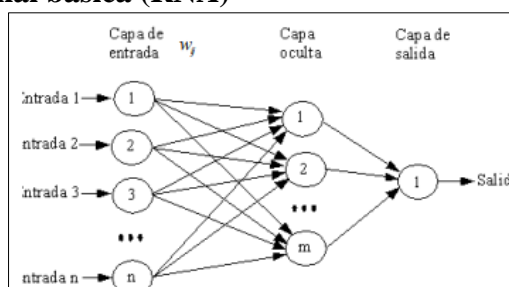
Fuente: (Ligdi González 2018)

3.7.1.1 Estructura de la red neuronal

Las RNA están compuestas por un conjunto de elementos simples (neuronas) que se interconectan unas entre otras masivamente, en paralelo y con una organización jerárquica. Estas características hacen que este tipo de metodologías puedan describir muchos fenómenos del mundo real y que puedan ser aplicables y equiparables a los modelos estadísticos utilizados para la predicción, clasificación y agrupamiento (Menacho Chiok 2014).

En su forma más básica esta clase de algoritmos se compone de tres capas de neuronas que constituyen un sistema que básicamente se compone de la capa de entrada (punto de entrada de los datos), una capa oculta (donde se procesa la información), la capa de salida (dónde el sistema decide cómo proceder en función de los datos).

Figura 3.5. Red neuronal básica (RNA)



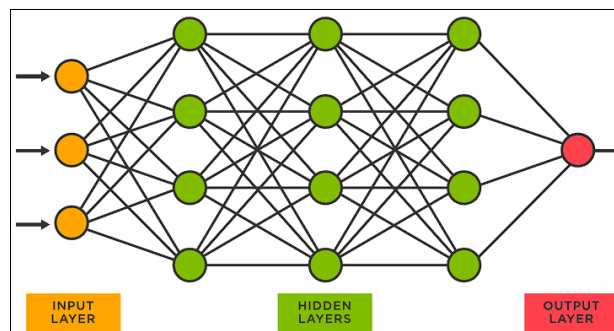
Fuente: (Menacho Chiok 2014)

Como se observa en la figura previa, la RNA se puede representar de manera gráfica (Grafo) y que está compuesta de:

- **Neuronas:** son elementos constitutivos más básicos de una RNA, cada neurona según su etapa de procesamiento posee un valor numérico o estado de activación. A su vez cada neurona puede tener múltiples entradas y opcionalmente una entrada constante conocida como sesgo o umbral
- **Conexiones entre neuronas:** Las conexiones son representadas a través de líneas o arcos dirigidos al cual se asocia un peso numérico W_{ij} que es el parámetro que refleja la influencia y fuerza de la sinapsis de conexión.
 - Si W_{ij} es positivo, indica que la interacción entre la neurona i y j es inactivada, en otras palabras, siempre que la neurona i esté desactivada, la neurona recibirá una señal de j que tenderá a activarla.
 - Si W_{ij} es negativo, indica que la interacción es inhibidora, esto implica que si i está activada, está enviando una señal a j que tenderá a desactivarla.
 - Si W_{ij} es cero, esto implica la no conexión.
- **Capa de entrada:** es la capa que recibe directamente la información proveniente de las fuentes externas (variables independientes)
- **Capas ocultas:** Las neuronas de esta capa son internas y no tienen contacto con el entorno exterior (por esta razón se las denomina ocultas). Estas neuronas se pueden interconectar de distintas maneras, determinando las diferentes topologías de las RNA.
- **Capa de salida:** Esta capa corresponde al resultado calculado de la variable de salida (Menacho Chiok 2014).

Partiendo de estas generalidades se puede indicar que existen redes neuronales artificiales más complejas que se componen de múltiples capas (ocultas). A continuación, en la siguiente figura se puede visualizar una red neuronal compleja y el funcionamiento de cada nodo que se compone de entradas, pesos, un sesgo o umbral y una salida (TIBCO 2023a).

Figura 3.6. Esquema de funcionamiento de redes Neuronales



Fuente: (TIBCO 2023a)

3.7.1.2 Funciones de activación

Funciones de entrada: Una neurona depende de un conjunto de señales (entradas) que representan la información del estado de activación de la red de conexión. Considerando esto se tiene que cuando existe el valor de entrada de la neurona i que mantiene una conexión con una neurona j ponderada por un peso W_{ij} , se tiene un efecto aditivo entre cada una de las conexiones. Dicha operación se resume en la siguiente ecuación (Menacho Chiok 2014).

$$f(X_j) = W_0 + \sum_{i=1}^N W_{ij} * X_i$$

Dónde:

- W_0 : es el sesgo del peso (intercepto)
- W_{ij} : Peso sinapsis entre neurona i y j
- X_i : Valores de ingreso de la neurona i
- X_j : Valores de salida de la neurona j

Funciones de salida: Este tipo de función se las conoce como función de activación, transferencia o paso y se convierte en gran medida en la responsable de la topología de la red, su bondad de ajuste y sus subsecuentes análisis de sus indicadores (Menacho Chiok 2014).

Las funciones de salida son elegidas en base al tipo de comportamiento de los datos y su adaptabilidad al modelo de predicción, generalmente se realizan varias iteraciones entre diferentes funciones de activación de las cuales destacan las siguientes:

Tabla 3.6. Funciones de salida comunes para RNA

Función	Detalle
Binaria:	$y_j = \begin{cases} 0 & \text{si } X_j < 0 \\ 1 & \text{si } X_j \geq 0 \end{cases}$
Linear:	$y_j = X_j$
Logística sigmoide	$y_j = \frac{1}{1 + e^{-X_j}}$
Gaussiana	$y_j = e^{-\frac{X_j^2}{2}}$
Trigonométricas – Trigonométricas inversas	$y_j = \text{sen}(X_j)$ $y_j = \text{cos}(X_j)$ $y_j = \text{tan}(X_j)$

Fuente: elaborado por el autor con base a información de Menacho Chiok (2014)

3.7.1.3 Algoritmos de aprendizaje

Existen diferentes algoritmos de entrenamiento o mejor conocidos como algoritmos de aprendizaje, esos procesos computacionales se encargan de realizar una serie de iteraciones con el objetivo de hacer mejoras cuando se experimentan malos resultados en la regresión por RNA.

La red neuronal comienza con todos los pesos inicializados a valores aleatorios, por lo que en su primera iteración arroja un resultado muy lejano al que sea desea predecir, a medida que el algoritmo “aprende”, que consiste en un proceso de minimización de sus errores se ajusta hasta obtener una convergencia entre los pesos resultantes.

El algoritmo de entrenamiento de redes neuronales más común y conocido es la propagación hacia atrás. Este algoritmo distribuye el error que existe entre la variable de salida y la deseada hacia cada una de las neuronas de las diferentes capas de procesamiento. Este es uno de los procesos más utilizados. Sin embargo, existen diferentes variaciones de este que se centran en la: regularización, restricciones de normas, abandono, lotes y mini lotes, entre otros.

Una de estas variaciones es conocida como RPROP (*Resilient Backpropagation With Backtracing*) cuya principal diferencia con la propagación hacia atrás es que cada peso y sesgo tiene una tasa de aprendizaje implícita, variable y diferente. En cada iteración cada peso tiene un valor delta que aumenta cuando el gradiente no cambia de signo (convergencia) o disminuye cuando el gradiente cambia de signo (no convergencia).

3.7.2 Árbol de decisión

Este tipo de métodos crean modelos de regresión o clasificación en forma de una estructura de árbol, ya que descompone a un conjunto de datos en subconjuntos cada vez más pequeños (ramas). Estos algoritmos son uno de los enfoques prácticos más utilizados en el aprendizaje supervisado (Ligdi González 2019).

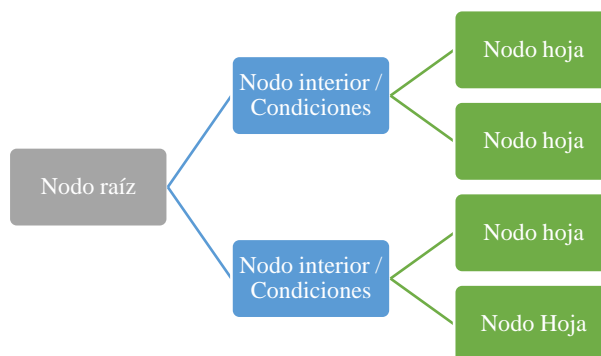
3.7.2.1 Estructura del árbol de decisión

Los árboles de decisión se componen de tres tipos de nodos:

- El nodo raíz: que es el nodo inicial que representa la muestra completa y puede dividirse en más nodos
- Los nodos interiores: que representan las características de un conjunto de datos, a estos nodos se encuentran asociadas las ramas que son las reglas de decisión
- Los nodos de hoja: que representan el resultado final.

En la siguiente figura se muestra el tipo de gráfica resultante de este tipo de metodología:

Figura 3.7. Estructura árbol de decisión



Fuente: adaptado por el autor con base a información de L. González (2019b)

La idea básica del algoritmo es encontrar un punto en el que las variables independientes permiten dividir el conjunto de datos en 2 partes para lo cual considera como punto de decisión a aquellos valores que minimicen el error cuadrático medio. En otras palabras, este tipo de regresiones observa las características de un conjunto de datos y entrena un modelo con la estructura de árbol previamente vista, en la que cada una de las ramas de decisión representan la verdad o falsedad de la declaración (2 partes).

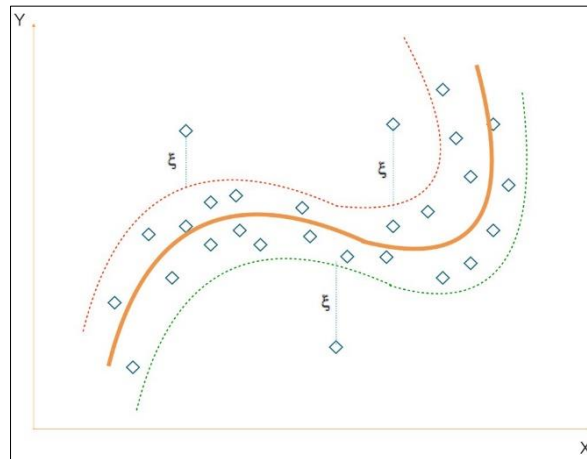
Para la búsqueda y selección de este valor, el algoritmo desarrolla un proceso iterativo en la que toma como posible solución al dominio numérico de cada una de variables dentro de un conjunto de datos, hasta determinar aquella cifra que permita subdividir de mejor forma el conjunto de datos.

3.7.3 Máquinas de vector soporte (SVM)

El método Máquinas de Vector Soporte (SVM por sus siglas en inglés) es un tipo de algoritmo utilizado para la clasificación o regresión de datos, se lo considera como un referentes en el aprendizaje estadísticos y del *machine learning*. Originalmente fue desarrollado como un método de clasificación binaria en la década de 1990 y su fundamento se basa en el concepto de la generación de un hiperplano.

Este algoritmo de regresión se basa en predecir valores numéricos, generando una función (lineal o no lineal) que minimice el error, para lo cual construye un hiperplano o una subfunción paralela que optimice la inclusión de observaciones y teniendo en cuenta que se tolera parte del error. En este sentido, SVM siempre tratará de modelar el comportamiento de los datos a través de una función primaria a la cual se le acompañará un rango (tanto positivo como negativo) que tendrá el mismo comportamiento de la función primaria. A pesar de generar un hiperplano, existirán datos que se encuentren fuera de su rango y serán considerados como errores, por lo cual es necesario calcular su distancia de forma reiterativa para optimizar el modelo matemático. A continuación, se representa una figura que ejemplifica de forma gráfica la metodología SVM (Gonzales 2019).

Figura 3.8. Representación gráfica SVM, función principal, hiperplano y errores.



Fuente: (Gonzales 2019)

3.7.3.1 Representación del SVM

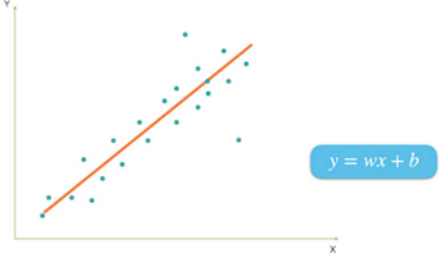
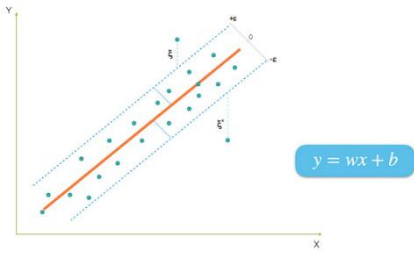
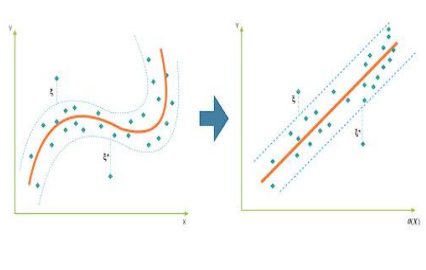
La forma más práctica de entender el funcionamiento y el comportamiento de la regresión por SVM es a través de su representación más básica que es la regresión lineal. Es necesario mencionar que, en esta representación básica el hiperplano correspondería a una línea. Sin embargo, cuando se trabaja con más datos e inclusive con datos no lineales el hiperplano puede ser mucho más complicado que el de una función lineal lo que representaría una mayor demanda computacional.

En este sentido el algoritmo se compone de 4 etapas que se detallan a continuación:

1. La primera es la inicialización y generación de una función primaria
2. El segundo paso es la construcción de unas bandas paralelas al hiperplano (que abarque la mayor cantidad de observaciones (Vectores de apoyo o soporte))
3. El tercer paso es la evaluación. Para esta etapa el algoritmo identifica las distancias entre las bandas (hiperplano) y los puntos que se encuentren fuera los vectores de soporte, dicha distancia es conocida como ϵ .
4. El cuarto paso es la minimización de los errores, el algoritmo de forma iterativa evalúa una serie funciones y vectores de soporte los cuales minimicen la cantidad de errores y sus distancias ϵ , permitiendo así, generar un modelo con el mejor resultado.

En la siguiente figura se presentan las etapas del algoritmo SVM:

Figura 3.9. Procesos del método SVM, función principal, hiperplano y errores.

		
1. Inicialización / generación del hiperplano	2. Bandas soporte	3. Evaluación de distancias 4. Minimización de errores

Fuente: (Gonzales 2019)

Así, se presenta la ecuación que utiliza el algoritmo para la generación del modelo de regresión en el que se pretende maximizar el margen (hiperplano) y la minimización del error de entrenamiento:

$$f(x) = \min \frac{1}{2} \|w\|^2 + C \sum_{ij=1}^N (\varepsilon_{ij} + \varepsilon_{ij}^*)$$

Dónde:

- W es la matriz de pesos del vector o hiperplano
- C es la constante entre la regularidad de la función principal y la tolerancia de desviaciones de las bandas de soporte
- ε_{ij} y ε_{ij}^* son las variables que controlan el error cometido por la función de regresión al aproximadas (distancias)

Este tipo de modelo computacional se ajusta al comportamiento de datos lineales y no lineales, por lo cual, para que los resultados obtengan unos resultados óptimos se deben tener en cuenta varios aspectos como: Un conjunto de datos grandes pueden representar un tiempo de entrenamiento alto y la necesidad capacidad computacional asociada alta. Otro aspecto de interés es que el conjunto de datos debe estar depurados y normalizados previamente y además los datos no deben tener clases o características superpuestas muy similares.

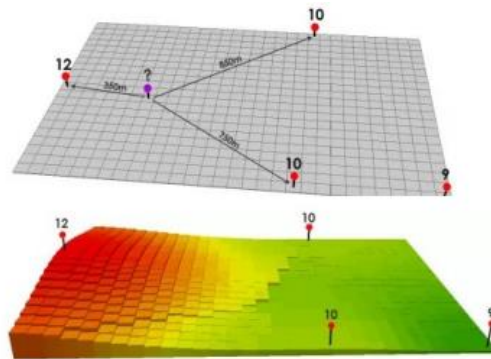
3.8 Representación espacial

3.8.1 IDW

IDW es un método de interpolación espacial conocida como distancia inversa ponderada (por sus siglas traducidas del inglés) y determina los valores de una superficie continua de tipo ráster a través de una combinación ponderada linealmente de un conjunto de puntos de muestra, esta ponderación parte del supuesto que la variable que se representa cartográficamente disminuye su influencia a medida que la distancia desde su ubicación aumenta

Por ser un método de interpolación estima valores desconocidos a partir de valores conocidos cuya matemática es de tipo determinista ya que asume que los valores más cercanos están más relacionados que otros que se encuentren alejados entre sí mismos (ArcGeek 2018). Esta distancia o radio de búsqueda entre puntos se puede especificar si se conoce la influencia o la cantidad de puntos que deben ser tomados en cuenta en función al fenómeno estudiado.

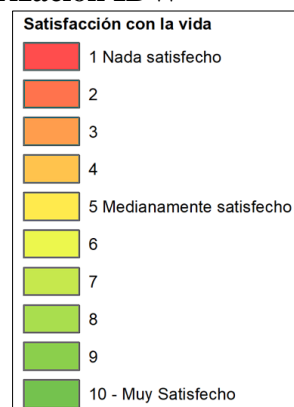
Figura 3.10. Representación visual IDW



Fuente: (ArcGeek 2018)

La herramienta IDW disponible en el conjunto de herramientas de interpolación del software ArcGIS 10.4 permite determinar la potencia que puede ser usada por el algoritmo, este valor puede variar desde 0.25 hasta 3 e indica qué tanto van a verse influenciados las celdas adyacentes al valor interpolar. Una potencia baja ocasiona que la superficie se alise mientras que una potencia alta ocasiona que la superficie se vea ligeramente elevada en las celdas continuas. En la previa se puede observar una representación gráfica de una superficie ráster cuyas celdas fueron interpoladas con IDW los valores en tonos rojos indican valores más bajos de satisfacción con la vida y a manera que se van interpolando a puntos cuyas cifras con más altas esta tonalidad cambia a color verde (valores altos de satisfacción con la vida). A esta representación se le asocia una escala de semaforización que se presenta en la siguiente figura:

Figura 3.11. Escala de semaforización IDW



Fuente: elaborado por el autor

Capítulo 4. Resultados

4.1 Estadísticas descriptivas sociodemográficas EMP

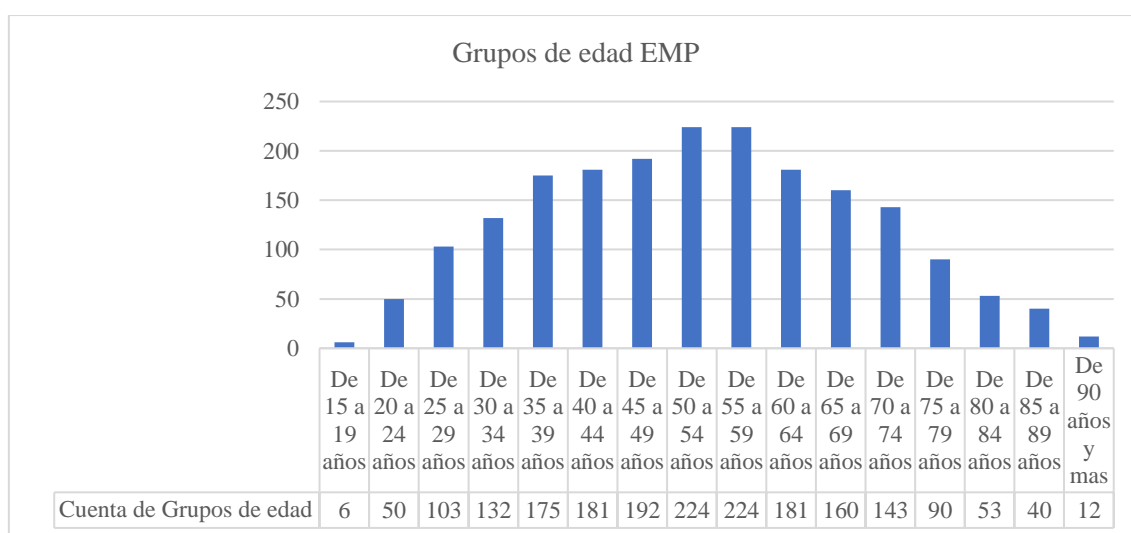
La encuesta multipropósito del año 2016 cuenta con un total de 1967 observaciones que corresponde al número de hogares que fueron parte de la selección muestral. De estos datos se presentan las principales estadísticas sociodemográficas.

La encuesta debía cumplir la condición de ser respondida por personas mayores de 18 años de edad, que sean los informantes jefes de familia o conyugues y que al menos haya estado residiendo en la vivienda un año a la fecha de la encuesta, en este sentido la muestra cumple con en un 100% las condiciones previamente descritas.

4.1.1 Género y edad

De las 1967 observaciones el 72.33% de personas eran de género masculino mientras que el 27.67% de género femenino, con la siguiente distribución según sus edades:

Figura 4.1. Grupos de edad Encuesta Multipropósito



Fuente: elaborado por el autor con base a los resultados de la EMP (2016)

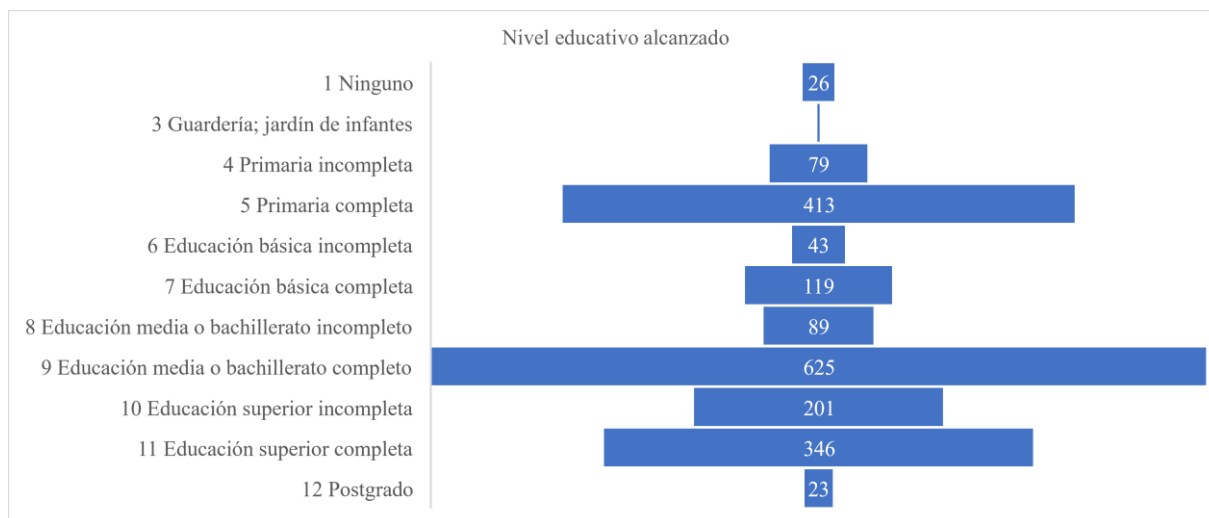
En la figura previa se puede observar que 1309 personas, que corresponden al 66.58% de la muestra a se encuentran en un rango etario considerado como adulto de 30 – 65 años y en su mayoría se encuentran entrando a una etapa de adultos mayores. También se puede observar que 498 personas se ubican en un rango de edad de 65 - 90 años lo que corresponde al 25.33% de la muestra y corresponde a una fracción importante de la muestra.

La edad promedio es de 52 años, con una desviación estándar de 16.14 una mediana de 53 y una moda de 60 años

4.1.2 Nivel educativo

En lo que respecta al nivel educativo alcanzado la muestra se distribuye de la siguiente forma

Figura 4.2. Nivel educativo alcanzado



Fuente: elaborado por el autor con base a los resultados de la EMP (2016)

Con base en la figura previa se puede observar que 625 personas encuestadas, que corresponden al 31.79% de la muestra, alcanzaron un nivel educativo de “Educación media o bachillerato completado”. También se observa que tan solo 369 (18.76%) personas han alcanzado un nivel de instrucción de tercer y cuarto nivel completado mientras que, en la proporción predominante de la muestra, que corresponde a 743 (37.79%) personas han alcanzado un nivel de instrucción de primaria e inclusive 28 personas no han alcanzado ningún nivel de educativo formal.

4.1.3 Provincia o país de origen

Más del 50% de personas encuestadas corresponden a la provincia de Pichincha. Sin embargo, 942 personas o el 47.91% de la muestra corresponden a personas migrantes de otras provincias del país que en su mayoría provienen de provincias de la Sierra central como Chimborazo (227), Cotopaxi (108) y Tungurahua (75).

Tabla 4.1. Provincia o país de origen

Provincia de origen	Frecuencia	Provincia de origen	Frecuencia
Pichincha	1024	Esmeraldas	23
Chimborazo	227	El oro	21
Cotopaxi	108	Los ríos	11
Manabí	84	Cañar	8
Tungurahua	75	Santo domingo	8

Provincia de origen	Frecuencia	Provincia de origen	Frecuencia
Carchi	67	Pastaza	4
Imbabura	66	Morona Santiago	3
Loja	61	Sucumbíos	3
Bolívar	52	Zamora Chinchipe	3
Azuay	44	Napo	2
Otro país	35	No sabe	2
Guayas	34	Orellana	1

Fuente: (ICQ 2016)

4.2 Estadísticas descriptivas de las variables empleadas

De este análisis se puede observar que en promedio de la satisfacción con la vida (general) obtiene un valor de 8.388 siendo este el dato más alto de todas las variables empleadas. En contraste, el valor correspondiente a la satisfacción con el Medio Ambiente es el más bajo con un 6.414 en la escala de satisfacción con la vida

A continuación, se presenta una matriz con las principales estadísticas descriptivas de las 8 variables que forman parte de los modelos de aprendizaje supervisado y no supervisado.

Tabla 4.2. Estadísticas descriptivas variables

Estadísticas descriptivas variables (n=1424)								
Satisfacción con la:	Educación	Salud	Economía	Vivienda	Trabajo	Vecinos	Ambiente	Vida
Moda	8.000	8.000	8.000	8.000	8.000	8.000	8.000	10.000
Mediana	7.000	8.000	7.000	8.000	8.000	8.000	7.000	8.000
Media	7.110	7.890	6.993	7.562	7.568	7.598	6.414	8.388
Desviación estándar	1.993	1.661	1.907	1.921	2.075	2.108	2.210	1.597
Mínimo	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Máximo	10.000	10.000	10.000	10.000	10.000	10.000	10.000	10.000

Fuente: elaborado por el autor con base a los resultados de la EMP (2016)

De esta misma matriz se debe tener en consideración que los valores de moda, mediana se ubican entre los datos de 7 y 8 en la escala de clasificación a excepción de la moda de la calidad de vida generalizada con un dato de 10

4.3 Autocorrelación espacial de la Satisfacción con la vida en el CHQ.

Mediante el uso del Software Geoda se obtuvo como resultado que: del análisis univariado *Moran's I* para la variable Satisfacción con la vida en el centro Histórico de Quito es de 0,092. A este valor se encuentra asociado el valor de Z de 6,0829 y el P-valor 0,001.

A continuación, se realiza la prueba de hipótesis:

- **Planteamiento de las pruebas de hipótesis nula y alternativa**

H_0 : No hay autocorrelación espacial

H_1 : Existe una autocorrelación espacial

- **Significancia o α :** 5% = 0,05
- **Estadístico de prueba Z:** 6,0536
- **P-Valor:** 0,0010

Regla de decisión del valor p:

Si P – valor < α se rechaza H_0

Si P – valor $\geq \alpha$ se rechaza H_1

P – valor < α

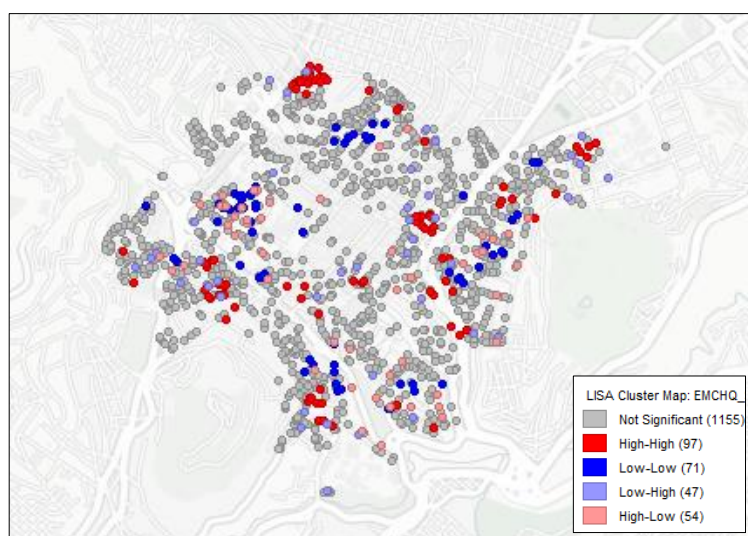
0,001 < 0,05

Se rechaza H_0 y se acepta la hipótesis alternativa que indica que existe una Autocorrelación espacial en la Satisfacción con la vida en el Centro Histórico de Quito.

En otras palabras, los patrones de asociación espacial de la Satisfacción con la vida en el Centro Histórico, de acuerdo con la evidencia estadística indican que existe una autocorrelación espacial entre los valores altos y los valores bajos en el conjunto de datos, y que los mismos se encuentran agrupados espacialmente y no responden a un proceso espacial subyacente aleatorio.

En la siguiente figura se observan los resultados del geoprocesamiento en el software Geoda:

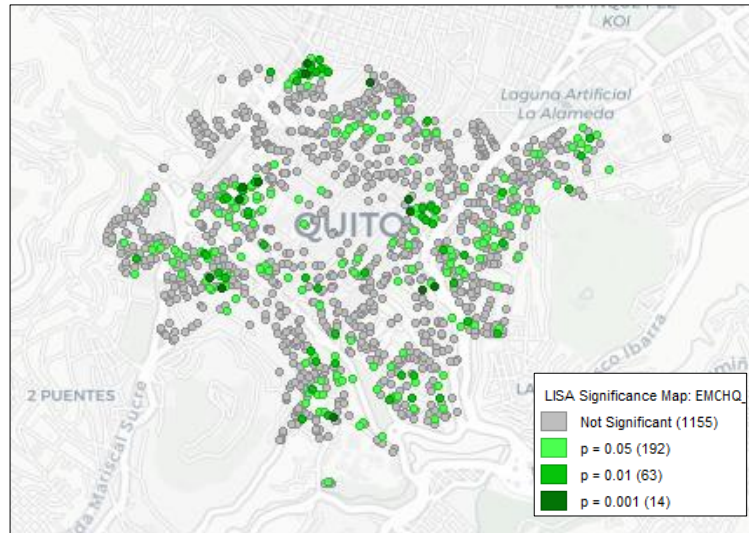
Mapa 4.1. Clústeres de LISA



Fuente: elaborado por el autor

En el mapa resultante se pueden observar zonas de interés entre valores altos – altos y valores bajos – bajos, también la combinación de escenarios entre valores bajos – altos y viceversa. Finalmente 1155 registros no representan significancia estadística para este análisis.

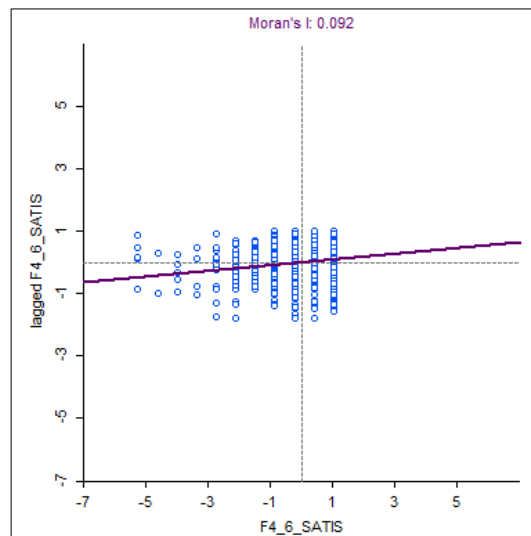
Mapa 4.2. Significancia de LISA



Fuente: elaborado por el autor

Asimismo, se obtuvo el mapa de significancia de los cuales 192 tienen una significancia de $p=0.05$, 63 registros $p=0.01$ y 14 registros $p=0.001$, este mapa es concordante a lo mostrado en la figura previa de *clusters* de LISA.

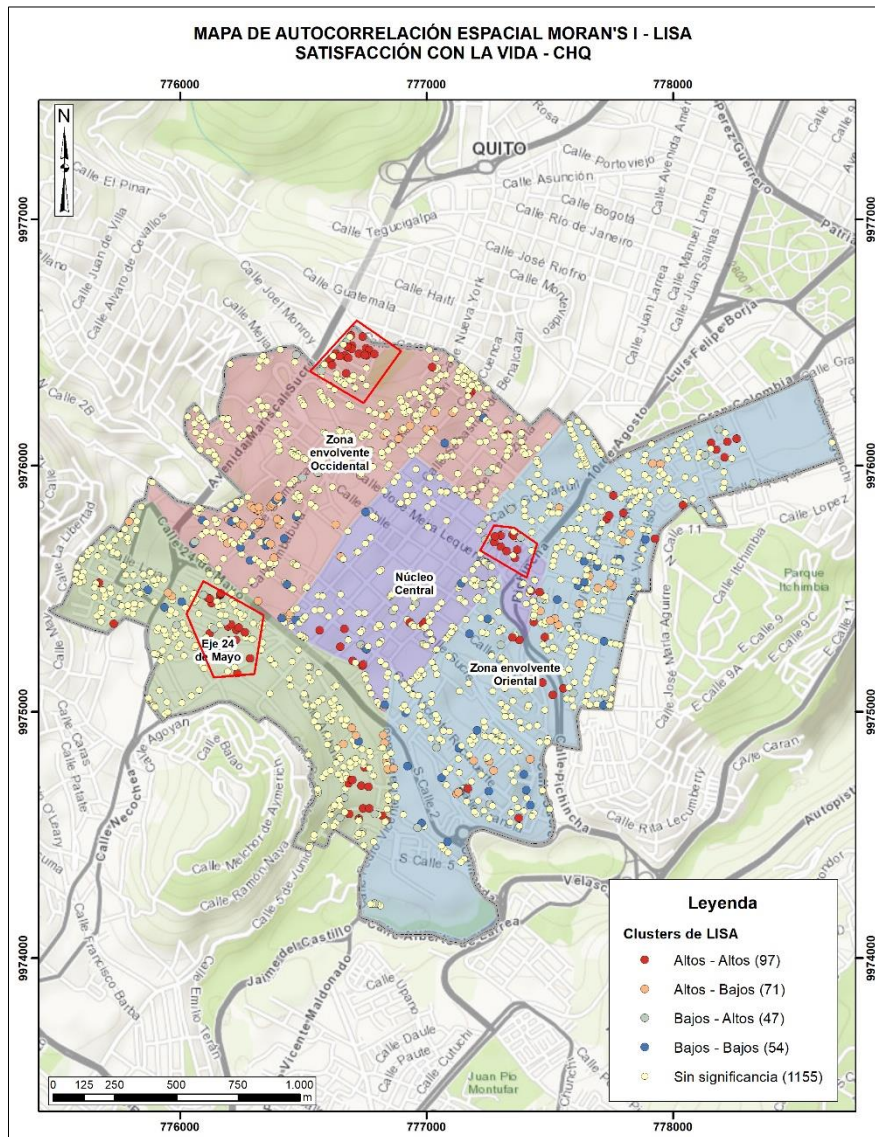
Figura 4.3. Esquema de Moran's I (Valores altos y Bajos)



Fuente: elaborado por el autor

De todas las gráficas y figuras resultantes, la que más información proporciona es la LISA Cluster Map que indica con un nivel de confianza del 95% las ubicaciones y regiones que poseen relaciones de interés Alto – Alto y Bajo – Bajo. De este modo se obtiene el siguiente mapa:

Mapa 4.3. Autocorrelación espacial Moran's I - LISA



Fuente: elaborado por el autor

De las 1424 observaciones se tiene que 97 corresponden a una categoría de valores altos de satisfacción con la vida en el CHQ se encuentran rodeados de valores similares. Existen 3 zonas (núcleos) bien definidos de esta categoría que corresponden a los ubicados en el norte de la zona envolvente occidental, en el noreste del núcleo central y en el centro del eje 24 de mayo.

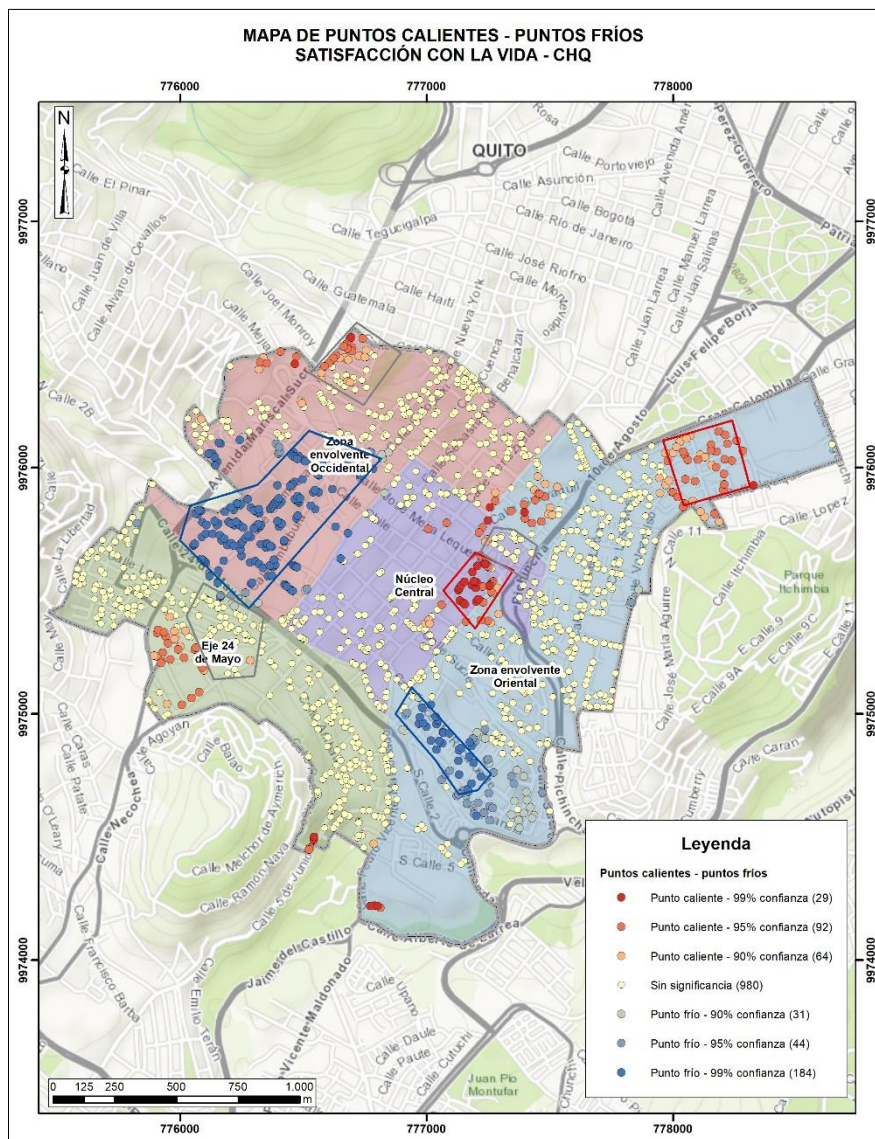
También se observa que existen datos que corresponden a esta categoría y que se distribuyen de forma dispersa en la zona envolvente oriental y en el sureste del eje 24 de mayo.

Para el caso de los valores bajos rodeados de valores bajos de satisfacción con la vida se evidencia que existen 54 observaciones que se encuentran ubicadas de forma dispersa en cada una de las zonas en la que se clasificó la Encuesta Multipropósito del año 2016.

4.3.1 Puntos calientes – Puntos fríos

A manera de comparación se desarrolló un análisis de puntos calientes (G_i^* de Getis – Ord) con el uso del software Arcgis Pro, con el objetivo de contrastar los resultados del análisis de autocorrelación espacial y determinar de forma significativa estadística la presencia de puntos calientes y puntos fríos de satisfacción con la vida en el CHQ.

Mapa 4.4. Puntos calientes – puntos fríos de la satisfacción con la vida en el CHQ.



Fuente: elaborado por el autor

A diferencia del análisis LISA se observan áreas de interés más extensas para el caso de puntos fríos de satisfacción con la vida, que se ubican en la zona envolvente occidental más cercanos a la Avenida Mariscal Sucre y una segunda zona en la parte sur de la zona envolvente oriental.

También se logra observar una zona de considerada de punto caliente de satisfacción con la vida que se ubica al oriente del núcleo central del centro histórico.

4.3.2 Interpretación del análisis de autocorrelación espacial

Esta clase de análisis espaciales hacen referencia al primer proceso estadístico al que se someten las observaciones muestrales. Dentro de un proceso de investigación, estos análisis facilitan el entendimiento del fenómeno y a través de un parámetro de significancia (confiabilidad estadística) sirven de indicador y sustento práctico para respaldar o rechazar análisis complejos posteriores.

Así, con el desarrollo de los dos análisis exploratorios de autocorrelación espacial de la satisfacción con la vida en el CHQ, se pretende determinar de que en el caso de que el fenómeno si responda a un proceso de aleatoriedad, no tendría sentido continuar con el desarrollo de la investigación y por otra parte, en el caso de que las variables no sigan un proceso aleatorio, esto significaría que están espacialmente autocorrelacionadas, y consecuentemente el análisis se podría clasificar en 2 escenarios (Anselin, 2020). Uno se llama LISA (indicador local de asociación) y el otro G general de Getis-Ord. Ambos tienen por objetivo determinar si el fenómeno sigue o no una distribución aleatoria en el espacio y son complementarios entre sí.

Cuando varias observaciones de valores altas de satisfacción con la vida se encuentran aglomeradas en una zona en específico se lo conoce como un punto caliente. Asimismo, cuando varias observaciones de valores bajos se encuentran aglomerados entre sí, se lo conoce como puntos fríos. Justamente, el resultado de este análisis es un mapa de *hot spots* y *cold spots*, que indica aquellas zonas en donde se presentan esta clase de fenómenos.

En el primer mapa (ver Figura 19), se observan 3 zonas en las que existen puntos calientes de satisfacción con la vida. Estas concentraciones se ubican en la zona envolvente occidental, en el barrio La Chilena entre las calles Tapi y José Yépez ubicada al lado oriental del túnel de San Juan y que se caracteriza por ser una zona residencial. La segunda zona se ubica en el Eje 24 de Mayo, barrio La Victoria, entre las calles Ambato y Rafael Barahona que comparte características similares con la anterior. Finalmente, la zona ubicada en el núcleo central, entre las calles Juan José Flores y Juan Pío Montufar con usos comerciales.

En perspectiva, en el segundo mapa (ver Figura 20) se identifican claramente dos puntos fríos de satisfacción con la vida que de distribuyen de forma más amplia, el primero se encuentra a lo

largo de la calle Chimborazo atravesando los barrios de San Roque y La Merced y también de forma transversal a la calle Vicente Rocafuerte y Simón Bolívar hasta la calle Chile. En este último sector es evidente un uso mixto y el deterioro del parque edilicio. El segundo punto frío se encuentra en la parte sur oriental del CHQ y se distribuye a lo largo de la calle Vicente Rocafuerte en la que prevalece un uso de carácter comercial.

4.4 Resultados de análisis de regionalización de satisfacción con la vida y variables predictoras

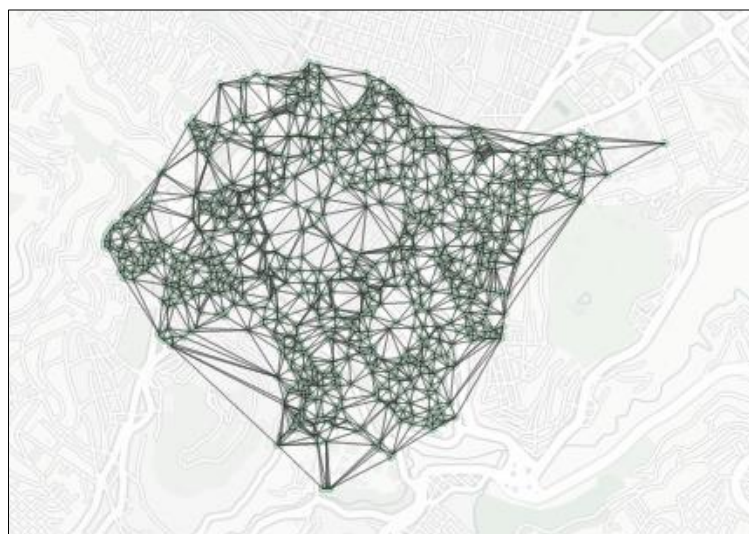
4.4.1 k-means

Como parte de este análisis se realizaron 2 modelos de regionalización espacial *k-means*: el primero únicamente ejecutado con la variable Satisfacción con la vida “F4_6” y el segundo modelo con la inclusión de las variables, educación alcanzada, salud actual, situación económica, vivienda actual, su trabajo, relación con los vecinos y calidad ambiental.

Cabe mencionar que para llevar a cabo el análisis *k-means* solo se ocupa de encontrar la similitud entre los atributos y no garantiza que los *clusters* resultantes sean espacialmente contiguos. Para solucionar este tema es necesario utilizar un archivo de pesos (ponderaciones espaciales), que aporte el criterio de contigüidad al resultado para lo cual se empleó el criterio de *Queen contiguity* del software Geoda que hace analogía a los movimientos permitidos para las piezas en un tablero de ajedrez.

De este modo se obtiene una red y entrelazada para cada uno de los puntos de la satisfacción con la vida en el CHQ que matemáticamente representa una matriz de pesos:

Figura 4.4 Gráfica de contigüidad tipo Reina Satisfacción con la vida CHQ.



Fuente: elaborado por el autor

4.4.1.1 Análisis de sensibilidad

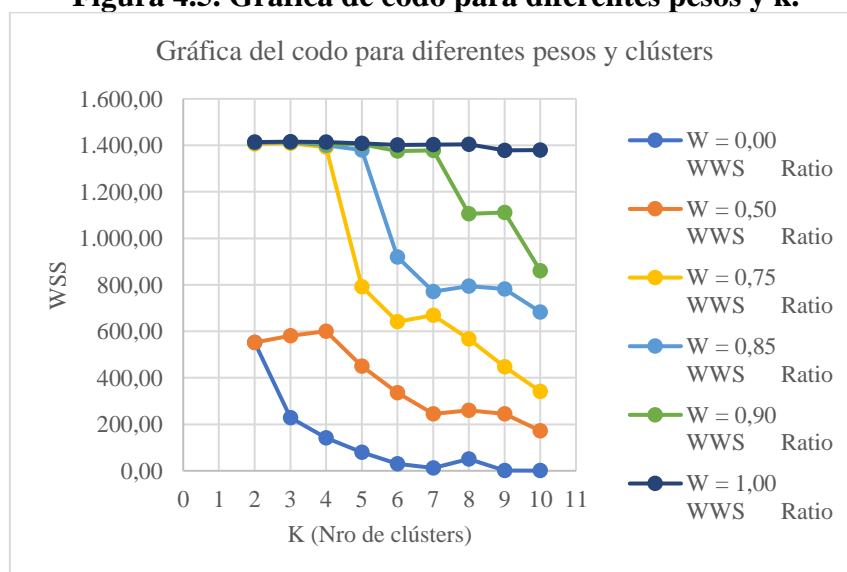
Como primer paso se determinó el valor para el peso en la configuración de los centroides geométricos, esto significa que una mejor solución para la similitud de atributos implica la nulidad o la tendencia a una matriz de pesos de cero ($w=0$, k-means puros) y una mejor solución en la similitud de la ubicación a una matriz de pesos que tienda a 1 ($w=1$, k-means con restricciones espaciales). Para lo cual de forma manual y reiterativa fue necesario encontrar una relación que permita optimizar el valor del peso (contigüidad espacial), la suma total de los cuadrados entre grupos y en el número óptimo de clústers mediante la gráfica de la regla del codo (valores subrayados). En la siguiente tabla y figura se resumen los principales resultados de este proceso:

Tabla 4.3. Optimización del modelo K-Means – datos para la gráfica del codo

Número de Clústers	Pesos											
	W = 0,00		W = 0,50		W = 0,75		W = 0,85		W = 0,90		W = 1,00	
	WWS	Ratio	WWS	Ratio	WWS	Ratio	WWS	Ratio	WWS	Ratio	WWS	Ratio
2	552,28	0,61	552,28	0,61	1405,97	0,01	1411,40	0,01	1412,2	0,01	1414,40	0,01
3	228,36	0,84	580,38	0,59	1409,19	0,01	1413,71	0,01	1414,29	0,00	1415,90	0,00
4	141,86	0,90	600,45	0,58	1392,65	0,02	1399,90	0,01	1405,76	0,01	1414,10	0,01
5	78,96	0,94	450,61	0,68	790,939	0,44	1379,88	0,02	1404,27	0,01	1408,38	0,01
6	30,30	0,98	335,68	0,76	640,935	0,55	919,89	0,22	1375,3	0,03	1401,81	0,01
7	12,37	0,99	244,82	0,83	668,805	0,53	771,38	0,27	1377,93	0,02	1402,64	0,01
8	50,96	1,00	260,66	0,82	567,752	0,60	794,11	0,26	1106,06	0,02	1404,03	0,01
9	1,43	1,00	245,16	0,83	447,082	0,69	781,39	0,43	1111,07	0,06	1378,59	0,03
10	0,56	1,00	172,22	0,88	341,691	0,76	683,49	0,49	860,516	0,07	1379,54	0,03

Fuente: elaborado por el autor

Figura 4.5. Gráfica de codo para diferentes pesos y k.



Fuente: elaborado por el autor

Dónde

- WSS: *With in Sum of Squares* es la distancia total de los puntos de datos desde sus respectivos centroides de clúster
- BSS: *Between Sum of Squares* es la distancia ponderada total de varios centroides de clúster a la medida global de datos
- TSS: *Total Sum of Squares* es la distancia total de los puntos de datos desde la media global de datos, para un conjunto de datos dado, este valor es una constante
- R²-Ratio: BSS/TSS es la varianza total explicada por el ejercicio de agrupamiento.

Como regla de decisión se espera que los clústers sean compactos (WSS debería ser más bajo) y homogéneos (BSS debería ser más alto), mientras que con el R² se puede evaluar el progreso entre las diferentes iteraciones y en este sentido se elige el valor más alto que cumpla los criterios de contigüidad y gráfica de codo. Y que para este caso particular es el de k=6 con W=0.85 y R²=0.22. Todo este análisis permite identificar de forma específica la diferenciación espacial y diferenciación entre en conjunto de datos utilizados para este algoritmo.

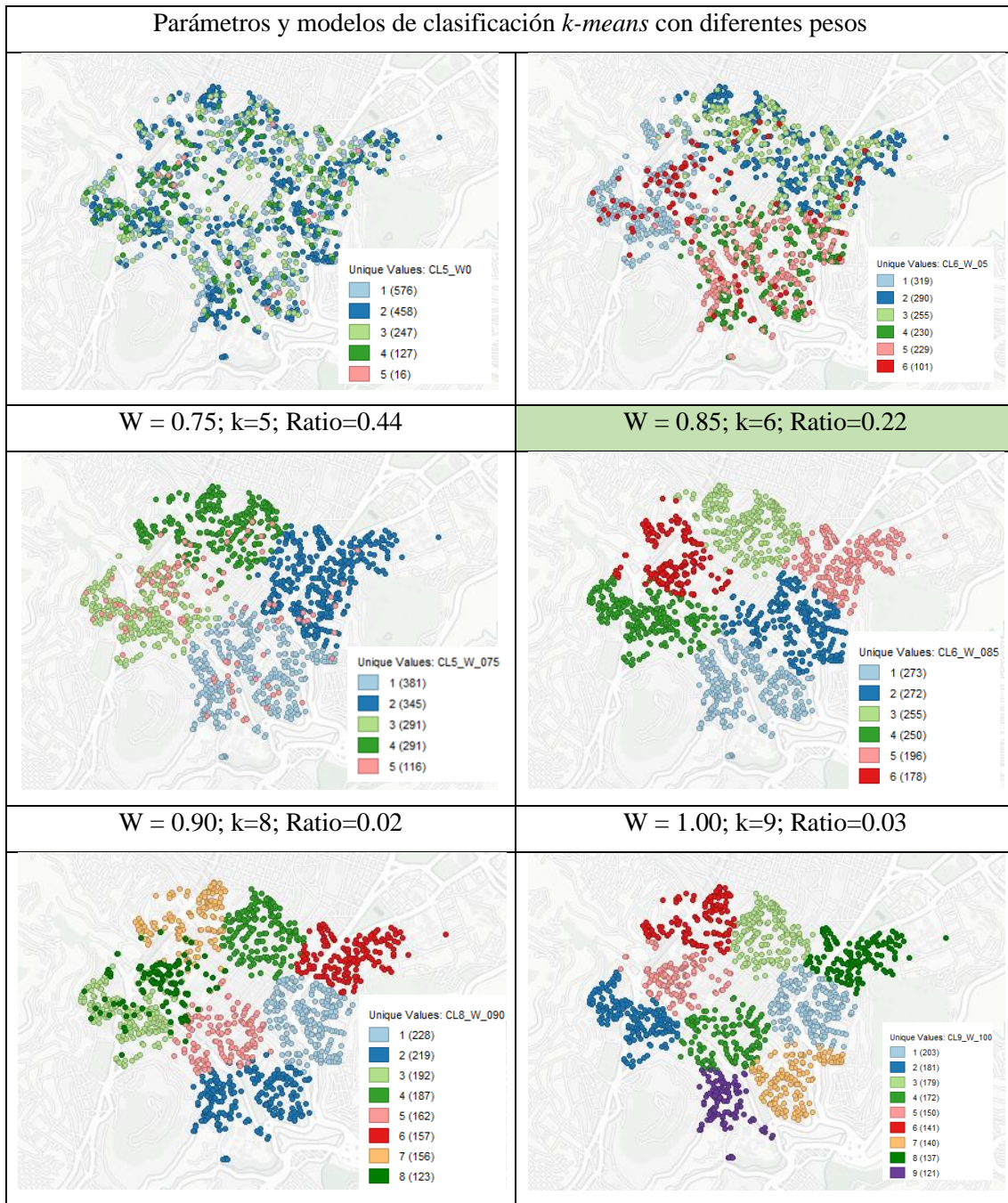
4.4.1.2 K-means contigüidad espacial para diferentes pesos

El análisis de optimización se condensa en la representación espacial de los clústers y justamente la generación de modelos de clasificación cuyos parámetros cumplan y brinden las mejores prestaciones para una clasificación correcta. Además de estos parámetros la contigüidad espacial se evalúa de forma visual cuando cada uno de los datos se encuentra clasificado de tal forma en que cada observación corresponda únicamente a un clúster cuya delimitación espacial sea bien definida.

En la siguiente tabla se presentan los principales modelos con la mejor clasificación según su peso, gráfica de codo y R² (Ver valores subrayados en la Tabla 9 y Figura 22).

Tabla 4.4. Parámetros y modelos de clasificación k-means

Parámetros y modelos de clasificación <i>k-means</i> con diferentes pesos	
W = 0.00; k=5; Ratio=0.94	W = 0.50; k=6; Ratio=0.76



Fuente: elaborado por el autor

4.4.1.3 K-means modelo de clasificación para satisfacción para la vida

Partiendo de este análisis visual y estadístico se obtuvo que el mejor modelo de clasificación corresponde a aquel que tiene una matriz de peso de 0,85 ($W=0,85$) con 6 clústers ($k=6$) y un R^2 de 0,22. El mapa resultante se clasifica de forma óptima y se evidencia que cada clúster se conforma de la siguiente cantidad de observaciones

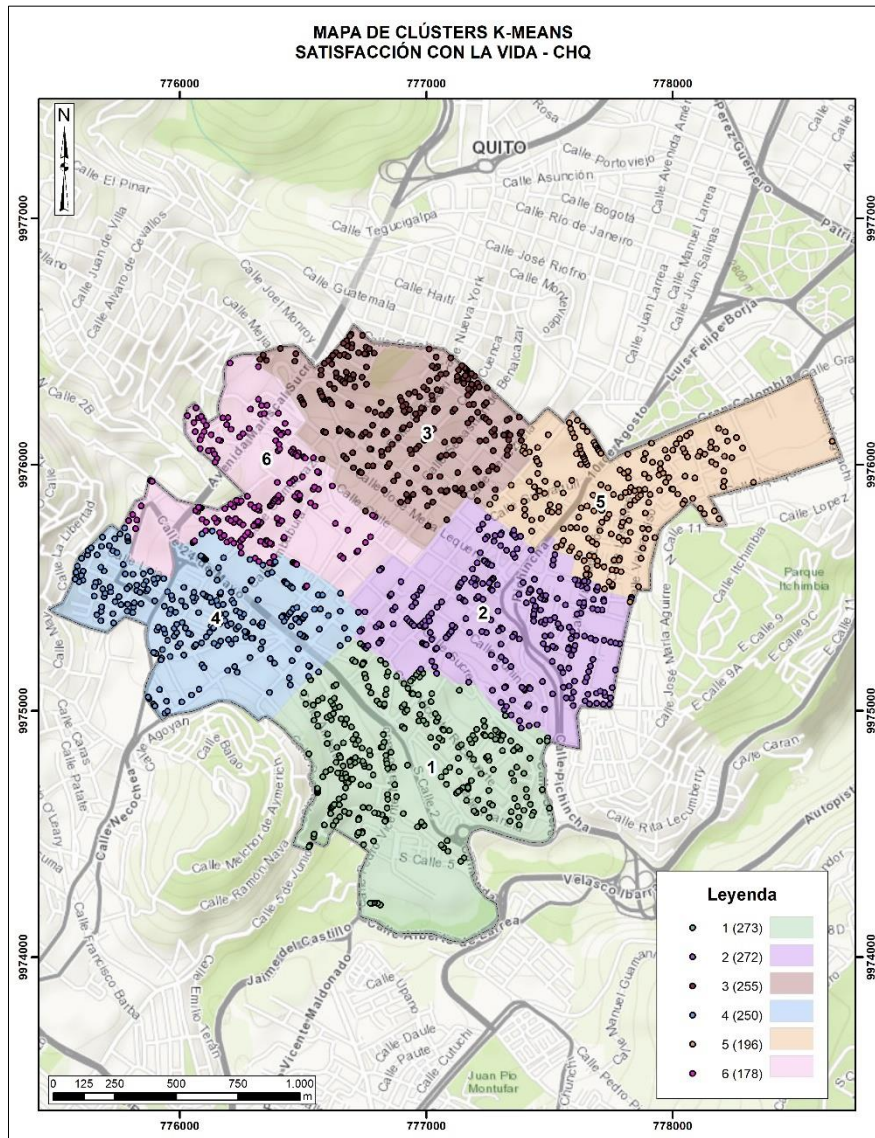
- Cluster 1: 273
- Cluster 2: 272

- *Cluster 3*: 255
- *Cluster 4*: 250
- *Cluster 5*: 196
- *Cluster 6*: 178

Es necesario indicar que este modelo de clasificación cumple con los parámetros de optimización. Sin embargo, su resultado es perfectible ya que entre los Clúster 4 y 6 existen 5 observaciones que se encuentran al límite de su región espacial.

Además, se observa una semejanza entre las zonas que fue clasificada la encuesta multicriterio del 2016. Por ejemplo, los clústers 4 y 1 se ubican en la zona conocida como eje 24 de mayo. Parte del clúster 1 y los clústers 2 y 5 corresponderían a la zona envolvente oriental y de forma similar lo que ocurre con los clúster 4, 6 y 3 que corresponderían a la zona envolvente occidental. Sin embargo, de todas la iteraciones realizadas en ninguna se distingue de forma clara una que clasifique la zona de núcleo central. A continuación, se presenta el mapa resultante:

Mapa 4.5. Clústeres k – means Satisfacción con la vida en el CHQ



Fuente: elaborado por el autor

4.4.1.4 K-means modelo de clasificación para satisfacción para la vida y variables independientes.

Para este modelo se tuvo en cuenta la misma matriz de pesos del modelo previo con el objetivo de generar una comparación entre resultados ($W=0.85$). Para este modelo además de la satisfacción con la vida se incluyeron variables como:

Educación alcanzada, salud actual, situación económica actual, su vivienda actual, su trabajo, relación con los vecinos, calidad ambiental.

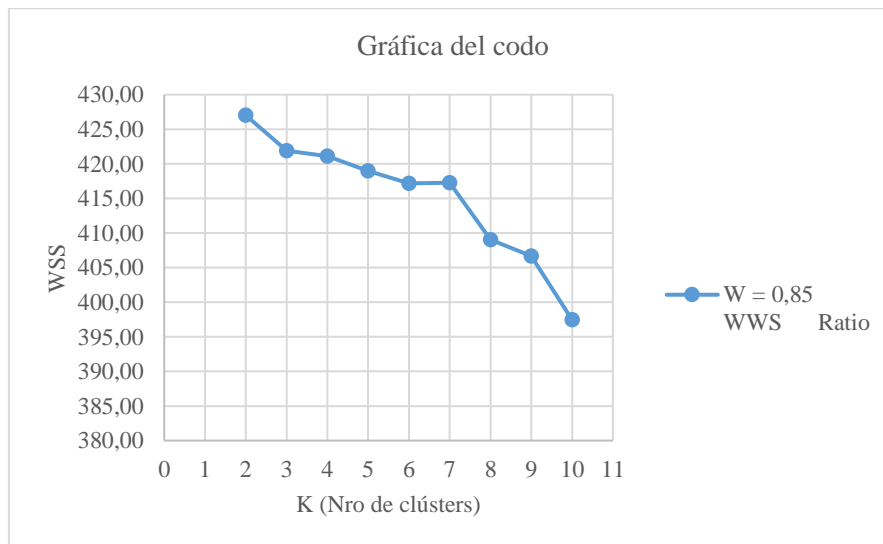
Como primer paso se desarrolló en análisis de la gráfica del codo en la que se observa el punto de inflexión en $k = 6$, los datos se resumen en la siguiente tabla y figura.

Tabla 4.5. Modelo *k-means* (todas las variables) – datos para la gráfica del codo

Número de clústers	W = 0,85	
	WSS	Ratio
2	427,02	0,01
3	421,90	0,02
4	421,13	0,02
5	419,00	0,03
6	417,20	0,03
7	417,26	0,03
8	409,03	0,05
9	406,66	0,05
10	397,46	0,08

Fuente: elaborado por el autor

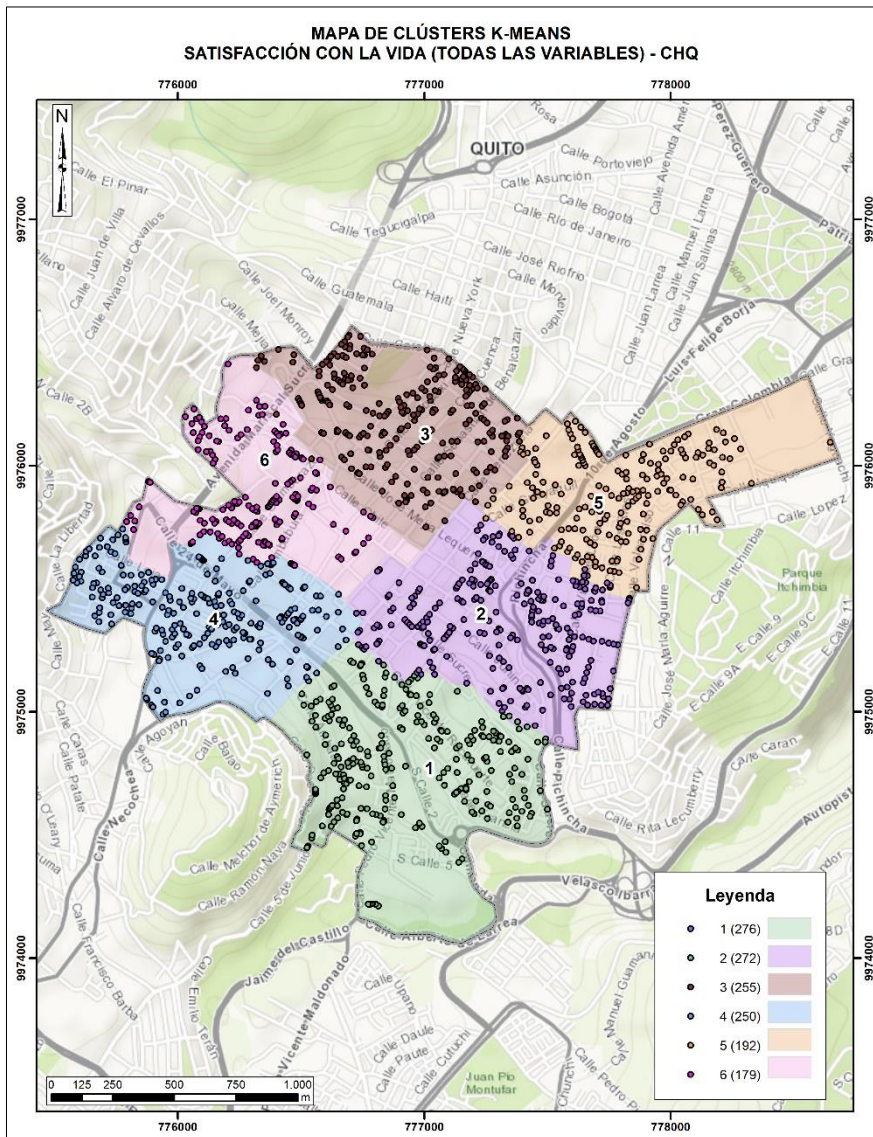
Figura 4.6. Gráfica de codo para *k-means* (todas las variables)



Fuente: elaborado por el autor

Este modelo de clasificación a pesar de compartir ciertas semejanzas como el número de clústers y una gráfica de codo similar se diferencia por su valor de R^2 que solo alcanza el valor de 0,0304815 lo que representa una disminución la varianza total explicada por el ejercicio de agrupamiento respecto al modelo que se desarrolló únicamente con la variable de satisfacción con la vida. En otras palabras, este modelo clasifica de forma correcta la contigüidad espacial pero la clasificación de sus datos no es tan buena como el modelo previo. En el siguiente mapa se observa el mapa con los clústers para este modelo:

Mapa 4.6. Clusters k – means Satisfacción con la vida (todas las variables) en el CHQ



Fuente: elaborado por el autor

Respecto a este resultado no se observan diferencias significativas respecto al modelo que fue realizado únicamente con la variable satisfacción con la vida y todas las variables, se mantiene el mismo número de registros para cada clúster resultante.

4.4.2 SKATER

Este enfoque toma en cuenta de forma explícita las restricciones de contigüidad en el proceso de agrupamiento, de forma operacional (sus parámetros de configuración en el software) es un proceso similar al realizado con el algoritmo *k-means* con la consideración que este algoritmo genera un árbol de expansión mínimo y depende de una matriz de pesos tipo reina cuyo peso no puede ser modificado ($W=1$). Dicho árbol se poda en sus bordes con la eliminación de aquellas

ramas que aumentan la disimilitud entre grupos. Además de esa consideración se debe tener en cuenta que para lograr un resultado homogéneo tanto en el número de observaciones por cada clúster y un buen criterio de contigüidad es necesario determinar el tamaño mínimo de observaciones que debe conformarse cada clúster. Para fines prácticos se ha elegido el 5% de los valores de la variable satisfacción con la vida (66 observaciones) y que justamente corresponde a la variable espacialmente extensiva para este estudio.

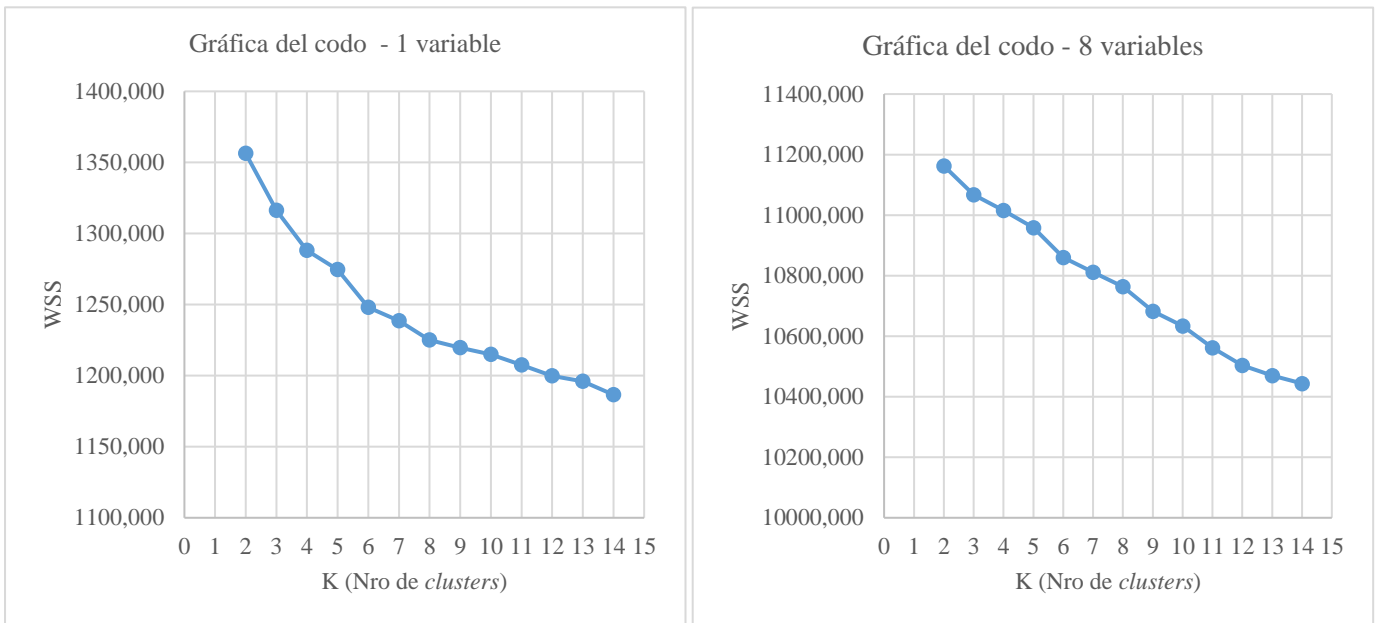
Al igual que en el algoritmo *k-means* es necesario realizar un análisis del WSS (gráfica del codo) y el Ratio de la varianza R^2 . Este proceso se detalla a continuación para el modelo de clasificación con 1 y con 8 variables, los resultados del WSS y ratio se resumen en la siguiente tabla y figura.

Tabla 4.6. Modelo Skater (1 y 8 variables) – datos para la gráfica del codo

Número de <i>clusters</i>	1 variable		8 variables	
	WWS	Ratio	WWS	Ratio
2	1356,320	0,047	11163,000	0,019
3	1316,190	0,075	11067,800	0,028
4	1288,030	0,095	11015,700	0,032
5	1274,580	0,104	10958,400	0,037
6	1248,090	0,123	10859,800	0,046
7	1238,640	0,130	10811,400	0,050
8	1224,990	0,139	10763,900	0,054
9	1219,670	0,143	10681,900	0,062
10	1214,870	0,146	10633,700	0,066
11	1207,360	0,152	10561,900	0,072
12	1199,890	0,157	10503,900	0,077
13	1195,900	0,160	10469,600	0,080
14	1186,550	0,166	10443,800	0,083

Fuente: elaborado por el autor

Figura 4.7. Gráficas de codo para skater (1 y 8 variables)

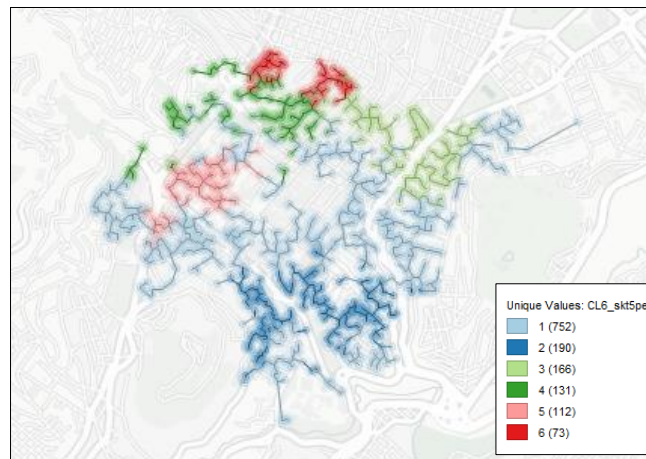


Fuente: elaborado por el autor

En las gráficas previas de forma visual se distingue de forma leve un codo entre los valores de $k=5, 6$ y 7 , de este modo resulta necesario tomar en cuenta aquel valor del ratio que tenga una mayor diferencia entre observaciones seguidas y que corresponde a los valores de $k = 5$ y $k = 6$. De esta forma se determinó que para este modelo de clasificación el número óptimo clústers es 6 .

Posteriormente se obtiene la gráfica del árbol de expansión mínima y la clasificación correspondiente para 6 clústers que se presentan en la siguiente figura:

Figura 4.8. Árbol de expansión mínima SKATER



Fuente: elaborado por el autor

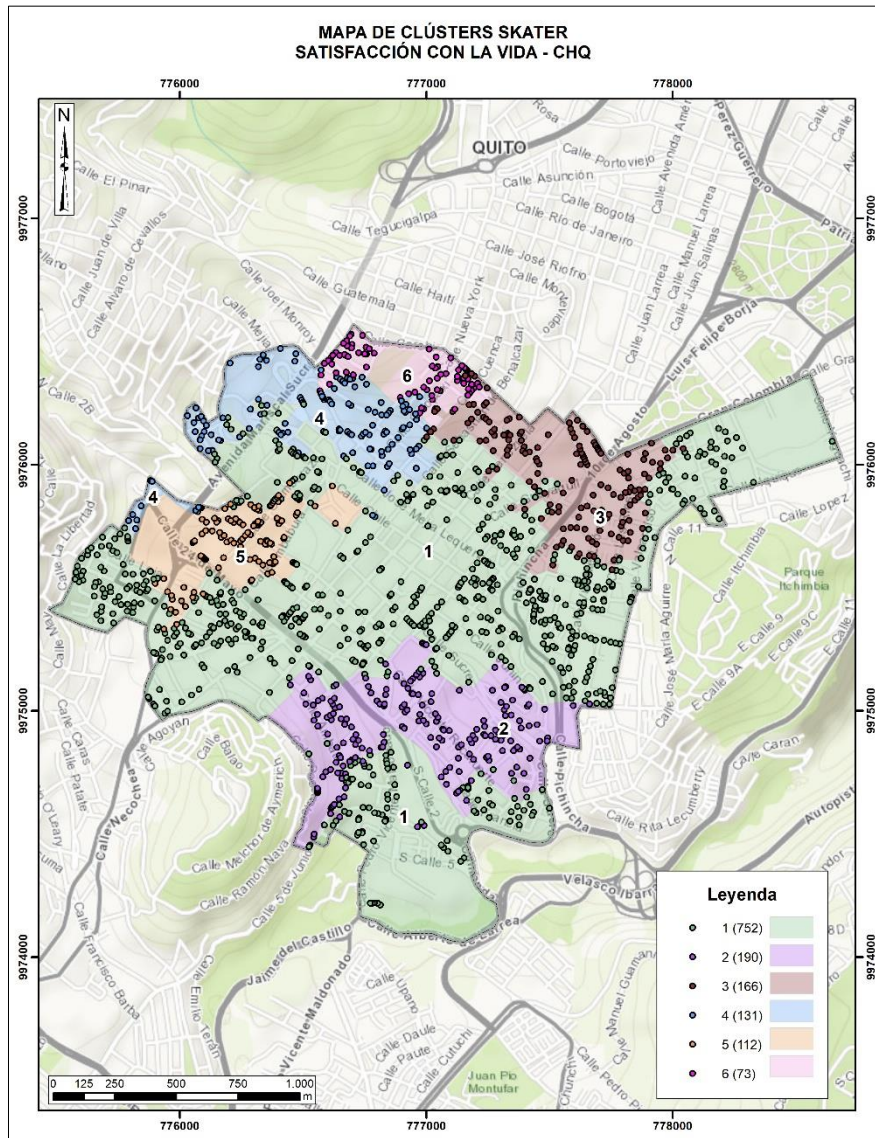
El método *skater* tiene un enfoque divisivo ya que comienza con un solo grupo y el mismo encuentra la división óptima en subgrupos hasta satisfacer el valor ingresado para k . Debido a esta naturaleza jerárquica una vez que se corta el árbol en un punto, todos los cortes posteriores se limitan a los ramales resultantes como se puede apreciar en la zona sur de la figura previa, ya que existe una discontinuidad con los valores del clúster número 1. De este modo, se puede decir que aquellas observaciones pertenecientes al *cluster* 1 ubicados en la rama sur no pueden volver a conectarse a sus ramas adyacentes y las observaciones del clúster número 2 se convierten en una especie de isla para este modelo de clasificación, esto a su vez se convierte en una limitación de este algoritmo.

Los mapas resultantes se clasifican con 6 *clusters* y se evidencia nuevamente la jerarquía del árbol de expansión mínima ya que cada clúster va reduciendo el número de observaciones a medida que aumenta su valor de k

- *Cluster* 1: 752
- *Cluster* 2: 190
- *Cluster* 3: 166
- *Cluster* 4: 131
- *Cluster* 5: 112
- *Cluster* 6: 73

A continuación, se presentan los mapas resultantes para 1 y 8 variables:

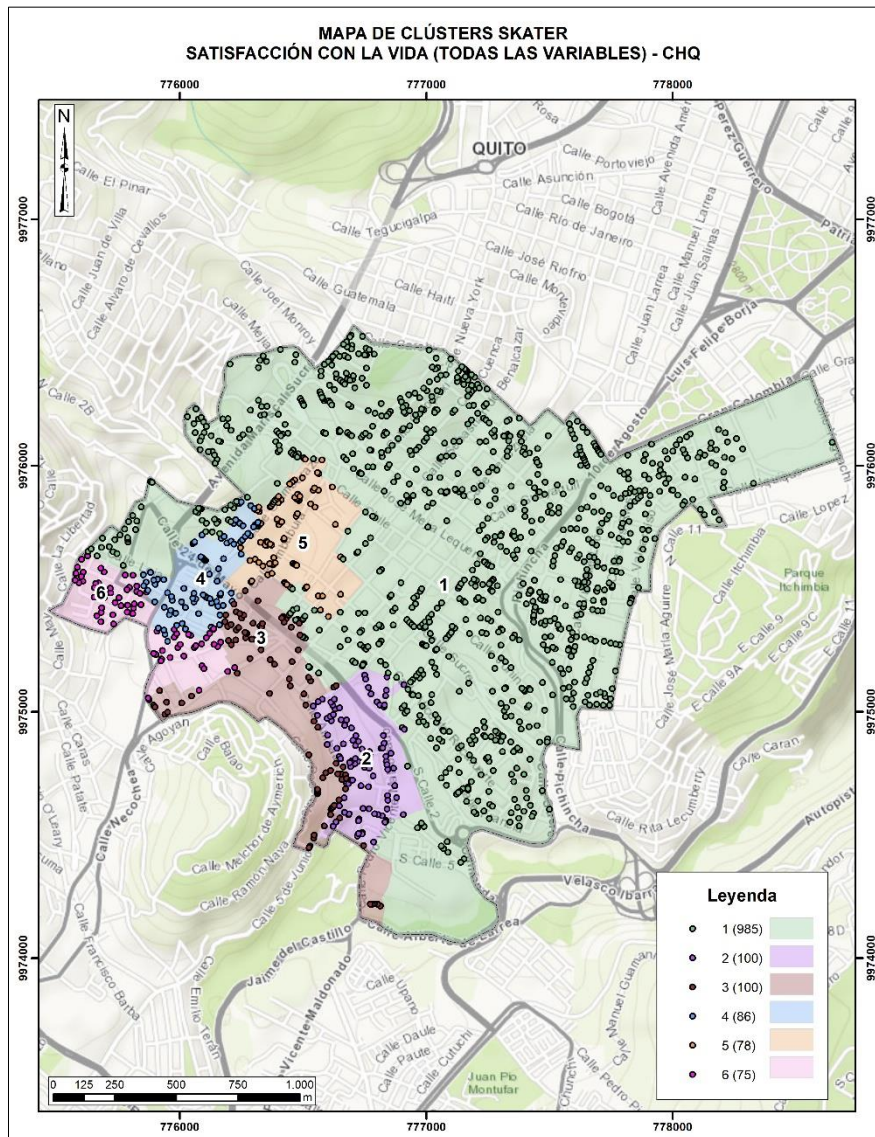
Mapa 4.7. Clústers SKATER Satisfacción con la vida en el CHQ



Fuente: elaborado por el autor

El modelo de clasificación *skater* con 8 variables difiere completamente del que se realiza únicamente con la variable satisfacción con la vida, la cantidad de observaciones no se distribuyen de forma homogénea y el clúster número 1 abarca el 69.17% de todas las observaciones. Sin embargo, este modelo respeta completamente las restricciones de contigüidad y no se presenta ninguna clase de isla en su clasificación.

Mapa 4.8. Clústers Skater (8 variables) Satisfacción con la vida en el CHQ



Fuente: elaborado por el autor

4.4.3 Hierarchical clustering

Previo a la ejecución del algoritmo, se debe procesar un dendrograma el mismo que permite interactuar con el programa para determinar el número de clústers en los cuáles se puede clasificar los datos. En este sentido y con el objetivo de obtener *clusters* bien equilibrados (en el número de observaciones) se empleó la matriz de pesos y el método *Ward* para la determinación del centroide de agrupamiento. También se evaluaron los parámetros WWS y radio de

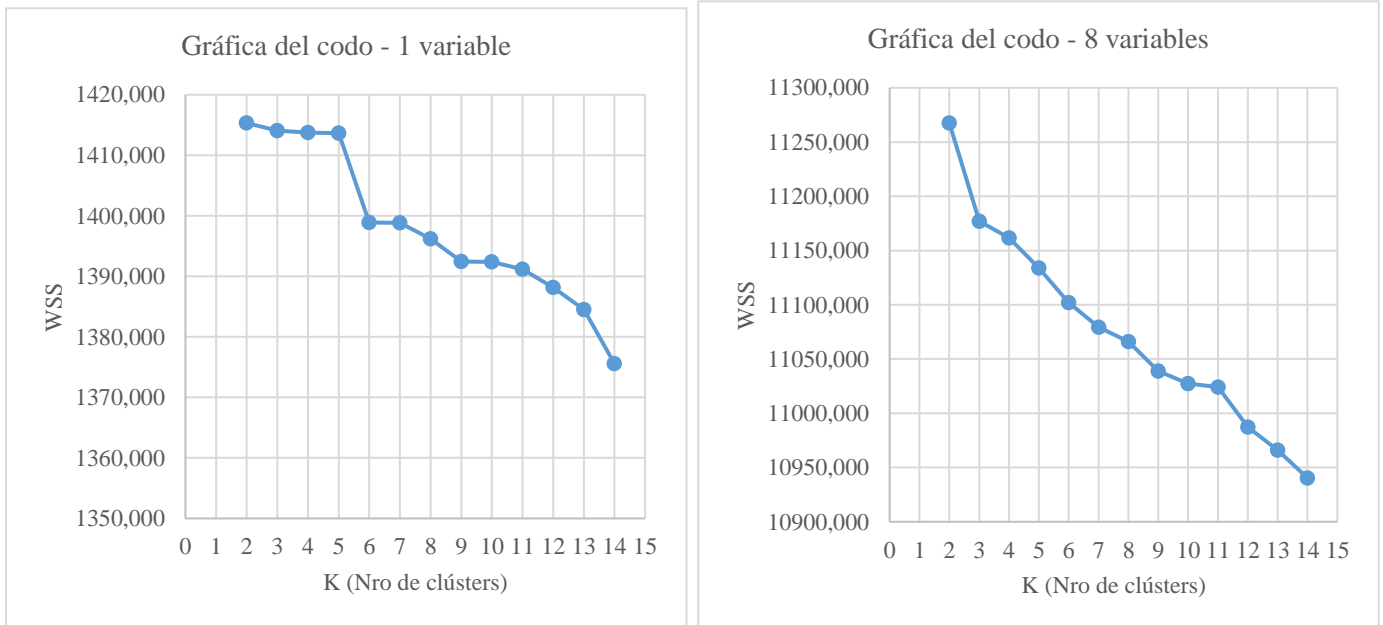
convergencia para determinar la cantidad de *clusters* siguiendo el principio del método del codo y se detalla en la siguiente tabla y figura:

Tabla 4.7. Modelo *clúster* jerárquico (1 y 8 variables) – datos para la gráfica del codo

Número de clústers	1 Variable		8 variables	
	WWS	Ratio	WWS	Ratio
2	1415,320	0,005	11267,600	0,010
3	1414,050	0,006	11176,900	0,018
4	1413,770	0,006	11161,600	0,020
5	1413,660	0,007	11134,000	0,022
6	1398,900	0,017	11102,200	0,025
7	1398,850	0,017	11079,400	0,027
8	1396,200	0,019	11066,200	0,028
9	1392,440	0,021	11038,900	0,030
10	1392,370	0,022	11027,300	0,031
11	1391,190	0,022	11024,300	0,032
12	1388,150	0,024	10987,200	0,035
13	1384,520	0,027	10966,000	0,037
14	1375,570	0,033	10940,400	0,039

Fuente: elaborado por el autor

Figura 4.9. Gráficas de codo para Skater (1 y 8 variables)



Fuente: elaborado por el autor

Con base en el análisis previo se determinó que el número de *clusters* óptimo para el modelo con la variable satisfacción con la vida es de $k=6$, al igual que el modelo *k-means* y *skater*. Sin

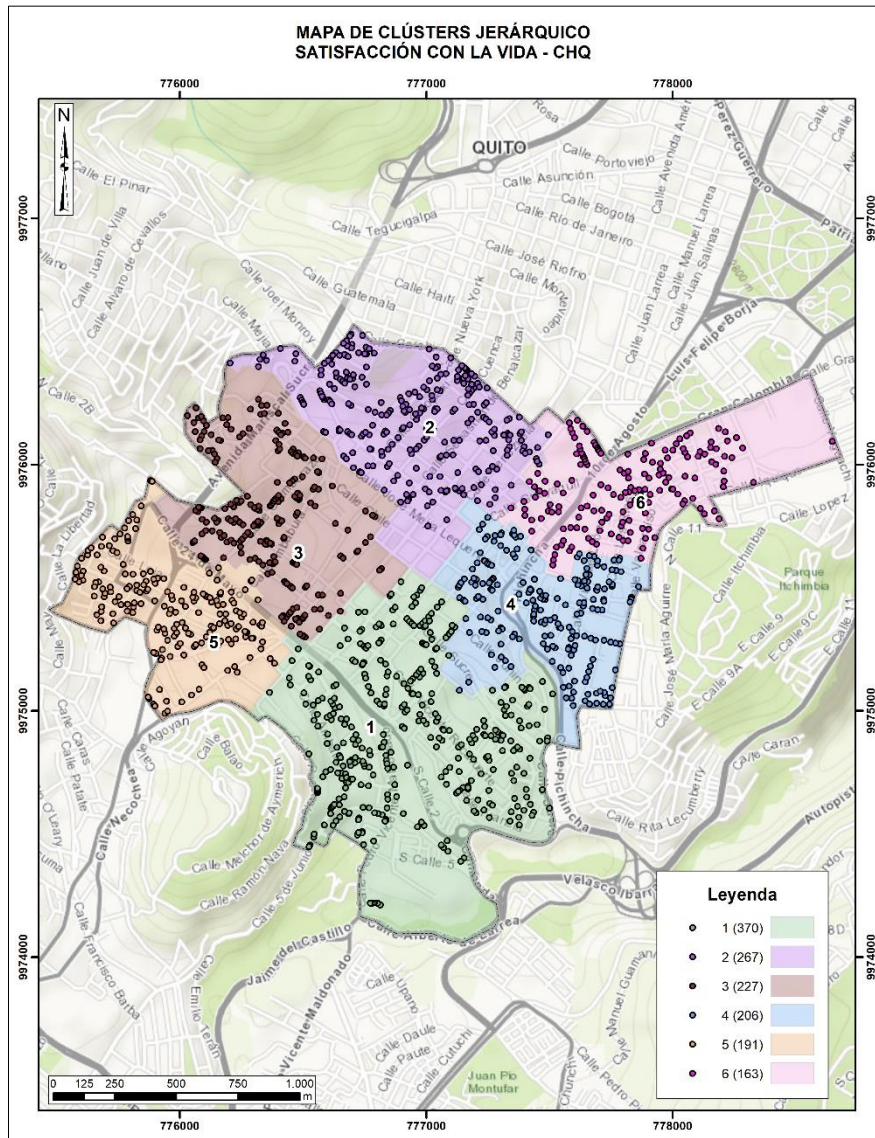
embargo, para el modelo en el que se emplean las 8 variables se pudo observar que los datos sufren un corte para su segmentación entre $k = 10$ y $k = 11$, ante esta situación se decidió tomar como referencia $k=10$ para su posterior representación gráfica.

En lo que respecta al modelo espacial, se observa que el algoritmo para la variable satisfacción con la vida se comporta de forma similar que mal modelo Skater, porque parte de un primer conglomerado de contiene el mayor número de observaciones, y a partir de este los siguientes clústers van reduciendo su cantidad hasta cumplir con el parámetro de $k=6$ ingresado previamente. A continuación, se presenta el número de observaciones por clústers

- *Cluster 1*: 370
- *Cluster 2*: 267
- *Cluster 3*: 227
- *Cluster 4*: 206
- *Cluster 5*: 191
- *Cluster 6*: 163

La representación espacial del método de clasificación jerárquico presenta un modelo relativamente bien distribuido y homogéneo tratando de mantener el número de observaciones constantes y comparte semejanzas importantes entre el modelo *k-means* en la forma que segmenta las observaciones. En este marco, se puede indicar que tanto el modelo jerárquico como el *k-means* se respaldarían entre sí lo que otorgaría robustez al análisis.

Mapa 4.9. Clústers jerárquico Satisfacción con la vida en el CHQ

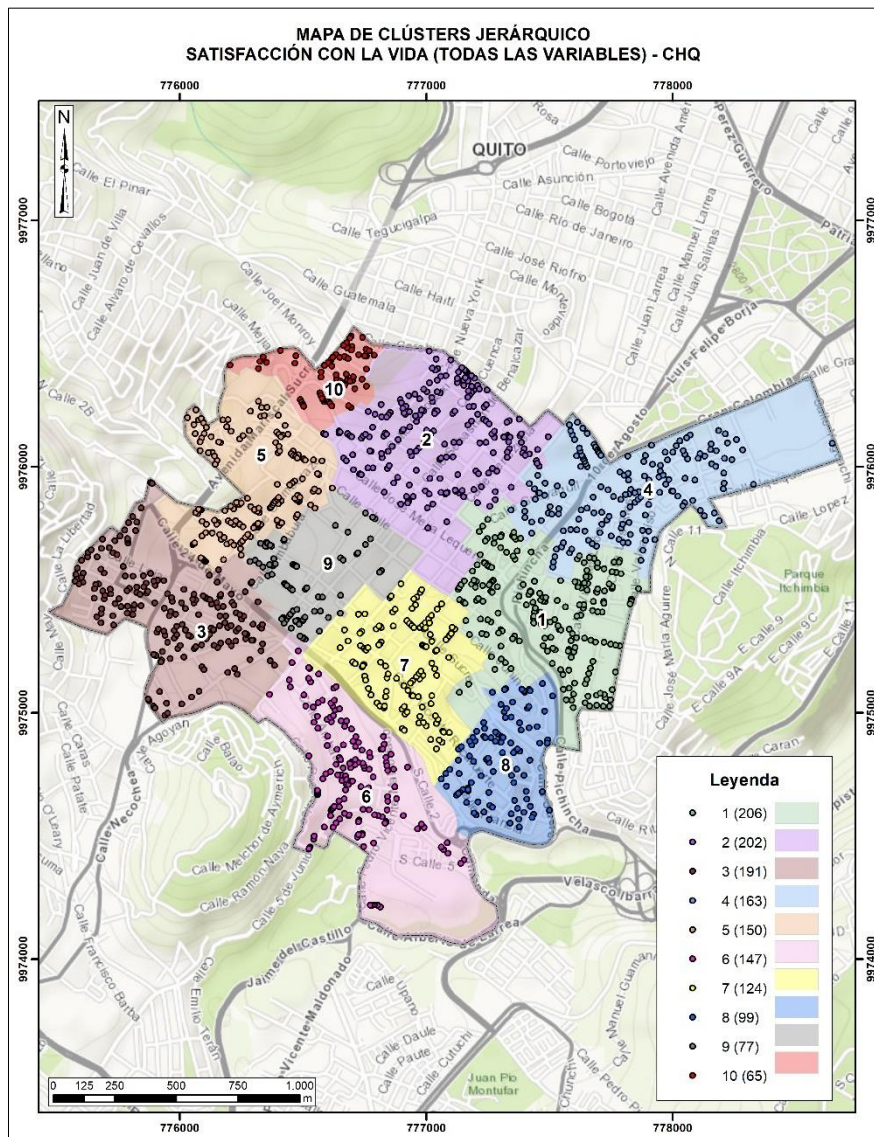


Fuente: elaborado por el autor

En lo que respecta al modelo de clasificación con todas las variables el modelo jerárquico es el que más difiere en la cantidad del número de grupos óptimos para su representación, ya que según el método del codo sugiere que sea $k=10$. Razón por la cual el modelo genera clústers muy heterogéneos en la cantidad de observaciones, ya que en la práctica tienden a estar con datos menores 150 a partir del clúster número 5.

Lo que ocasiona con esto es que el modelo tenga un excedente en el número de clústers que no aporta significativamente a su representación espacial y más bien se creen segmentos de muy pocas observaciones como el caso del clúster número 10 que posee 65 datos.

Mapa 4.10. Clusters jerárquico (8 variables) Satisfacción con la vida en el CHQ



Fuente: elaborado por el autor

4.5 Resultados de regresión con aprendizaje supervisado

4.5.1 Neural Network

Los parámetros de configuración de aprendizaje del algoritmo en el software JASP se calibraron con los siguientes valores:

- Información de entrenamiento 80%
- Información de validación 20%
- Algoritmo de aprendizaje: RPROP+
- Número máximo de repeticiones: 100000
- Funciones de activación: Lineal, logística binaria y binaria.

Se realizaron un total de 3 modelos de redes neuronales artificiales con diferentes funciones de activación como logística binaria, lineal y tangente inversa. A continuación, se describen los principales resultados de cada una de ellas.

4.5.1.1 Evaluación de parámetros estadísticos

La forma más objetiva de evaluar el rendimiento de los modelos es a través de la comparación de sus métricas de evaluación estadísticas, razón por la cual se ha elaborado una tabla que resume los principales indicadores resultantes como RMSE que indica la desviación estándar de los valores residuales o mejor considerados errores de predicción y también el R^2 que indica qué tanto explican las variables independientes la variabilidad de su variable dependiente en términos de la varianza. A continuación, se presenta la tabla resultante

Tabla 4.8. Métricas de evaluación rendimiento estadístico de los modelos de redes neuronales

Métricas de evaluación	Lineal	Logístico Sigmoides	Tangente inversa
MSE	0.647	0.763	2.003
RMSE	0.804	0.873	1.429
MAE / MAD	0.589	0.630	1.229
MAPE	96.7%	98.59%	87.78%
R^2	0.366	0.238	0.197
Capas ocultas	6	3	4
Nodos	26	19	32

Fuente: elaborado por el autor

En lo que respecta al modelo lineal, se observa que tiene un valor de RMSE de 0.804, siendo este dato el de menor de los tres modelos ejecutados empleando redes neuronales. Este dato nos indica el promedio de qué tan cerca están los puntos observados de los valores predichos por el modelo, manteniendo sus unidades. En este sentido, esta métrica estadística presenta una medida de precisión entre el dato real observado y el dato predicho considerando la escala de calificación de satisfacción con la vida (1 nula satisfacción con la vida y 10 mucha satisfacción con la vida)

A pesar de que lo ideal es un valor cercano a cero, el RMSE del modelo de redes neuronales con función de activación lineal nos indica que (a manera de ejemplo), teniendo un valor de 9 como valor observado esperado, el modelo puede darnos como resultado un dato de 9.804 u 8.196 siendo estos valores aceptables, considerando que el tipo de fenómeno que se ha modelado es una percepción subjetiva. Este mismo análisis puede realizarse a las funciones logística sigmoide y tangente inversa lo que connotaría un nivel de precisión inferior al del modelo lineal.

A diferencia de los modelos estadísticos convencionales, los modelos de redes neuronales, al tener arreglos matriciales iterativos no presentan valores de significancia (*p-value*) para cada una de las variables dependientes. A pesar de esto, los modelos de redes neuronales si presentan su valor de R^2 y se lo puede analizar de forma equivalente como si se tratase de un modelo estadístico de regresión.

Teniendo en cuenta que un valor de R^2 de 1 indica un valor de bondad de ajuste perfecta, mientras que un valor cercano a 0 indica modelo de ajuste imperfecto se puede analizar los resultados de la siguiente forma: El modelo lineal presenta un valor de R^2 de 0.366 siendo este resultado el mayor de todos los datos en comparación a los modelos de redes neuronales logístico sigmoide y tangente inversa con valores de 0.238 y 0.197 respectivamente.

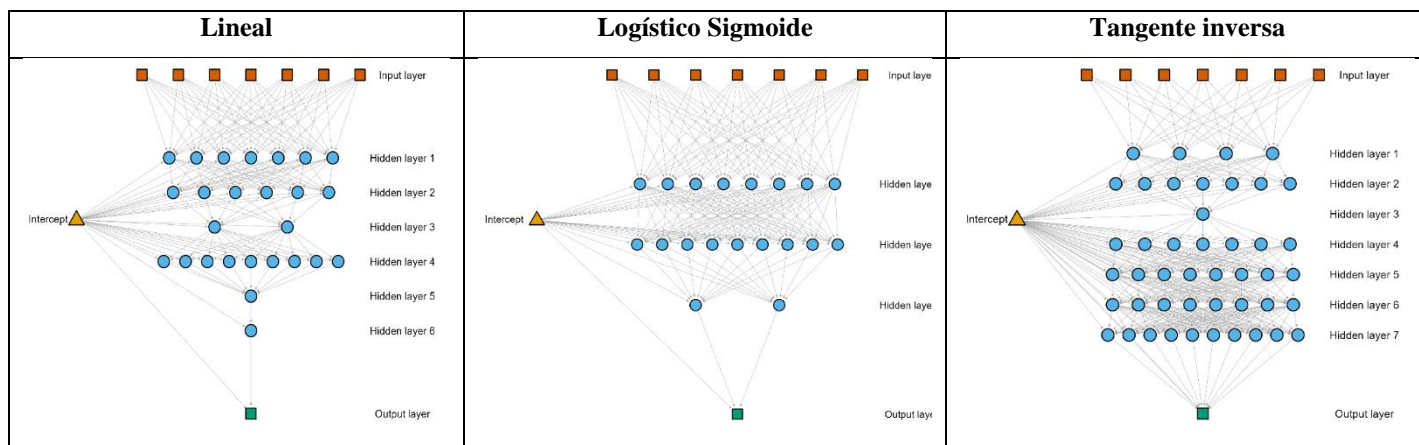
Generalmente cuando se tratan variables en lo que respecta a leyes físicas, relacionados procesos industriales, mecánica, procesos econométricos se esperan procesos en los que los valores de R^2 sean muy altos (0.95) o más. Sin embargo, en lo que respecta a análisis social y socio espacial y de forma específica al modelamiento de variables subjetivas se consideran como aceptables valores de R^2 con valores inferiores a 0.5 y hasta 0.25. Así, interpretando los resultados del modelo redes neuronales con función de activación lineal se puede indicar que basados en el coeficiente de determinación del modelo lineal, un 36.6% de la variación de la satisfacción con la vida es explicada por la satisfacción con la salud, trabajo y economía, su relación con los vecinos y la percepción del medio ambiente. Cada una de estas variables independientes aportan en mayor y menor magnitud al modelo, esta particularidad será analizada en el acápite de importancia relativa de las variables predictoras.

4.5.1.2 Diseños de las redes

Otro aspecto de interés al evaluar para evaluar los resultados de los modelos de redes neuronales es entender la topología de la red, esto permite entender el grado de complejidad bajo el cual se está realizando la regresión y a su vez entender la complejidad computacional que demandará más o menos recursos. En la siguiente tabla se puede observar los resultados de cada red, la topología, el número de capas ocultas y nodos que existe entre cada una de ellas.

Tabla 4.9 Resultados del diseño de redes neuronales

Lineal	Logístico Sigmoide	Tangente inversa
Capas ocultas: 6	Capas ocultas: 3	Capas ocultas: 4
Nodos: 26	Nodos: 19	Nodos: 32



Fuente: elaborado por el autor

Como se puede observar, todos los modelos propuestos respetan el modelo conceptual de tener siete variables independientes (satisfacción en diferentes ámbitos de la vida) como entrada y una variable dependiente (satisfacción general con la vida). Además, se observa que todos los modelos son multicapa porque presentan más de una capa oculta. Sin embargo, cuando una red se compone de más de tres multicapas es considerada como una red neuronal profunda como el caso de las redes con función de activación lineal y tangente inversa.

En términos de necesidad de procesamiento computacional el modelo de tangente inversa es el más complejo porque se compone de un total de 270 conexiones teniendo 4 capas y 32 nodos, seguido del modelo logístico sigmoide con 148 conexiones y finalmente con el modelo lineal de 132 conexiones, siendo este último el que mejores resultados presentan en función a lo observado en el acápite anterior.

4.5.1.3 Importancia relativa

La importancia relativa del predictor se entiende como la mejora porcentual de cada una de las variables empleadas para la realización de un modelo de redes neuronales, por tal motivo su análisis permite identificar aquellas variables que influyen en mayor o menor magnitud hacia la variable dependiente.

Es necesario mencionar que su análisis no tiene ninguna relación con si la predicción es o no precisa en términos de rendimiento.

En la siguiente tabla se detalla la importancia relativa de cada uno de los modelos de redes neuronales son su porcentaje de importancia relativa correspondiente:

Tabla 4.10. Importancia relativa modelos redes neuronales

Importancia relativa					
Lineal		Logístico sigmoide		Tangente inversa	
Variable	Importancia	Variable	Importancia	Variable	Importancia
Salud	23,90%	Salud	31,70%	Trabajo	25,60%
Trabajo	18,20%	Trabajo	19,30%	Economía	23,70%
Economía	14,20%	Vivienda	12,50%	Vecinos	11,30%
Vecinos	12,90%	Vecinos	12,40%	Salud	11,20%
Ambiente	11,30%	Economía	9,40%	Ambiente	9,80%
Vivienda	10,50%	educación	9,30%	educación	9,30%
Educación	9,00%	Ambiente	5,40%	Vivienda	9,10%

Fuente: elaborado por el autor

En función al tipo de modelo de regresión por redes neuronales se tienen diferentes resultados respecto a la importancia relativa de cada una de las variables independientes. Para los modelos lineal y logístico sigmoide existe una semejanza respecto a las variables satisfacción con la salud y satisfacción con el trabajo ocupando la primera y segunda posición de importancia respectivamente.

De forma específica, analizando las importancias relativas del modelo lineal, se observa que la satisfacción con la salud ocupa un 23.90% de relevancia para el modelo general. A esta importancia le siguen dos variables altamente correlacionadas entre sí que son la satisfacción con el trabajo y economía con un 18.20% y 14.20% respectivamente. A partir de aquí las variables restantes como la satisfacción con la relación con los vecinos, el medio ambiente, el estado de la vivienda y educación tienen importancias relativas inferiores al 12.90%.

La importancia relativa es la cuantificación de la contribución individual de un conjunto de variables predictoras al modelo de regresión, de allí se puede indicar que existe un análisis de dominancia que puede ser interpretado de fácil manera.

Por ejemplo, si se conoce que la variable satisfacción con la salud, incide mayoritariamente a la percepción con la satisfacción general con la vida, se convierte inmediatamente como factor de interés al cuál se le debe dar importancia y que su eventual mejoría o empeoramiento influiría directamente a un mejor o peor nivel de satisfacción con la vida general. De igual forma si se conoce que la satisfacción con el trabajo y economía ocupa una importancia elevada, resulta necesario priorizarlo al igual que la salud para que en conjunto sean factores decisivos que aporten a la mejoría de la satisfacción con la vida de forma general.

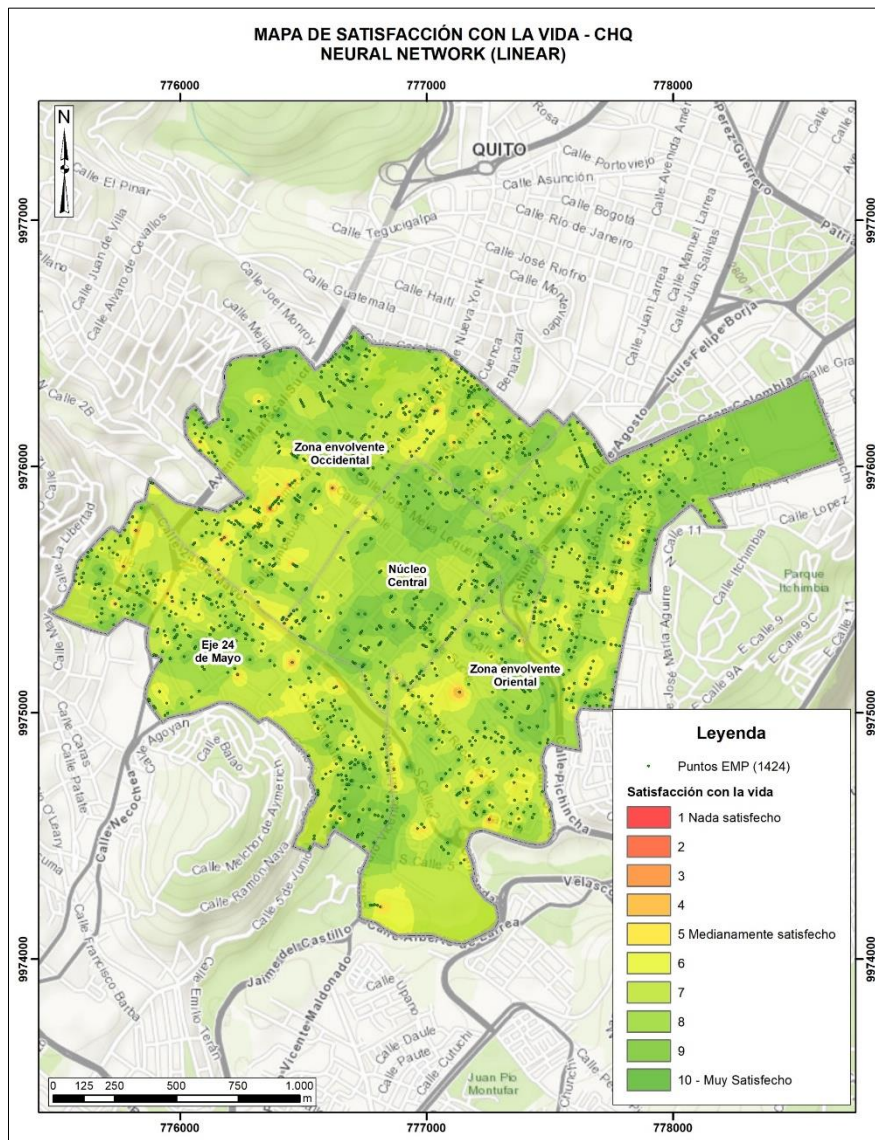
Resulta relevante entender que esta misma forma de análisis podría replicarse de forma sistemática hacia el resto de los modelos y a aquellas variables que cuya importancia relativa es inferior.

4.5.1.4 Modelamiento espacial Redes Neuronales (Lineal, Logístico Sigmoide, Tangente Inversa)

Para cada uno de los modelos de redes neuronales se realizó su correspondiente representación espacial, a través de la interpolación de los valores estimados con cada una de las funciones de activación que fueron: lineal, logístico sigmoide y tangente inversa. Los datos fueron representados a través de un modelo ráster enmascarado en el dominio del área de estudio e interpolado con la técnica de IDW cuya clasificación corresponde a la escala de valoración con la que fueron levantados los datos según la encuesta multipropósito en la que 1 significa nada satisfecho y se le asocia un color rojo, mientras que 5 significa medianamente satisfecho con color amarillo y 10 que implica muy satisfecho con colores verdes. La clasificación se mantiene para todos los modelos espaciales generados, con el objetivo que puedan ser comparables entre sí.

- **Modelo lineal:** en el mapa se puede observar que el modelo se interpola en gran parte de la zona del núcleo central y hacia la zona nororiental del CHQ en tonalidades verdes con valores superiores a 8 en los que respecta a satisfacción con la vida. En contraposición a esta distribución, se identifican zonas en tonalidades amarillas con tendencia a rojas en la calle que divide a la zona del eje 24 de mayo con el resto de las zonas estudiadas.

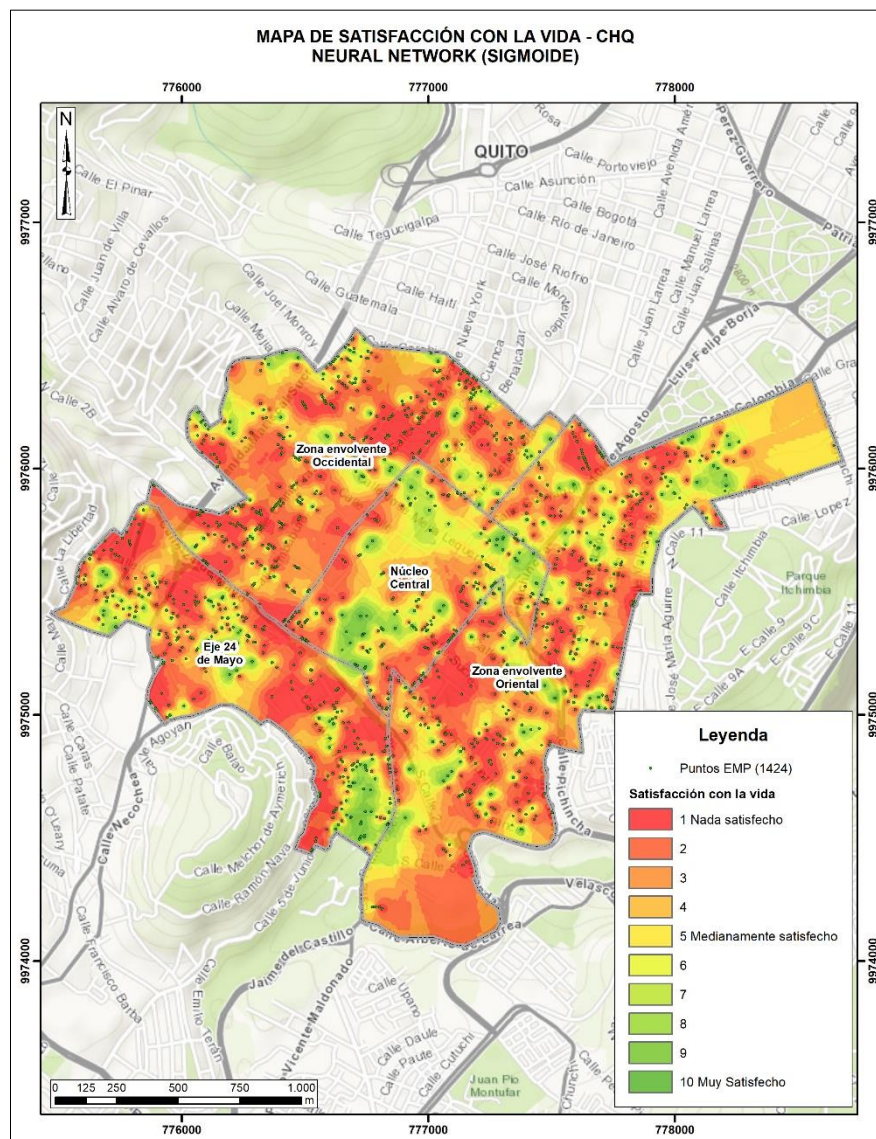
Mapa 4.11. Satisfacción con la vida modelo lineal redes neuronales



Fuente: elaborado por el autor

- Modelo logístico sigmoide: la representación del modelo connota una sobreestimación en aquellas cifras de satisfacción con la vida inferiores a 5, esto principalmente se debe al dominio que tiene la función de activación que corresponde a los reales positivos que van de 0 a 1. Dado que los datos predichos se aglomeran en cifras de 1 y 2, la interpolación muestra un modelo en el que prácticamente la mayoría del CHQ presenta una prácticamente nula satisfacción con la vida. Por esta razón y por lo indicado en el acápite de evaluación de las métricas estadísticas, se puede mencionar que el modelo logístico binario no representa de forma correcta al fenómeno principal estudiado.

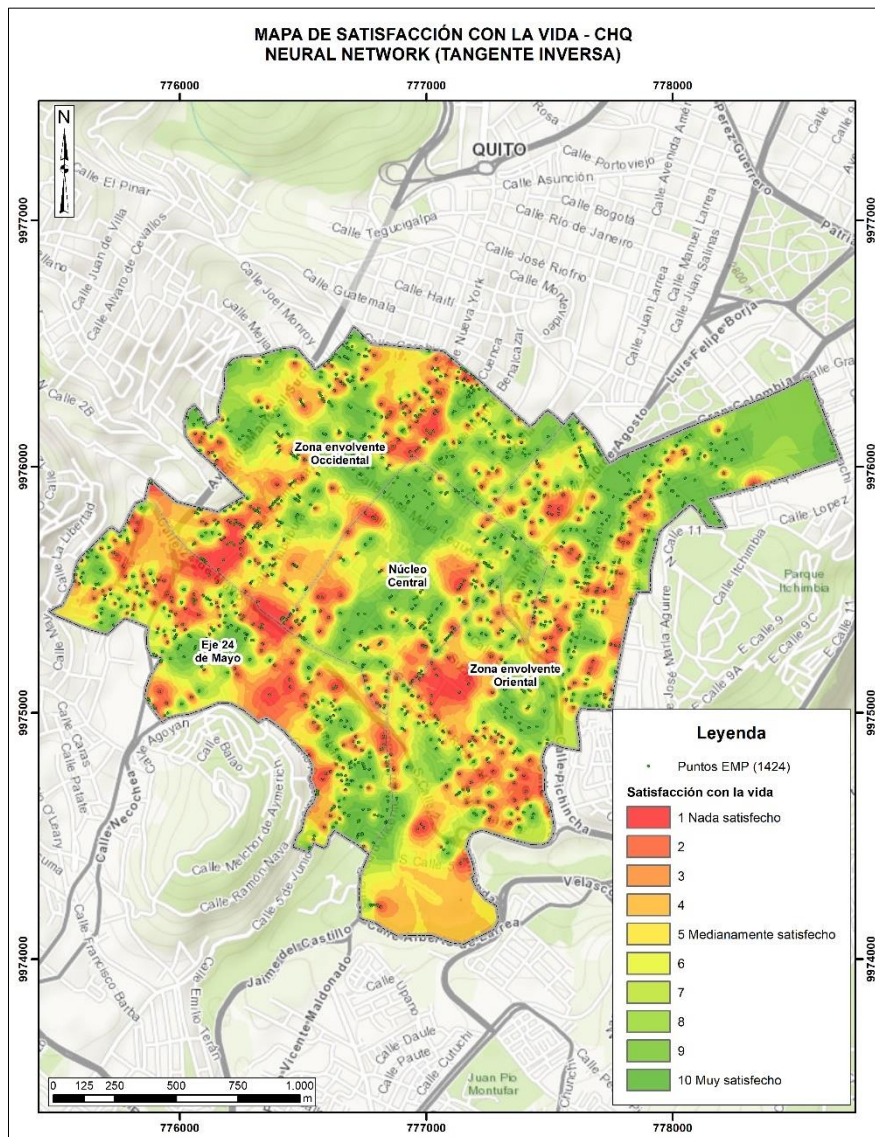
Mapa 4.12. Satisfacción con la vida modelo logístico sigmoide redes neuronales



Fuente: elaborado por el autor

- Modelo tangente inversa: en comparación a los anteriores dos modelos, la presentación con la función de activación tangente inversa presenta una distribución más contrastada entre valores altos y bajos de satisfacción con la vida. En perspectiva, si se compara a este modelo con el modelo lineal, se vuelven más evidentes aquellas zonas con una tendencia mediana y nada satisfecho con la satisfacción con la vida, que se muestran en la misma zona del sur de la zona envolvente occidental y el límite entre el eje 24 de mayo, también se presentan focos de bajos valores en la zona envolvente oriental e inclusive en el norte de la zona envolvente occidental

Mapa 4.13. Satisfacción con la vida modelo tangente inversa redes neuronales



Fuente: elaborado por el autor

4.5.2 Resultados del Árbol de decisión

4.5.2.1 Evaluación de parámetros estadísticos

En lo que se refiere a los resultados de los parámetros estadísticos del árbol de decisión se presentan y analizan en la siguiente tabla:

Tabla 4.11. Métricas de evaluación rendimiento estadístico del modelo árbol de decisión

Métricas de evaluación	Árbol de decisión
MSE	0.675
RMSE	0.822

Métricas de evaluación	Árbol de decisión
MAE / MAD	0.599
MAPE	80.66%
R²	0.367

Fuente: elaborado por el autor

Los resultados del RMSE y R² del árbol de decisión revelan que este modelo tiene una semejanza y prácticamente equivalencia estadística con el modelo de redes neuronales lineal porque sus métricas de evaluación estadísticas son iguales para sus valores de desviación estándar (precisión) y coeficiente de variación, a pesar de que existe un minúsculo incremento de 0.001 en su valor de R².

Tanto las redes neuronales como los árboles de decisión son técnicas metodológicas diferentes, pero se puede indicar que es un buen indicador que sus métricas estadísticas de evaluación sean equivalentes, porque facilitan al análisis comparativo, ya que en términos de rendimiento ambos modelos se respaldarían entre sí, lo que a su vez respaldaría a sus correspondientes representaciones y modelos espaciales otorgándole robustez al análisis.

4.5.2.2 Importancia relativa

Al igual que los modelos de redes neuronales, el modelo de árbol de decisión presenta los resultados de la importancia relativa para cada una de las variables independientes que fueron utilizadas para la predicción de modelos.

Tabla 4.12. Importancia relativa modelo árbol de decisión

Importancia relativa	
Árbol de decisión	
Variable	Importancia
Trabajo	26,06%
Salud	18,55%
Economía	17,21%
Vivienda	15,35%
Vecinos	13,18%
Educación	9,42%
Ambiente	0,23%

Fuente: elaborado por el autor

Con base en los resultados de la importancia relativa, se observa que para el modelo la variable satisfacción con el trabajo aportan con un 26.06% de importancia, seguido de la variable satisfacción con la salud con un 18.55% y economía con un 17.21%. Es necesario mencionar que estas tres variables también ocupan los primeros lugares del modelo de redes neuronales

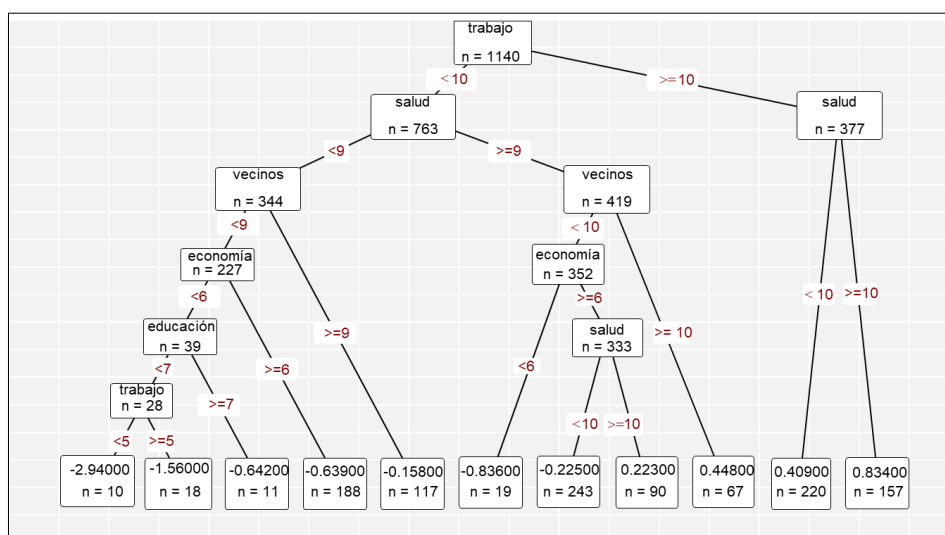
con función de activación lineal, y aglomeran más de 50% de la importancia relativa de los modelos. Sin embargo, diferente en sus porcentajes de participación orden. Por otra parte, las variables de satisfacción con la vivienda, vecinos, educación y ambiente aportan con una participación inferior al 15.35% en cada una de sus variables.

De este análisis también llama la atención el bajo porcentaje de participación que tiene la variable satisfacción con el medio ambiente, ya que prácticamente no aporta a la construcción del modelo y esto se evidencia en la construcción del árbol de decisión.

4.5.2.3 Diseño del árbol de decisión

La gráfica de árbol de decisión resultante presenta de forma explícita y en concordancia con la matriz de importancia relativa, aquellas variables consideradas de mayor importancia además de los cortes que existen para ir construyendo las ramas de decisión. A continuación, se presentan el diseño de árbol de decisión resultante:

Figura 4.10. Árbol de decisión



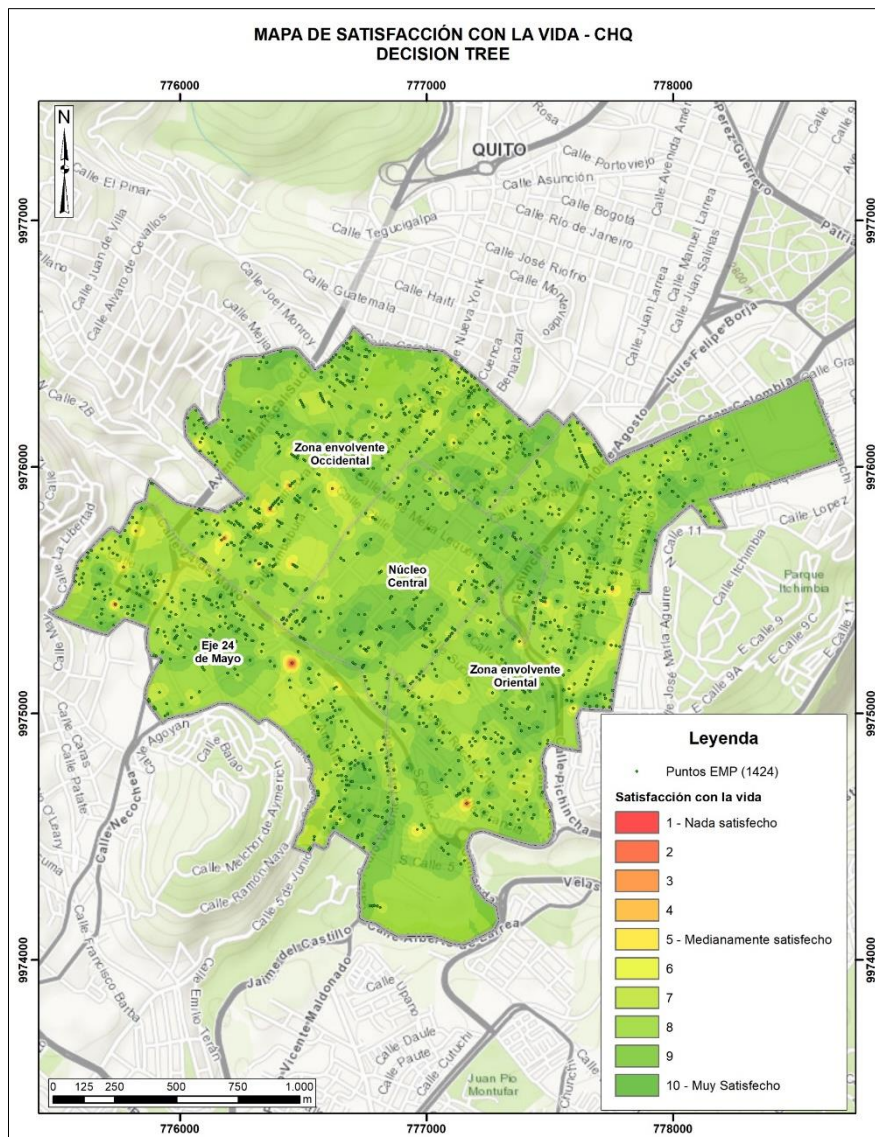
Fuente: elaborado por el autor

En este sentido y a manera de ejemplo, se observa que, si una persona hubiese respondido que se siente muy satisfecho con su trabajo (10) y muy satisfecho con su salud (10), automáticamente se podría afirmar que esta persona se encuentra satisfecha con su vida, sin la necesidad de realizar ninguna clase de pregunta adicional. Sin embargo, y lo verdaderamente interesante se encuentra en aquellas personas que en un primer momento calificarían inferior a 10 en satisfacción con el trabajo e inferior a 9 en la satisfacción con la salud, ya que en esos casos el árbol de decisión toma otras ramificaciones y combinaciones como una baja satisfacción con la economía (<math>< 6</math>) y educación (<math>< 7</math>) que incidirían sistemáticamente hasta obtener una baja calificación con la satisfacción con la vida.

4.5.2.4 Modelamiento espacial Árbol de Decisión

El modelo espacial resultante de esta técnica de *machine learning* presenta una distribución relativamente homogénea y uniforme de satisfacción con la vida cuyos valores predichos rondan entre 8 – 10. Sin embargo, se logran distinguir zonas con valores muy bajos de 1 y 2 que se ubican precisamente en la zona aledaña al eje vial de la 24 de Mayo, específicamente entre las calles Gabriel García Moreno y Ambato. Adicionalmente se observa que este modelo tiene concordancia en las zonas que presentan valores bajos de satisfacción con la vida en los análisis de *hot spots* y *cold spots* de la técnica G* de Getis Ord ya que de ambos análisis se muestran las zonas que presentan una autocorrelación espacial de valores bajos de satisfacción con la vida.

Mapa 4.14. Satisfacción con la vida modelo tangente inversa redes neuronales



Fuente: elaborado por el autor

4.5.3 Resultados Support Vector Machine

4.5.3.1 Evaluación de parámetros estadísticos

Los resultados estadísticos del modelo de Máquina de Vectores Soporte presentan un valor de RMSE de 0.766 siendo este dato el menor de todos los modelos ejecutados en la presente tesis, lo cual implica que este modelo tiene un intervalo de confianza más acertado al momento de modelar la variable dependiente. Sin embargo, este modelo no presenta el valor más alto de R^2 ya que con 0.352 este dato se encuentra por debajo de los resultados de Redes Neuronales Lineales y Árboles de Decisión.

Tabla 4.13. Métricas de evaluación estadística del modelo support vector machine

Métricas de evaluación	Support Vector Machine
MSE	0.586
RMSE	0.766
MAE / MAD	0.565
MAPE	99.62%
R^2	0.352
Vectores de soporte	1111

Fuente: elaborado por el autor

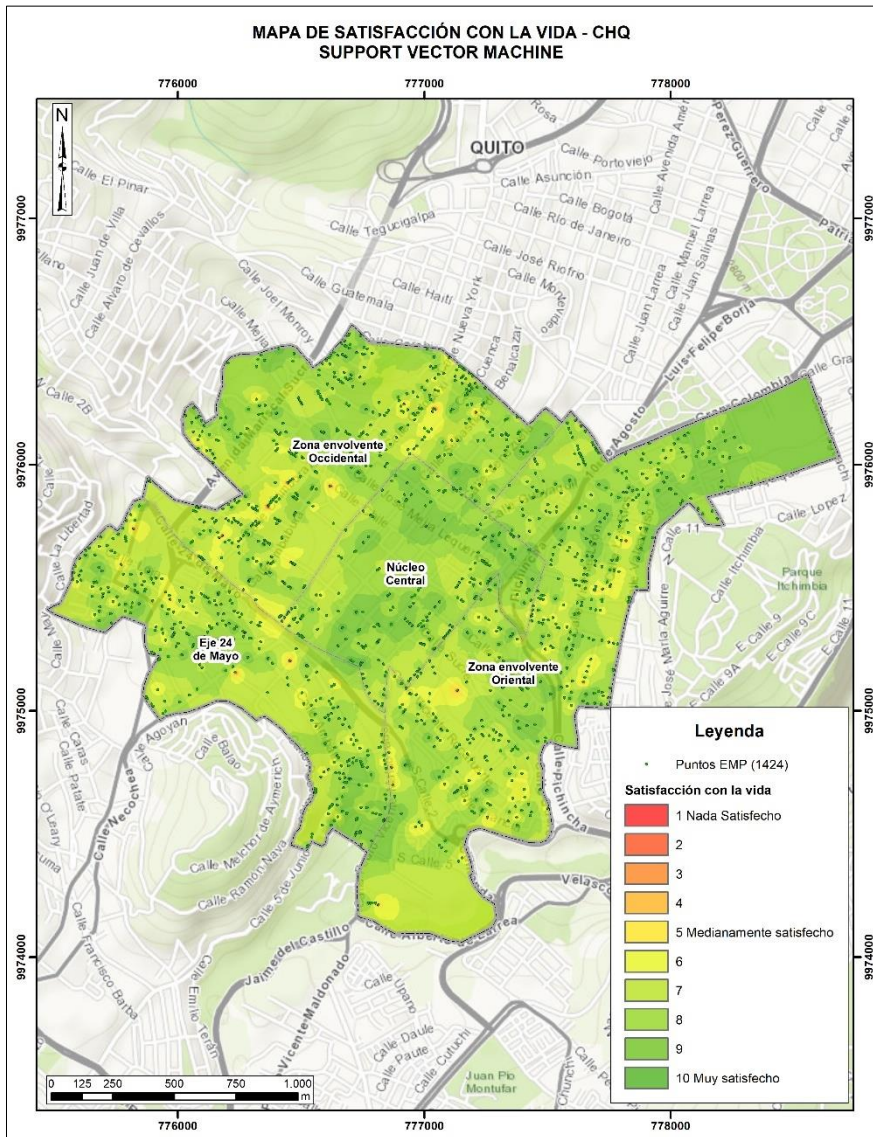
Por las características propias este modelo, no presenta un resultado de los porcentajes de importancia relativa, ya que utiliza arreglos matriciales, 1424 vectores, de pesos que van asignados a cada una de las observaciones, esto a su vez difiere de las otras metodologías de interpolación ya que no se asigna un único peso a las variables independientes lo cual no permite calcular una matriz de importancia relativa.

La razón principal de esto es que el proceso interno de las SVM no se basa en la evaluación de la importancia individual de las variables de la misma manera en que lo hacen, por ejemplo, los árboles de decisión. Las SVM buscan encontrar el hiperplano que mejor se ajusta a los datos en el caso de regresión, y se enfocan en el espacio de características y las relaciones entre las observaciones de datos más que en la importancia relativa de cada variable individual.

4.5.3.2 Modelamiento espacial *Support Vector Machine*

El modelo espacial resultante de método de *Support Vector Machine* presenta una distribución uniforme de satisfacción con la vida cuyos valores predichos rondan entre 6-10. Al igual que los modelos de redes neuronales (función lineal) y árbol de decisión, coinciden aquellas áreas con valores más altos y otros con valores intermedios, y también homogeniza una zona de media satisfacción con la vida en lo que la EMP considera como la Zona envolvente occidental.

Mapa 4.15. Satisfacción con la vida modelo support vector machine



Fuente: elaborado por el autor

Capítulo 5. Discusión y conclusiones

5.1 Discusión

5.1.1 Escalabilidad

Al trabajar con datos de la Encuesta Multipropósito realizada en el año 2016 y que estaba limitada a una zona específica, como el centro histórico de Quito, presenta ciertas restricciones que deben ser consideradas al interpretar los resultados y extrapolar conclusiones. En primer lugar, el factor temporal puede afectar la relevancia de los datos en el contexto actual. Los patrones sociales, económicos y demográficos pueden haber experimentado cambios significativos en los años transcurridos desde el desarrollo de encuesta, lo que podría llevar a una representación desactualizada de la realidad presente en la ciudad. Además, la extrapolación de resultados de una zona específica a la totalidad de Quito podría no tener en cuenta las variaciones geoespaciales y socioeconómicas que existen en diferentes partes de la ciudad. La diversidad de características entre el centro histórico y otras zonas urbanas puede llevar a generalizaciones inexactas y a la omisión de particularidades importantes en los análisis a mayor escala. Por lo tanto, aunque los datos de la encuesta multipropósito proporcionan una visión valiosa de la realidad en ese momento y en esa zona específica, no permiten extrapolar conclusiones para comprender la complejidad total y actual de la satisfacción con la vida en la ciudad de Quito.

Extrapolar un modelo de análisis espacial de una zona relativamente pequeña a una cobertura más amplia, como la ciudad de Quito, es fundamental para comprender de manera holística los patrones y tendencias que subyacen en los datos geoespaciales. Si bien los análisis detallados en una zona específica como el caso del CHQ pueden proporcionar información valiosa sobre las interacciones locales, al ampliar el enfoque a una escala más grande permitiría capturar relaciones más complejas y emergentes que no serían evidentes en una perspectiva limitada.

Al extrapolar las técnicas metodológicas de este análisis, se podrían identificar nuevos patrones o variables de interés como la movilidad, distribución de recursos, densidad poblacional y otros factores clave que pueden tener implicaciones significativas en la percepción de la satisfacción con la vida y esto a su vez en la planificación urbana, la toma de decisiones gubernamentales y la asignación de recursos a nivel municipal.

Además de su contribución a la planificación urbana, la extrapolación de un modelo de análisis espacial relacionado con la percepción subjetiva a una escala mayor como la ciudad de Quito resultaría esencial para anticipar y abordar desafíos relacionados con el crecimiento y la expansión urbana ya que los sistemas urbanos son intrínsecamente interconectados, y las

acciones tomadas en una parte de la ciudad pueden tener efectos en otras áreas aparentemente distantes.

La extrapolación y escalabilidad de esta clase de investigaciones y de modelos de análisis espacial aplicados a un dominio espacial más extenso, podría convertirse en un instrumento esencial para una gestión urbana informada y prospectiva en beneficio de la calidad de vida de los habitantes de la ciudad de Quito.

5.1.2 Análisis exploratorios

La aplicación de métodos de autocorrelación espacial, como LISA y G de Getis-Ord, en el análisis de variables subjetivas como la percepción de la satisfacción con la vida, ofrece una perspectiva única y enriquecedora para comprender las dinámicas espaciales de estos fenómenos. Aunque las mediciones subjetivas pueden ser difíciles de cuantificar y analizar debido a su naturaleza de los datos, la distribución geográfica de estas percepciones puede revelar patrones de agrupamiento que son esenciales para la formulación de políticas y programas de bienestar en aquellos sitios que presentan una baja percepción de satisfacción con la vida.

Los métodos de autocorrelación espacial permiten identificar áreas en las que la satisfacción con la vida tiende a ser similar, lo que puede sugerir la presencia de factores contextuales compartidos que influyen en las percepciones de las personas. Al revelar las relaciones espaciales entre las respuestas subjetivas, estos métodos brindan a los planificadores urbanos y a los responsables de la toma de decisiones la oportunidad de diseñar estrategias dirigidas a mejorar la calidad de vida en áreas específicas, abordando las necesidades y preocupaciones locales de manera más efectiva.

A menudo, las percepciones personales están influenciadas por factores socioeconómicos, culturales y ambientales únicos de un área determinada. Al utilizar LISA y G de Getis-Ord, fue posible identificar *clústers* espaciales de alta o baja satisfacción con la vida, lo que puede indicar la existencia de áreas donde la cohesión social, el acceso a servicios o la calidad del entorno influyen en la percepción del bienestar subjetivo. Además de que estos métodos ayudan a detectar patrones que podrían haber pasado desapercibidos en un análisis de agregación global, permitiendo una toma de decisiones más precisa y dirigida en políticas que buscan mejorar la calidad de vida en áreas específicas del CHQ.

5.1.3 Aprendizaje no supervisado

El aprendizaje no supervisado desempeña un papel fundamental en el análisis de datos de carácter subjetivo, como lo es la satisfacción con calidad de vida, debido a su capacidad para abordar la inherente complejidad y variabilidad de tales fenómenos.

Estas mediciones pueden abarcar una amplia gama de factores subjetivos, desde la satisfacción personal con el estado de salud, económico, trabajo, medio ambiente y el bienestar general. En este sentido, su discusión toma relevancia ya que estas variables a menudo no se ajustan a estructuras lineales o predefinidas, lo que hace que los enfoques de aprendizaje supervisado sean limitados en su aplicación, debido a dicha complejidad. El aprendizaje no supervisado, en contraste, permite descubrir patrones ocultos y estructuras subyacentes en los datos sin la necesidad de que estos estén etiquetados previamente. Este en perspectiva puede resultar ventajoso ya que permite la detección de relaciones complejas y no lineales que pueden ser cruciales para entender la calidad de vida en toda su amplitud. Además, al no imponer restricciones preconcebidas, el aprendizaje no supervisado puede identificar factores inesperados que podrían influir significativamente en la calidad de vida, ofreciendo así una visión más completa y holística de estas variables y conceptos subjetivos.

El uso de métodos de aprendizaje no supervisado como *k-means*, *skater* y *cluster* jerárquico puede ser valioso en un primer acercamiento para analizar la representación espacial de la satisfacción con la vida y las variables dependientes. Estos métodos permiten agrupar los datos de manera automática, lo que puede ayudar a identificar patrones subyacentes en los conjuntos de datos. Sin embargo, de los tres métodos mencionados, *k-means* se destaca como el más eficiente en la representación de los datos.

La ventaja principal de *k-means* radica en su capacidad para crear grupos bien definidos y compactos, minimizando la varianza entre grupos clasificados. Esto significa que *k-means* puede identificar de manera más precisa grupos cohesivos de datos que representan diferentes niveles de satisfacción con la vida y las variables dependientes. A diferencia del *cluster* jerárquico y *skater*, que pueden generar agrupaciones más dispersas y no tan claras, *k-means* ofrece una representación más contrastada, nítida y utilizable de la estructura subyacente en los datos, lo que facilita su interpretación y análisis posterior. A pesar de ello, es importante mencionar que los tres modelos tienen sus aplicaciones y, en algunos casos, *cluster* jerárquico y *skater* pueden ser preferibles cuando se buscan estructuras más complejas o interpretables en los datos.

Si se dispusiera de un conjunto de datos más amplio que abarcara todo el dominio espacial de la ciudad de Quito, es probable que el método de *cluster* jerárquico se vuelva más eficiente y preferible en comparación con *k-means*. Esto se debe a dos razones fundamentales. En primer lugar, el *cluster* jerárquico tiene la ventaja de capturar estructuras de agrupación a múltiples escalas, lo que sería beneficioso en una ciudad diversa como Quito, donde diferentes niveles de satisfacción con la vida y variables dependientes pueden variar significativamente según la ubicación geográfica, el estrato socioeconómico y otros factores como la percepción de salud y

medio ambiente. En segundo lugar, el *cluster* Jerárquico permite una mayor flexibilidad en la interpretación de los resultados, ya que puede generar dendrogramas que revelan la jerarquía de agrupaciones, lo que sería valioso para comprender las relaciones entre grupos en una ciudad grande y heterogénea como Quito.

5.1.4 Aprendizaje supervisado

Una vez analizados los resultados de la presente Tesis, se puede indicar que el modelo de árbol de decisión emerge como una herramienta excepcional para analizar la satisfacción con la vida, siendo esta metodología la mejor de todas las empleadas para analizar en conjunto múltiples variables independientes, como la satisfacción con la salud, la economía, el trabajo, el medio ambiente, la educación y la vivienda.

Se destaca su destreza para representar visualmente las complejas interacciones entre estas variables proporciona una comprensión clara y accesible de cómo influyen en la satisfacción con la vida. Además de su capacidad para la visualización, los árboles de decisión también sobresalen en el análisis estadístico inferencial. Al evaluar la importancia relativa de cada variable independiente, estos modelos permiten identificar cuáles factores tienen el mayor impacto en la satisfacción con la vida como lo son la satisfacción con el trabajo (26,06%), satisfacción con la salud (18,55%), satisfacción con la economía (17,21%), satisfacción con la vivienda (15,35%), satisfacción con la educación y ambiente (9,42% y 0,23% respectivamente).

Estos resultados son considerados fundamentales, ya que permiten la priorización de intervenciones y el desarrollo de políticas públicas basadas en evidencia científica y que se enfocan en mejorar el bienestar de la población desde diferentes aristas siempre teniendo en cuenta que, una mejoría en la variable más influyente podría aportar más hacia una mejoría integral de la satisfacción con la vida.

En perspectiva, los modelos de redes neuronales con función de activación lineal, logístico sigmoide y tangente inversa ofrecen una flexibilidad excepcional en el aprendizaje de patrones complejos y son capaces de capturar relaciones de carácter lineal como no lineal en los datos, lo que los hace adecuados para tareas de alta dimensionalidad y complejidad.

Además, su capacidad de generalización puede resultar en un buen rendimiento en datos no vistos. Sin embargo, su desventaja frente al modelo de árbol de decisión radica en su falta de interpretabilidad, ya que las decisiones tomadas por una red neuronal pueden ser difíciles de entender. Adicionalmente, este tipo de algoritmos tienden a requerir más datos para entrenar al modelo con éxito, como el caso de la satisfacción con la vida cuya base de datos fueron 1424 observaciones de las cuales el 80% se empleó para el entrenamiento y el 20% restante se empleó como validación y prueba.

Por otro lado, el modelo de *support vector machine* se destaca por su capacidad para lidiar con problemas de alta dimensionalidad y su capacidad de separar eficazmente clases en un espacio de características. Sin embargo, la principal desventaja de SVM es su dificultad para lidiar con conjuntos de datos relativamente grandes, ya que su entrenamiento puede ser computacionalmente costoso. Además, la interpretación de los resultados puede ser desafiante, especialmente en dimensiones superiores. En términos de representación espacial, SVM puede ser efectivo para separar clases geoespaciales, pero la interpretación de cómo se toman estas decisiones en el espacio puede requerir técnicas adicionales.

Finalmente, al ser datos georreferenciados tanto los árboles de decisión como los otros modelos permiten una representación cartográfica y desempeña un papel importante en el entendimiento del fenómeno dentro de un dominio territorial como el caso del Centro Histórico de Quito. Es así como realizando análisis geoespaciales adicionales y de interpolación como IDW, es posible identificar áreas geográficas específicas con niveles significativamente altos o bajos de satisfacción con la vida. Esto tiene implicaciones directas en la planificación urbana y la asignación de recursos a nivel local, permitiendo abordar desigualdades geográficas en el bienestar de la población de manera más efectiva.

5.1.5 Comparación de Resultados de la investigación frente a otros estudios cuantitativos sobre satisfacción con la calidad de vida.

En una comparativa, los hallazgos de esta tesis muestran una correlación significativa con investigaciones previas, como la de Tonón (2008), en la que se establece que los estudios sobre la calidad de vida y el bienestar están vinculados a la percepción individual de cada persona con respecto a su posición en el contexto cultural y sistema en el que se desenvuelve. Estos factores se ven influidos por la situación laboral, la salud física y psicológica, así como las relaciones sociales y la interacción con el entorno ambiental.

Los resultados presentados a lo largo de este documento contribuyen a la creciente evidencia que respalda la necesidad de explorar nuevos modelos explicativos de la satisfacción con la vida, tal como lo sugieren investigaciones anteriores, incluyendo los trabajos de Moreno & Vera (2011), Crous et al. (2018) y Sarriera et al. (2018). A pesar de que estos estudios tienen en común el grupo demográfico de interés, que son niños y adolescentes, todos coinciden en la identificación de las variables clave que pueden influir en la satisfacción con la vida. En conjunto, concluyen que aquellos estudios que incorporan tanto variables de carácter intraindividual (subjetivas) como factores ambientales y sociales (objetivas) son los que mejor se ajustan y explican esta variable dependiente.

Dentro de este marco en el que se plantea la búsqueda de nuevas herramientas o modelos de explicación es que se incorpora a las técnicas de *machine learning* y analítica socioespacial

como los desarrollados y presentados en esta tesis y en otros trabajos publicados como los de Shen et al. (2023) en el que se emplean varios modelos de aprendizaje automático como vectores de máquina soporte (*SVM*) o árboles de decisión (*DT*) se puede indicar que modelos inferenciales de satisfacción con la vida son empleados de forma más generalizada en otras partes del mundo. Ya que tanto en esa investigación como en el desarrollo de esta discusión se coincide que los estudios de satisfacción con la vida se han realizado ampliamente a través de investigaciones empíricas, cuestionarios y análisis teóricos y la gran mayoría de estos análisis basan sus análisis en relaciones simples entre variables. Sin embargo, la mayoría de las relaciones de la vida real, son complejas y no pueden aproximarse con correlaciones simples como el caso del análisis del bienestar subjetivo.

Al igual que en desarrollo de esta tesis, en el estudio de Shen et al. (2023) se emplean diferentes técnicas metodológicas que en la evaluación y comparación de modelos se someten a un análisis respecto a su valor de R^2 para evaluar el nivel de ajuste de cada uno de ellos, en ese caso se emplean diferentes variables predictoras y se observa que el mejor de todos los empleados corresponde al el modelo vecinos más cercanos (*KNN* por sus siglas en inglés), alcanzando un valor de R^2 de 0.506 en comparación a los valores de 0.441 y 0.447 de máquinas de vectores soporte y árboles de decisiones respectivamente. En este sentido, sería recomendable ahondar en esta técnica para que se incluya como modelo de interés para futuras investigaciones.

En lo que respecta al análisis socioespacial, los estudios de bienestar subjetivo se han abordado desde un análisis con un nivel de agregación que responde a unidades espaciales como manzanas y barrios (Lucero et al. 2007). Si bien son correctas las formas de representación espacial esta tesis ha demostrado que esta clase de variables pueden ser proyectadas y estudiadas con la unidad de análisis más básica que son puntos y estos a su vez posteriormente permiten generar superficies continuas (*rasters*) que permiten una mejor visualización y entendimiento de la información. Queda claro que de requerir un análisis con un dominio de geográfico más extenso como el caso de la ciudad de Quito lo correcto sería agregar la información a nivel de sectores o manzanas censales con el objetivo de tener más representatividad espacial.

En un estudio paralelo llevado a cabo por Rodríguez (2023) se obtuvieron resultados comparables, a pesar de que ambos trabajos aplican metodologías distintas. Al evaluar los resultados del modelo de satisfacción con la vida de Rodríguez y Revelo, se llega a la conclusión de que existen variables independientes que guardan una relación causal con la satisfacción general de vida. Entre las variables explicativas identificadas de manera consistente

se encuentran la satisfacción con la salud actual (coincide), la vivienda (difieren) y el trabajo (coincide).

Por otro lado, los modelos del estudio mencionado logran un R2 de hasta el 50%, destacándose por tener el coeficiente de determinación más elevado en comparación con el valor máximo de R2 de este trabajo, que es del 36%. Es crucial tener presente que se emplearon variables independientes diferentes en los modelos conceptuales propuestos, lo cual puede influir en las divergencias observadas en los resultados.

Ambos estudios coinciden en que estas metodologías representan un aporte significativo al ámbito de estudios relacionados a la evaluación de la calidad de vida subjetiva y también al uso de técnicas cuantitativas, espaciales y de machine learning en las ciudades. Y que se generaron aportes metodológicos que marcan la pauta para estudios posteriores.

5.1.6 Implicación de resultados para la planificación urbana y las políticas públicas.

Para Leva (2005b), el análisis de satisfacción con la calidad se entiende como un objetivo primordial de las políticas públicas, siendo el resultado de la satisfacción de las necesidades vinculadas al bienestar de los ciudadanos. En un entorno donde las demandas ciudadanas son cada vez más evidentes, la elevación de los estándares de vida se convierte en un tema crucial en la agenda de las administraciones locales y nacionales.

El desafío se concentra en la formulación de políticas que fomenten esta mejora, asegurando el acceso de la población a los recursos que satisfacen estas necesidades. En este contexto, la medición de la satisfacción de calidad de vida se vuelve esencial. Los indicadores y análisis diseñados para este fin no solo deben reflejar aspectos objetivos, como la disponibilidad de recursos, sino también considerar la dimensión subjetiva, que se relaciona con la percepción individual y las características personales que influyen en los niveles de vida de forma multidimensional (Tonon 2010).

El reconocimiento de una relación no perfecta entre las condiciones objetivas de vida y la percepción del individuo sobre su calidad de vida enfatiza la complejidad de esta evaluación. Aquí es donde la propuesta de emplear una serie de metodologías como *machine learning* y analítica socioespacial cobran protagonismo, ya que permiten la construcción de indicadores urbanos, índices de calidad de vida urbana y análisis espaciales complejos. Razón por la cual, los resultados de esta investigación se convierten en un recurso valioso para los responsables de tomar decisiones ofreciéndoles información más precisa y alineada con las problemáticas reales de los ciudadanos del Centro Histórico de Quito.

En lo que respecta al ámbito de la planificación urbana es necesario considerar lo mencionado por Mouratidis (2021), que básicamente indica que esta clase de estudios que ofrecen información para mejorar la calidad de vida en entornos urbanos, haciendo de esto un objeto de interés esencial en la planificación de las ciudades. El constante crecimiento demográfico y los procesos de urbanización han generado un aumento significativo en las poblaciones urbanas a nivel global, convirtiendo la calidad de vida en un asunto crucial para una cantidad cada vez mayor de residentes.

Paralelamente, las transformaciones en las características físicas, demográficas, económicas y sociales de las ciudades responden a las necesidades de sus habitantes residentes, razón por la cual, el comprender más a fondo la relación entre el entorno construido y la satisfacción de calidad de vida en contextos urbanos puede desempeñar un papel fundamental en moldear tanto el desarrollo presente como futuro de las ciudades.

5.1.7 Discusión sobre limitaciones del estudio, y sobre ideas futuras de investigación sobre el tema tratado en la Tesis

En el desarrollo de esta tesis se identificaron múltiples definiciones de calidad de vida. Varios estudios sostienen que la calidad de vida es un concepto multidimensional y, como es un concepto complejo, una estructura multidimensional es ventajosa al medirla (Dissart & Deller, 2000; Diener et al., 1985b; Rogge & Van Nijverseel, 2019). La calidad de vida urbana incluye condiciones materiales y no materiales, condiciones de vida individuales y colectivas, la dimensión objetiva de estas condiciones de vida (indicadores basados en métricas universales) y la dimensión subjetiva de estas condiciones de vida (satisfacción de las personas).

Así, es necesario indicar que el enfoque en la dimensión subjetiva de la calidad de vida, aunque es fundamental para comprender la percepción individual de las condiciones de vida, puede, de hecho, limitar la comprensión más amplia y holística de este concepto. La preponderancia de la satisfacción personal puede resultar insuficiente para capturar la complejidad multidimensional de la calidad de vida. Al centrarse en la satisfacción subjetiva, se podrían pasar por alto factores objetivos igualmente significativos que influyen en la calidad de vida urbana.

Este énfasis en la subjetividad puede restringir la comparabilidad y la aplicabilidad general de los resultados, ya que no todas las métricas subjetivas son fácilmente cuantificables o comparables entre diferentes grupos demográficos o contextos urbanos. Al descuidar indicadores objetivos, como las condiciones materiales, el acceso a servicios básicos o la infraestructura, se limita la comprensión completa de la calidad de vida.

Por ende, una evaluación más equilibrada y completa de la calidad de vida urbana debería considerar no solo la satisfacción subjetiva, sino también aspectos objetivos que son

cuantificables y comparables. Integrar indicadores tangibles y métricas objetivas en la evaluación de la calidad de vida puede ofrecer una visión más rica y representativa, lo que resulta crucial para el diseño e implementación de políticas públicas efectivas orientadas a mejorar la calidad de vida en entornos urbanos.

Adicionalmente a esto, otra de las limitaciones tiene que ver con la información que se empleó para la tesis, ya que al trabajar con datos de una encuesta correspondiente al Centro Histórico de Quito en el 2016 se puede presentar limitaciones al generalizar conclusiones. La restricción temporal puede ofrecer una visión desactualizada de la ciudad, ya que los cambios socioeconómicos y demográficos podrían no reflejarse. Además, extrapolar datos de una zona a toda la ciudad puede ignorar las variaciones geoespaciales y socioeconómicas entre áreas. Esta extrapolación podría conducir a generalizaciones inexactas y a omitir particularidades cruciales en un análisis amplio. Así, aunque valiosos, estos datos no capturan la complejidad total de la satisfacción con la vida en Quito.

Con miras hacia análisis futuros el traslado de un modelo de análisis de una zona pequeña a una escala más amplia es fundamental para capturar patrones complejos y emergentes en datos geoespaciales. Ampliar el enfoque permite revelar nuevas variables de interés, como movilidad, distribución de recursos y densidad poblacional, elementos clave para la planificación urbana. Esta extrapolación contribuiría a una gestión urbana más informada y sería crucial para anticipar y abordar desafíos de crecimiento urbano, considerando la interconexión de sistemas urbanos, permitiendo acciones más precisas y beneficiosas para toda la ciudad de Quito.

Del mismo modo, los modelos conceptuales aquí planteados y la revisión de la literatura pueden ofrecer orientación teórica y metodológica para futuras investigaciones empíricas. Y que de este modo se contribuya desde nuevos enfoques a la práctica de la planificación urbana. Además, se proporcionado sugerencias basadas en evidencia sobre estrategias de planificación urbana que podrían guiar a los profesionales, formuladores de políticas y tomadores de decisiones hacia un mejor y coherente desarrollo local tanto en el CHQ como en toda la ciudad.

5.2 Conclusiones

La satisfacción con el trabajo, la satisfacción con la salud y economía son factores influyentes al momento de determinar el bienestar subjetivo de una persona, si bien resulta algo lógico y esperable, no resulta sencillo o evidente el poder determinar de forma porcentual y con un nivel de confiabilidad estadística aquellos factores que verdaderamente influyen en mayor o menor proporción. Esto es precisamente, lo que se ha conseguido a través de esta tesis, presentando datos e información sustentados en evidencia cuantitativa mediante el uso de algoritmos de *machine learnig*.

Este tipo de investigaciones tiene un alto potencial de replicabilidad, ya que esta metodología de análisis puede extrapolarse a más áreas de interés a lo largo del área urbana y rural del DMQ, lo que a su vez dotaría a los tomadores de decisiones de información y herramientas que faciliten la propuesta y ejecución de planes de acción. Sobre esta misma línea, otro punto de interés es sobre lo que se podría hacer desde la Municipalidad para mejorar la satisfacción con la vida en aquellos sectores del Centro Histórico que presentan un bajo bienestar subjetivo.

En conclusión, tanto de los análisis de *machine learning* supervisados como de los no supervisados, el modelo de árbol de decisión se presenta como el mejor de todos, ya que por sus resultados estadísticos y por su enfoque altamente efectivo y versátil para analizar y representar a la satisfacción con la vida en el CHQ en función de múltiples variables independientes. En el ámbito del análisis socioespacial, los árboles de decisión en conjunto con técnicas de analítica espacial permiten identificar patrones geográficos de la satisfacción con la vida. Esto es de gran relevancia para la planificación urbana y la asignación de recursos a nivel local, ya que posibilita la detección de desigualdades geográficas en el bienestar de la población y orienta la toma de decisiones para abordar estas desigualdades de manera efectiva.

En conjunto, el modelo de árbol de decisión emerge como una herramienta completa y poderosa para comprender y mejorar la satisfacción con la vida desde diversas perspectivas, incluyendo su representación visual, análisis estadístico y análisis geográfico. Su versatilidad y capacidad para aportar claridad a la toma de decisiones lo convierten en un recurso invaluable en la analítica de *machine learning* y en la búsqueda de un mayor bienestar para la sociedad.

Históricamente, diversos estudios relacionados con la satisfacción con la vida, la felicidad y el bienestar se han explorado a través de miradas cualitativas y filosóficas, centrándose en narrativas individuales y en profundizar en la experiencia subjetiva tanto en su medición como interpretación. Sin embargo, los análisis cuantitativos brindan la capacidad de generalizar y encontrar patrones en gran escala, lo que permite identificar tendencias y relaciones que pueden haber pasado desapercibidas en el análisis cualitativo. Esto no solo enriquece la comprensión de las variables estudiadas, sino que también proporciona datos empíricos sólidos que respaldan las teorías y argumentos en campos como la psicología, la sociología y el análisis espacial.

Consecuentemente a lo indicado, la combinación de enfoques cuantitativos y cualitativos ofrece un panorama más completo y enriquecedor de los fenómenos sociales – urbanos, mientras que los análisis cualitativos pueden profundizar en la experiencia y la subjetividad, los enfoques cuantitativos agregan una dimensión clave para estas investigaciones, permitiendo un entendimiento más holístico y espacial que generan una base sólida para la formulación de políticas y la toma de decisiones.

Desde el lado de la gobernanza y planificación urbana local los ejes de salud, trabajo y economía no son competencias de entera responsabilidad del Municipio. Sin embargo, resulta conveniente entender que, desde las gestiones correspondientes se podrían diseñar propuestas, planes o programas de atención dirigidos hacia zonas específicas del Centro Histórico de Quito que tengan por objetivo mejorar la situación de la ciudadanía en temas de salud, a través de brigadas médicas, o en el caso del trabajo y economía generar espacios que promuevan el comercio ordenado en aquellas zonas donde se identifican bajos niveles de satisfacción para estas dimensiones, y que todo esto en conjunto mejore la situación de los habitantes residentes, evitando así el abandono y migración de más personas.

Por estas razones, en un primer momento, tanto los estudios como las propuestas de acción deben presentarse desde el Gobierno Municipal y que estos a su vez se eleven a las necesidades del Gobierno Nacional para desarrollar estrategias colaborativas en conjunto que permitan mejorar la gestión y accionar ante esta clase de problemáticas.

Finalmente, desde el ámbito de la planificación y política pública, es necesario entender que este tipo de investigaciones en el campo del análisis espacial del bienestar subjetivo son un buen indicador de la realidad que atraviesa el DMQ por tales razones, esta clase de datos deben ser levantados de forma periódica y extendidos a lo largo de la ciudad, ya que pueden convertirse en herramientas de información valiosa con miras hacia la actualización de los Planes de Desarrollo y Ordenamiento Territorial de Quito y la generación de recomendaciones de política pública basada en evidencia.

5.2.1 Objetivos, principales resultados y conclusiones.

Tabla 5.1. Matriz resumen de objetivos, resultados y conclusiones

Objetivos	Resultados	Conclusiones																				
<p>A través de ML, determinar qué factores demográficos, sociales y económicos son los que influyen más en la satisfacción con la vida de los habitantes del Centro Histórico de Quito</p>	<p>Con base a los análisis ML de aprendizaje supervisado y la metodología de árboles de decisiones (que obtuvo el valor más alto de R^2 con 0.367). Se puede determinar a través del análisis de la importancia relativa aquellos factores que influyen más a la satisfacción con la vida en el CHQ. Se detalla en la siguiente matriz:</p> <table border="1" data-bbox="647 703 1010 1091"> <thead> <tr> <th colspan="2" style="text-align: center;">Importancia relativa</th> </tr> <tr> <th colspan="2" style="text-align: center;">Árbol de decisión</th> </tr> <tr> <th style="text-align: left;">Variable</th> <th style="text-align: left;">Importancia</th> </tr> </thead> <tbody> <tr> <td>Trabajo</td> <td>26,06%</td> </tr> <tr> <td>Salud</td> <td>18,55%</td> </tr> <tr> <td>Economía</td> <td>17,21%</td> </tr> <tr> <td>Vivienda</td> <td>15,35%</td> </tr> <tr> <td>Vecinos</td> <td>13,18%</td> </tr> <tr> <td>Educación</td> <td>9,42%</td> </tr> <tr> <td>Ambiente</td> <td>0,23%</td> </tr> </tbody> </table> <p>Este mismo análisis se desarrolló para los modelos de redes neuronales con diferentes funciones de activación.</p>	Importancia relativa		Árbol de decisión		Variable	Importancia	Trabajo	26,06%	Salud	18,55%	Economía	17,21%	Vivienda	15,35%	Vecinos	13,18%	Educación	9,42%	Ambiente	0,23%	<ul style="list-style-type: none"> • El análisis de la importancia relativa revela que la satisfacción con el trabajo, la salud y la economía son factores preeminentes en la construcción del modelo, representando conjuntamente más del 50% de su influencia. Estas variables, al ocupar los primeros lugares tanto en el modelo lineal como en el de redes neuronales, subrayan su impacto significativo en la percepción general de la satisfacción con la vida. Sin embargo, resulta notable la escasa contribución de variables como la satisfacción con la vivienda, vecinos, educación y ambiente, todas ellas aportando menos del 15.35% cada una. • Adicionalmente, la baja participación de la variable satisfacción con el medio ambiente, evidenciada en la construcción del árbol de decisión, sugiere su limitada influencia en la determinación de la satisfacción general con la vida en el modelo propuesto.
Importancia relativa																						
Árbol de decisión																						
Variable	Importancia																					
Trabajo	26,06%																					
Salud	18,55%																					
Economía	17,21%																					
Vivienda	15,35%																					
Vecinos	13,18%																					
Educación	9,42%																					
Ambiente	0,23%																					

Objetivos	Resultados	Conclusiones
<p>Comparar diferentes modelos de ML para evaluar la satisfacción con la vida de los habitantes del Centro Histórico de Quito</p>	<ul style="list-style-type: none"> • Se desarrollaron en total 6 modelos de <i>machine learning</i>, 3 de aprendizaje no supervisado y los restantes de aprendizaje supervisado. • Para el caso de los modelos de aprendizaje no supervisado se realizaron corridas con la variable satisfacción general con la vida y también con las otras variables del modelo conceptual propuesto en la tesis. En todos los casos se realizó el análisis del gráfico del codo y su análisis de sensibilidad con diferentes pesos y número de <i>clusters</i>. • En lo que respecta a los modelos de aprendizaje supervisado, se planteó el mismo escenario de variable dependiente satisfacción general con la vida y el resto de las variables dependientes (trabajo, salud, economía, educación, vivienda, ambiente y relación con los vecinos). En todos los casos se evaluaron y compararon sus parámetros estadísticos para tener una contrastación objetiva entre diferentes metodologías. 	<ul style="list-style-type: none"> • Entre los métodos de aprendizaje no supervisado, <i>k-means</i> sobresale como el más eficiente para representar la estructura de los datos, generando grupos bien definidos y compactos que minimizan la varianza entre ellos. Esta capacidad se traduce en una representación más nítida y utilizable de la satisfacción con la vida y variables dependientes. No obstante, se reconoce que tanto el cluster jerárquico como skater tienen sus aplicaciones y pueden ser preferibles en situaciones que demanden estructuras más complejas o interpretables. • La elección entre k-means, cluster jerárquico y skater depende del contexto y los objetivos específicos del análisis. En un entorno más amplio y diverso, como la ciudad de Quito, el cluster jerárquico podría volverse más eficiente gracias a su capacidad para capturar estructuras de agrupación a múltiples escalas, especialmente relevante cuando la satisfacción con la vida varía según la ubicación geográfica, estrato socioeconómico y otros factores. Además, la flexibilidad interpretativa del cluster jerárquico, evidenciada en la generación de dendrogramas, sería valiosa para comprender las complejas relaciones entre grupos en una ciudad grande y heterogénea como Quito. • Entre los métodos de aprendizaje supervisado, el modelo de árbol de decisión destaca como la herramienta más efectiva para analizar la satisfacción con la vida, ofreciendo una representación visual clara de las complejas interacciones entre variables. Su capacidad para el análisis estadístico inferencial y la identificación de factores clave facilita la

Objetivos	Resultados	Conclusiones
		<p>priorización de intervenciones y políticas públicas basadas en evidencia científica, permitiendo abordar diversas áreas del bienestar de la población.</p> <ul style="list-style-type: none"> • En contraste, los modelos de redes neuronales, aunque flexibles en el aprendizaje de patrones complejos, carecen de interpretabilidad y pueden requerir grandes conjuntos de datos. El support vector machine destaca en la separación de clases, pero su manejo de conjuntos de datos grandes y su interpretación pueden ser desafiantes. Además, la representación geoespacial de datos, proporcionada por todos los modelos, resulta esencial para la planificación urbana, permitiendo abordar desigualdades geográficas de manera efectiva en el Centro Histórico de Quito.
<p>Expresar espacialmente los factores más influyentes en la satisfacción con la vida de los habitantes del Centro Histórico de Quito</p>	<ul style="list-style-type: none"> • Como parte de los análisis exploratorios se desarrollaron dos análisis de autocorrelación espacial teniendo en consideración la variable satisfacción con la vida y su distribución en el espacio. • El análisis de autocorrelación espacial <i>Moran's I</i> – LISA y puntos calientes y puntos fríos, brindaron una perspectiva espacial de los lugares en los que se forman conglomerados que comparten una alta/baja densidad de valores bajo de satisfacción con la vida en el Centro Histórico de Quito. 	<ul style="list-style-type: none"> • La aplicación de métodos de autocorrelación espacial, como LISA y G de Getis-Ord, en el análisis de la satisfacción con la vida, se enriquece aún más mediante la obtención de cartografía detallada que visualiza claramente las zonas identificadas en el Centro Histórico de Quito. Esta visualización geográfica destaca las áreas con agrupamientos espaciales de alta o baja satisfacción con la vida, proporcionando una representación gráfica crucial para la toma de decisiones. Al observar la cartografía resultante, los planificadores urbanos pueden identificar de manera efectiva patrones geoespaciales y factores contextuales que influyen en las percepciones subjetivas. Este enfoque espacial no solo complementa, sino que también potencia la capacidad de diseñar estrategias y políticas específicas dirigidas

Objetivos	Resultados	Conclusiones
		<p>a mejorar la calidad de vida en áreas particulares del Centro Histórico de Quito, permitiendo una intervención más precisa y efectiva.</p>
<p>Discutir la potencialidad de ML para analizar datos cualitativos, en el marco de las Ciencias Sociales y Estudios de la ciudad.</p>	<p>El desarrollo de la discusión se centra en 6 puntos de interés:</p> <p>1. Escalabilidad y limitaciones Temporales: La limitación de datos a una zona específica (Centro Histórico de Quito) y el factor temporal pueden afectar la actualidad y generalización de los resultados. La extrapolación de resultados a toda la ciudad puede omitir variaciones geoespaciales y socioeconómicas, limitando la comprensión completa de la calidad de vida.</p> <p>2. Análisis exploratorios con autocorrelación espacial: Los métodos de autocorrelación espacial proporcionan una perspectiva única para entender las dinámicas de la satisfacción con la vida. Identificación de clusters espaciales revela áreas con patrones similares, brindando oportunidades para diseñar estrategias específicas de mejora en políticas y programas de bienestar.</p> <p>3. Aprendizaje no supervisado y supervisado: El aprendizaje no supervisado, como k-means, skater, y cluster jerárquico, aborda la complejidad de datos subjetivos, identificando patrones ocultos. Los árboles de decisión se destacan como herramienta excepcional para analizar múltiples variables independientes y priorizar intervenciones basadas en la importancia relativa de cada factor.</p>	<ul style="list-style-type: none"> • Este análisis exhaustivo de la satisfacción con la vida en el Centro Histórico de Quito, empleando técnicas avanzadas como la autocorrelación espacial y el machine learning, ofrece valiosas perspectivas para la planificación urbana y las políticas públicas. A pesar de las limitaciones inherentes a la zona específica y al periodo temporal analizado, las metodologías aplicadas proporcionan una comprensión enriquecedora de las dinámicas espaciales y los determinantes de la calidad de vida. • El desafío de escalabilidad subraya la necesidad de extender el enfoque a una escala más amplia, permitiendo la captura de patrones complejos en toda la ciudad de Quito. El uso de técnicas no supervisadas y supervisadas, como k-means y árboles de decisión, resalta la importancia de identificar patrones ocultos y priorizar intervenciones basadas en factores clave, ofreciendo una visión clara y accesible de la influencia de diversas variables. • La comparación con estudios anteriores destaca la consistencia en la identificación de variables cruciales para la satisfacción con la vida, tanto a nivel individual como ambiental. La integración de datos georreferenciados y técnicas de interpolación facilita la visualización y comprensión de disparidades geográficas, proporcionando perspectivas valiosas para la asignación de recursos y la planificación urbana. • A pesar de limitaciones, como la focalización en la dimensión subjetiva de la calidad de vida, este análisis sienta las bases para investigaciones futuras

Objetivos	Resultados	Conclusiones
	<p>4. Comparación con estudios Previos: Los hallazgos coinciden con investigaciones anteriores, destacando la importancia de variables intraindividuales y factores ambientales y sociales en la satisfacción con la vida.</p> <p>5. Implicaciones para la planificación urbana: La información obtenida a través de técnicas de machine learning y análisis socioespacial es esencial para la formulación de políticas públicas y la planificación urbana. La complejidad de la calidad de vida requiere la integración de indicadores subjetivos y objetivos para una evaluación más equilibrada y completa.</p> <p>6. Limitaciones y futuras Investigaciones: La limitación en la representación de la calidad de vida solo a través de la satisfacción subjetiva podría limitar la comprensión total del concepto. La extrapolación de resultados a una escala mayor y la consideración de variables adicionales son esenciales para futuras investigaciones y análisis más completos.</p>	<p>que buscan una evaluación más equilibrada e integral. En última instancia, la información generada contribuye significativamente a la toma de decisiones informada, destacando la importancia de considerar tanto aspectos subjetivos como objetivos para abordar la complejidad de la calidad de vida en entornos urbanos en constante evolución.</p>

Fuente: elaborado por el autor

Referencias

- Abubakar, A. Mohammed. 2019. "Using hybrid SEM – artificial intelligence". *Personnel Review* 49 (1): 67-86. <https://doi.org/10.1108/PR-06-2017-0180>.
- Alarcón, R. 2006. "Desarrollo de una escala factorial para medir la felicidad". *Revista Interamericana de Psicología* 40: 99-106.
- . 2009. *Psicología de la felicidad. Introducción a la Psicología positiva*. II. Vol. 28. Lima.
- Alteryx. 2022. "Herramienta de diagnóstico de K-centroides". Agrupamiento predictivo. 2022.
- Anselin, Luc. 2017. "Spatially Constrained Clustering Methods". *Cluster Analysis* (3). 2017.
- . 2018. "Classic Clustering Methods". Geoda. 19 de noviembre de 2018.
- . 2020. "Local Spatial Autocorrelation (1) LISA and Local Moran". Geoda An introduction to espatual data science. 10 de diciembre de 2020.
- ArcGeek. 2018. "IDW ArcGeek". 2018. <https://acolita.com/interpolacion-con-la-distancia-inversa-ponderada-idw/>.
- Assunção, R. M., M. C. Neves, G. Câmara, y C. Da Costa Freitas. 2006. "Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees". *International Journal of Geographical Information Science* 20 (7): 797-811. <https://doi.org/10.1080/13658810600665111>.
- Barragán, Ahmad. 2013. "Aproximaciones científicas al estudio de la felicidad: ¿qué podemos aprender de la felicidad?" *Intercontinental de Psicología y Educación* 15 (julio): 7-24.
- Bialowolski, Piotr, y Dorota Weziak-Bialowolska. 2021. "Longitudinal Evidence for Reciprocal Effects Between Life Satisfaction and Job Satisfaction". *Journal of Happiness Studies* 22 (3): 1287-1312. <https://doi.org/10.1007/s10902-020-00273-1>.
- Bustamante-Ubilla, Miguel, María del Carmen Lapo-Maza, Laura Zambrano-Chumo, y María de los Ángeles Núñez-Lapo. 2020. "Factores de satisfacción con la vida en adultos mayores en zonas urbanas y rurales de Guayas, Ecuador". *Empresarial* 14 (1): 23-29. <https://doi.org/10.23878/empr.v14i1.174>.
- Carrión, Fernando. 2001. "Centro histórico: relación social, globalización y mitos". *Políticas y gestión para la sostenibilidad del patrimonio urbano*, 25-53.
- Caycho Rodríguez, Tomás. 2010. "Reseña de "Psicología de la felicidad. Introducción a la Psicología positiva" de Alarcon, R". *Red de revistas Científicas de América Latina, el Caribe, España y Portugal*.
- Chen, Yunsong, Xiaogang Wu, Anning Hu, Guangye He, y Guodong Ju. 2021. "Social prediction: a new research paradigm based on machine learning". *The Journal of Chinese Sociology* 8 (1): 15. <https://doi.org/10.1186/s40711-021-00152-z>.
- Crous, Gemma, Ferran Casas, y Mònica González-Carrasco. 2018. "What Aspects are Important to Adolescents to Achieve Full Satisfaction in Life?" *Child Indicators Research* 11 (6): 1699-1718. <https://doi.org/10.1007/s12187-018-9535-6>.
- Diener, Ed. 2000. "Subjective well-being: The science of happiness and a proposal for a national index." *American Psychologist* 55 (1): 34-43. <https://doi.org/10.1037/0003-066X.55.1.34>.
- Diener, Ed, Robert A. Emmons, Randy J. Larsen, y Sharon Griffin. 1985a. "The Satisfaction With Life Scale". *Journal of Personality Assessment* 49 (1): 71-75. https://doi.org/10.1207/s15327752jpa4901_13.
- . 1985b. "The Satisfaction With Life Scale". *Journal of Personality Assessment* 49 (1): 71-75. https://doi.org/10.1207/s15327752jpa4901_13.
- Dissart, J. -C., y Steven C. Deller. 2000. "Quality of Life in the Planning Literature". *Journal of Planning Literature* 15 (1): 135-61. <https://doi.org/10.1177/08854120022092962>.
- Duque, J, R Ramos, y J Surinach. 2007. "Métodos de regionalización supervisada: Una encuesta". *Revisión Internacional de la Ciencia Regional*, 195-220.
- Elastic. s. f. "¿Qué es el análisis de sentimiento?"
- ESRI - Arcgis Pro. 2022. "Cómo funciona Autocorrelación espacial (I de Moran global)". 2022.
- Evans, Michael, Dev Oliver, Xun Zhou, y Shashi Shekhar. 2010. *Big Data: Techniques and Technologies in Geoinformatics*.

- Franco, Giovanni di, y Michele Santurro. 2021. "Machine learning, artificial neural networks and social research". *Quality & Quantity* 55 (3): 1007-25. <https://doi.org/10.1007/s11135-020-01037-y>.
- Gamero Maldonado, Harold Yusit. 2013. "La Satisfacción Laboral como Dimensión de la Felicidad". *Ciencia & trabajo* 15 (47): 94-102. <https://doi.org/10.4067/S0718-24492013000200010>.
- Geograma. 2021. "Geo Inteligencia Artificial (GeoIA) y el futuro de los datos geográficos". 2021.
- Gómez, Josefina, Santiago Fernández, y Rafael Mata. 2001. "El Paisaje, calidad de vida y territorio". *Prisma Social*, 27-40.
- Gonzales, Ligdi. 2019. "Vectores de Soporte Regresión – Teoría". AprendIA. 1 de febrero de 2019.
- González, Andrés. 2019. "¿Qué es Machine Learning? by Andrés González". cleverdata. 2019.
- González, Lidgi. 2019. "Algoritmo de agrupamiento jerárquico". aprendeia. 2019.
- González, Ligdi. 2018. "Evaluando el error en los modelos de regresión". aprendeia. 23 de noviembre de 2018.
- . 2019. "Árboles de Decisión Regresión – Teoría". 2019.
- Gordón, Sara, Sandra Murillo, y Selene Hernández. 2018. "Satisfacción con la vida y desempeño social en México: un enfoque multidimensional". *Sociológica (México)* 33 (mayo).
- Gracia, Enrique, y Juan Herrero. 2006. "La comunidad como fuente de apoyo social: Evaluación e implicaciones en los ámbitos individual y comunitario". *Revista Latinoamericana de Psicología* 38: 327-42.
- Gualda, Estrella. 2022. "Social big data, sociología y ciencias sociales computacionales". *Empiria. Revista de metodología de ciencias sociales*, n.º 53 (enero). <https://doi.org/10.5944/empiria.53.2022.32631>.
- IBM. 2023. "Cómo funciona el aprendizaje supervisado". ¿Qué es el aprendizaje supervisado? enero de 2023.
- ICQ. 2016. "Análisis de la Encuesta Multipropósito DMQ". Encuesta multipropósito en el CHQ. 2016.
- Leva, Germán. 2005a. *Indicadores de calidad de vida urbana*. Buenos Aires: Universidad Nacional de Quilmes.
- . 2005b. "INDICADORES DE CALIDAD DE VIDA URBANA". *Hábitat metropolis*, 2005.
- Locke, E.A, y M.D Dunnette. 1976. "The Nature and Causes of Job Satisfaction." *Handbook of Industrial and Organizational Psychology*, 1297-1349.
- López, María José. 2007. "La calidad de vida subjetiva y su relación con las experiencias recreativas en los espacios naturales". En *VIII Jornadas Nacionales y II Simposio Internacional de Investigación Acción en Turismo*. Buenos Aires: Univerisdad Nacional de Mar del Plata.
- Lucero, Patricia, Claudia Mikkelsen, Fernando Sabuda, Silvina Aveni, y Ariel Ondartz. 2007. "Calidad de vida y espacio: Una Mirada Geográfica desde el Territorio Local". *Universidad Nacional de Mar del PLata* 7: 99-125.
- Lyubomirsky, S. 2008. "La ciencia de la felicidad. Un método probado para conseguir el bienestar". *Urano* .
- Macků, Karel, Jan Caha, Vít Pászto, y Pavel Tuček. 2020. "Subjective or Objective? How Objective Measures Relate to Subjective Life Satisfaction in Europe". *ISPRS International Journal of Geo-Information* 9 (5): 320. <https://doi.org/10.3390/ijgi9050320>.
- Martínez, Antón, María Buelga, Sofía Cava, y Jesús Marín. 2007. "La satisfacción con la vida en la adolescencia y su relación con la autoestima y el ajuste escolar". *Anuario de psicología* 38 (septiembre): 293-303.
- MDMQ. 2003. *Centro Histórico de Quito Plan Especial*.
- . 2011. *ECCO Distrito Metropolitano de Quito*.
- Menacho Chiok, Cesar Higinio. 2014. "Modelos de regresión lineal con redes neuronales". *Anales Científicos* 75 (2): 253. <https://doi.org/10.21704/ac.v75i2.961>.

- MIDUVI. 2016. “Revitalización del Centro Histórico de Quito”. *Subsecretaría de Hábitat y Asentamientos Humanos*, 2016.
- Mitchell, Andy. 2005. *La Guía de Esri para el análisis SIG*. Editado por Esri Press. Vol. 2.
- Moreno, Juan, y José Vera. 2011. “Modelo causal de la satisfacción con la vida en adolescentes de educación física”. *Revista de Psicodidáctica* 2: 367-80.
- Moreta-Herrera, Rodrigo, Claudio López-Calle, Paulina Gordón-Villalba, William Ortíz-Ochoa, y Ismael Gaibor-González. 2018. “El bienestar subjetivo, psicológico y social como predictores de la salud mental en ecuatorianos”. *Actualidades en Psicología* 32 (124). <https://doi.org/10.15517/ap.v32i124.31989>.
- Mouratidis, Kostas. 2021. “Urban planning and quality of life: A review of pathways linking the built environment to subjective well-being”. *Cities* 115 (agosto): 103229. <https://doi.org/10.1016/j.cities.2021.103229>.
- Navarro Astor, Elena. 2008. “Aportación al estudio de la satisfacción laboral de los profesionales técnicos del sector de la construcción: una aplicación cualitativa en la Comunidad Valenciana.” Valencia (Spain): Universitat Politècnica de València. <https://doi.org/10.4995/Thesis/10251/2189>.
- Pacione, Michael. 2003. “Urban environmental quality and human wellbeing—a social geographical perspective”. *Landscape and Urban Planning* 65 (1-2): 19-30. [https://doi.org/10.1016/S0169-2046\(02\)00234-7](https://doi.org/10.1016/S0169-2046(02)00234-7).
- Parra, Andrea. 2021. “Análisis de sentimiento. Qué es y cómo realizarlo”. QuestionPro. 2021.
- Rodríguez, Sebastián. 2023. “Evaluando la calidad de vida urbana en el centro histórico de Quito aplicando técnicas cuantitativas, espaciales y de machine learning”. Quito: Facultad Latinoamericana de Ciencias Sociales, FLACSO Ecuador.
- Rogge, Nicky, y Ilse Van Nijverseel. 2019. “Quality of Life in the European Union: A Multidimensional Analysis”. *Social Indicators Research* 141 (2): 765-89. <https://doi.org/10.1007/s11205-018-1854-y>.
- Rojas, Andreina. 2011. “Calidad de vida, calidad ambiental y sustentabilidad como conceptos urbanos complementarios”. *Revista venezolana de sociología y antropología*, mayo de 2011.
- Rojas, Mariano. 2012. “Quality of Life in Latin America and the Caribbean”. En *Handbook of Social Indicators and Quality of Life Research*, 529-45. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-2421-1_24.
- Rouhiainen, Lasse. 2018. *INTELIGENCIA ARTIFICIAL 101 COSAS QUE DEBES SABER HOY SOBRE NUESTRO FUTURO*. 1.^a ed. Barcelona: Planeta S.A.
- Sanz, Francisco. 2022. “K-Means Clustering: algoritmo, aplicaciones y desventajas”. K-Means Clustering: algoritmo, aplicaciones y desventajas. 2022.
- . 2023. “Algoritmo K-Means Clustering – aplicaciones y desventajas”. Algoritmo K-Means Clustering – aplicaciones y desventajas. 20 de mayo de 2023.
- Sarriera, Jorge Castellá, Ferran Casas, Bibiana Ramos dos Santos, Lívia Maria Bedin, y Mònica González. 2018. “Subjective well-being and personal relationships in childhood: Comparison of Brazilian and Spanish children”. *Interpersona: An International Journal on Personal Relationships* 12 (1): 91-106. <https://doi.org/10.5964/ijpr.v12i1.284>.
- Secretaría técnica Ecuador crece sin desnutrición infantil. s. f. “Acceso a vivienda adecuada y digna con pertenencia cultural”. Acceso a vivienda adecuada y digna con pertenencia cultural.
- Shen, Xiaofang, Fei Yin, y Can Jiao. 2023. “Predictive Models of Life Satisfaction in Older People: A Machine Learning Approach”. *International Journal of Environmental Research and Public Health* 20 (3): 2445. <https://doi.org/10.3390/ijerph20032445>.
- Smith, David Marshall. 1973. *The geography of social well-being in the United States: An introduction to territorial social indicators*. 1.^a ed. New York: McGraw-Hill.
- TIBCO. 2023a. “What is a Neural Network?” 2023.
- . 2023b. “¿Qué es el aprendizaje supervisado?” Ramas del Machine Learning. enero de 2023.

- Tkach, Chris, y Sonja Lyubomirsky. 2006. "How Do People Pursue Happiness?: Relating Personality, Happiness-Increasing Strategies, and Well-Being". *Journal of Happiness Studies* 7 (2): 183-225. <https://doi.org/10.1007/s10902-005-4754-1>.
- Tonón, Graciela. 2008. "Investigar la calidad de vida en Argentina". *Psicodebate* 8 (0): 141. <https://doi.org/10.18682/pd.v8i0.422>.
- Tonon, Graciela. 2010. "La utilización de indicadores de calidad de vida para la decisión de políticas públicas". *Polis (Santiago)* 9 (26). <https://doi.org/10.4067/S0718-65682010000200017>.
- Wolfewicz, Arne. 2022. "How Do Machines Learn? A Beginners Guide". *Levity*.
- . 2023. "Deep Learning vs. Machine Learning – What's The Difference?" *LEVITY*. 15 de febrero de 2023.