



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

MODELAMIENTO DE CAMBIOS EN LA DEMANDA DE TRANSPORTE PÚBLICO EN
SANTIAGO DE CHILE USANDO TÉCNICAS DE APRENDIZAJE DE MÁQUINA E
INTELIGENCIA ARTIFICIAL

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO/A CIVIL EN COMPUTACIÓN

SEBASTIÁN ALEJANDRO MONTEIRO PARADA

PROFESOR GUÍA:
EDUARDO GRAELLS-GARRIDO

SANTIAGO DE CHILE

2025

Capítulo 1

Introducción

El sistema de transporte público en Santiago de Chile es un componente esencial para el funcionamiento de la ciudad. Cambios en su oferta —sean planificados o inesperados— pueden generar impactos significativos en la movilidad de las zonas aledañas, tanto a corto como a largo plazo. Las motivaciones para estudiar estos cambios son diversas: desde promover un uso más eficiente de los recursos públicos al construir nuevas líneas de metro, hasta anticipar qué recorridos de buses podrían saturarse ante la suspensión parcial del servicio subterráneo. Comprender cómo estos eventos redistribuyen la carga dentro del sistema es clave para una planificación urbana más informada.

Una exploración bibliográfica sugiere que el campo de la predicción de la demanda usando técnicas de Machine Learning / Inteligencia Artificial (entendiendo que la segunda contiene a la primera) ha crecido notablemente. Un trabajo muy importante que ayudo mucho a la creación de este informe es el de Torrepadula et al. [5], quienes recopilaron cientos de papers de investigación resolviendo el problema de la demanda, resumiendo muy bien el enfoque de la solución, la fuente de sus datasets, la solución per sé, y conclusiones de los desempeños de las soluciones. Torrepadula menciona que el problema de la demanda es de tipo pronóstico de series de tiempo. En ese sentido, se abren varias soluciones, como el uso de RNN (Redes Neuronales Recurrentes), CNN (Redes Neuronales Convolucionales), SVR (Regresión de vectores de soporte), SVM (Máquinas de vectores de soporte), ELM (Máquinas de aprendizaje extremo) AE (Autoencoders) y Transformers, usados generalmente para lenguaje natural igual encontraron su uso en predicción de demanda. Lo mas interesante para mi, y en lo que basaré mi trabajo probablemente es en la solución usando Redes Neuronales de Grafos (GNNs) o en su forma convolucional, con RNN para capturar correlaciones temporales y espaciales.

Actualmente, algunos de los estudios que abordan esta problemática desde Chile lo hacen desde enfoques estadísticos y/o a nivel macro. Estos suelen analizar el antes y el después de una intervención, sin capacidad real de predicción. Otros modelos tienen una orientación más predictiva, pero se encuentran desactualizados y no reflejan adecuadamente las dinámicas actuales del transporte urbano. También existen enfoques centrados en el transporte privado, que estudian cómo factores como la infraestructura, las tarifas o las políticas públicas afectan la movilidad general. Sin embargo, estos trabajos no se enfocan en cambios estructurales de la red de transporte público, sino que operan sobre la oferta ya existente.

Estas limitaciones, sumadas al avance de nuevas tecnologías como la inteligencia artificial y al creciente acceso a datos relevantes —como las validaciones con tarjeta Bip!, el CENSO 2024 y registros de movilidad provistos por Entel— representan una oportunidad para plantear un nuevo enfoque. Este proyecto propone la creación de un modelo predictivo, transparente y flexible, capaz de emular la demanda de transporte público a partir de modificaciones en la oferta. El objetivo es anticipar el impacto de intervenciones en la red antes de que ocurran, brindando así una herramienta útil para la planificación y la toma de decisiones.

La solución propuesta se basa en el uso de técnicas de aprendizaje automático, modelando el sistema de transporte como un grafo en el que se representen recorridos, paradas y transbordos. Este modelo tendrá que aprender a predecir el comportamiento de los usuarios en función de múltiples factores, como la duración del viaje, el número de transbordos y el tiempo de espera. Estos modelos y sus resultados se compararán con datos reales de uso, para afinar el modelo y su precisión. Finalmente, se realizarán simulaciones de diferentes escenarios, como la introducción de nuevas líneas o la eliminación de recorridos, para observar cómo estos cambios afectan la demanda y la distribución de usuarios en la red obteniendo datos de la red y su uso modelado usando técnicas de ML y grafos.

1.1 Situación actual

Una gran motivación para este proyecto es la optimización en el uso de los recursos públicos. Modificar dinámicamente la frecuencia de los buses, crear nuevos recorridos o eliminar aquellos que han quedado obsoletos son decisiones que, actualmente, requieren largos procesos de análisis y evaluación. En este contexto, contar con un modelo que permita representar la red de transporte público de Santiago de Chile y simular cambios en su oferta, se vuelve una herramienta valiosa para la toma de decisiones más ágil y fundamentada.

Siguiendo el trabajo de Torrepadula mas a fondo [5] se abren muchas soluciones y consideraciones: La primera , el sujeto de la predicción.

1.1.1 Sujeto de la predicción :

Diversos trabajos se enfocan tanto en:

1. Cantidad de personas en una parada en la ruta.
2. Cantidad de personas en la ruta.
3. Cantidad de personas en un vehículo.
4. Cantidad de personas en un área.

Notar que cada enfoque o sujeto requiere un set de datos distintos, por ejemplo, para saber cuanta gente hay en un momento dado en un vehículo, debemos de usar cámaras o sensores, en cambio, para saber un estimado de gente en la ruta, podemos usar los datos de las validaciones de la tarjeta bip! de la ruta.

1.1.2 Tipo de datos:

El tipo de datos es importante. Algunos ejemplos son:

1. Datos de validación de la tarjeta Bip! (que se puede usar para saber cuanta gente hay en una ruta, o en un área).
2. Datos de sensores (que se pueden usar para saber cuanta gente hay en un vehículo).
3. Datos de cámaras (que se pueden usar para saber cuanta gente hay en un vehículo, o en un área o un paradero).
4. GPS para el flujo de personas en un área.

Citando a Torrepadula [5] , En general, los datos de validación de la tarjeta , como la bip o sus equivalentes en otros países son los más utilizados, ya que son fáciles de obtener y tienen una buena cobertura geográfica. Sin embargo, también tienen limitaciones, como la falta de información sobre el origen y destino de los viajes. Los datos de sensores y cámaras son más precisos, pero son más difíciles de obtener y tienen una cobertura geográfica limitada. Los datos de GPS son muy precisos, pero también son difíciles de obtener y tienen una cobertura geográfica limitada.

1.1.3 Factores

Los factores que afectan la demanda son diversos y pueden variar según el contexto. Algunos de los más relevantes son:

1. **Tarifas:** El costo del transporte público puede influir en la demanda, especialmente en áreas donde existen alternativas de transporte privado.
2. **Frecuencia:** La cantidad de buses o trenes disponibles en una ruta puede afectar la demanda, ya que una mayor frecuencia puede atraer a más usuarios.
3. **Tiempo de viaje:** La duración del trayecto es un factor clave en la decisión de utilizar el transporte público. Un tiempo de viaje más corto puede aumentar la demanda.
4. **Comodidad:** La calidad del servicio, como la limpieza, el confort y la seguridad, puede influir en la decisión de utilizar el transporte público.
5. **Accesibilidad:** La facilidad de acceso a las paradas o estaciones, así como la disponibilidad de servicios complementarios (como estacionamientos o bicicletas compartidas), puede afectar la demanda.
6. **Condiciones climáticas:** Factores como la lluvia, el frío o el calor extremo pueden influir en la decisión de utilizar el transporte público.
7. **Eventos especiales:** La realización de eventos masivos, como conciertos o ferias, puede generar picos de demanda en ciertas rutas.

8. **Fiestas y feriados:** La demanda de transporte público puede variar significativamente durante días festivos o feriados, lo que puede afectar la planificación de la oferta.
9. **Búsqueda Web** Los turistas, generalmente, se informan de las rutas y horarios de los buses en la web, por lo que el tráfico web puede ser un buen indicador de la demanda.

1.1.4 Modo de transporte

En distintos trabajos, se exploró la predicción de distintos métodos de transporte. Por ejemplo, de todos los papers analizados, la mayoría se enfocó en el Metro (por su cantidad de datos disponibles probablemente) y le seguía el bus.

1.1.5 Preprocesado de datos.

En este mismo paper, se habla de muchas formas de preprocesado. Entre ellas, crear una matriz o grafo para la red, Ingeniería de Características, normalización de datos, etc.

1.1.6 Técnicas de predicción

Lo mas importante, dada la cantidad de datos y la complejidad del problema, es el uso de técnicas de Machine Learning. En el trabajo de Torrepadula se habla de muchas técnicas, entre ellas:

La mayoría de los trabajos analizados usaron técnicas como RNN, GNN/GCNN.

Algunas ventajas y desventajas de las 3 mencionadas son:

- **RNN:** Ventajas: Captura correlaciones temporales, buena para series de tiempo multivariadas. Desventajas: No está diseñada para usarse con correlación espacial, es intensiva en recursos y tiene procesamiento paralelo limitado.
- **GNN/GCNN:** Ventajas: Captura correlaciones espaciales, buena para series de tiempo multivariadas. Desventajas: No captura correlaciones temporales, es intensiva en recursos, necesita la construcción del grafo.

Como podemos ver , una es el complemento de la otra. Segun Torrepadula, la mejor solución es usar una combinación de ambas, usando RNN para capturar correlaciones temporales y GNN/GCNN para capturar correlaciones espaciales.

De hecho, algunos autores han explorado hipergrafos, es decir, la topología de la red en un grafo y otro por encima que capture los caminos peatonales. Mas aún, se suelen usar LSTM para el espacio del tiempo.

1.1.7 En Chile...

Hoy en día, la red está enfrentando transformaciones importantes. La construcción e implementación de nuevas líneas de metro, como la Línea 7 y la futura Línea 8, tendrá un efecto profundo sobre el uso de ciertos recorridos de buses. Algunos servicios podrían volverse redundantes, mientras que otros —como los recorridos locales tipo [LETRA]-XX— podrían experimentar un aumento significativo en la demanda, al convertirse en alimentadores hacia las nuevas estaciones. Esta situación presenta una oportunidad para replantear frecuencias, redistribuir flotas y mejorar la eficiencia general del sistema.

Estos fenómenos han sido objeto de análisis en trabajos previos. Un ejemplo representativo es el de Ramírez [3], quien estudia el cambio espacial en la demanda de transporte público tras la apertura de una nueva línea de metro, empleando un enfoque estadístico. Si bien su análisis es útil para evaluar efectos pasados, no permite anticipar escenarios futuros ni explorar condiciones hipotéticas. El estudio concluye, entre otros puntos, que la cantidad de transbordos y la demanda por servicios locales aumentan tras la introducción de un servicio estructurante como una línea de metro.

Por otra parte, el trabajo de Camus [1] propone una simulación basada en agentes dentro de la red de transporte público. Sin embargo, dicho modelo considera la oferta como un elemento estático y no contempla escenarios en los que esta pueda ser modificada. Aun así, su enfoque representa un punto de partida interesante, ya que podría ser extendido para evaluar diferentes configuraciones de red.

También existe el modelo desarrollado para el Directorio de Transporte Público Metropolitano (DTPM) [2], mediante la consultora EMME de INRO (actualmente Bentley Systems), el cual segmenta la demanda en tres franjas horarias: punta mañana, bajo mañana y punta tarde. Este modelo, sin embargo, presenta limitaciones importantes: no es de código abierto, omite información relevante (como los aforos del sector oriente), y está basado en datos anteriores a la pandemia de COVID-19, específicamente de 2020, lo que afecta su vigencia y aplicabilidad.

Asimismo, existen modelos de demanda agregada, como el desarrollado por Méndez [4], que se apoyan en técnicas econométricas y estudian elasticidades en función de variables como tarifas o cantidad de servicios disponibles. Aunque valiosos, estos trabajos no abordan cambios estructurales en la red, sino que se enfocan en la oferta existente.

En resumen, los trabajos existentes suelen centrarse en enfoques estadísticos retrospectivos o en simulaciones que no permiten modificar dinámicamente la oferta. Esto deja un vacío importante: no existe una herramienta que permita analizar, de forma flexible y anticipada, cómo un cambio específico genera efectos en cascada sobre la red de transporte. En este contexto, se propone una nueva aproximación que permita comparar distintos estados de la red, con un enfoque predictivo y adaptable, apoyado en técnicas modernas de representación como grafos y aprendizaje automático.

1.2 Objetivos

1.2.1 Objetivo general

Modelar la red de transporte público de Santiago de Chile permitiendo hacer cambios en la oferta y evaluar su impacto en la demanda.

1.2.2 Objetivos específicos

1. Disponer de datos actualizados sobre el uso de transporte publico, como frecuencias e itinerarios y los destinos/origenes de los usuarios, como también, a de ser posible, de flujos de transporte.
2. Modelar la red de transporte publico en un grafo o hipergrafo de ser necesario, que permita representar la topología de la red de transporte y las combinaciones de ellas.
3. Modelar la demanda en sus dos aspectos, espacial y temporal. Para ello, se utilizará un GNN para capturar la topología de la red y una RNN para capturar la temporalidad de los datos. Se espera que el modelo sea capaz de predecir la demanda en función de múltiples factores, como la duración del viaje, el número de transbordos y el tiempo de espera.
4. Cambiar la topología de la red y observar cómo cambia la demanda . Cambiar la topología involucrará cambios de infraestructura (agregar, quitar o modificar rutas existentes) como también cambios en la frecuencia de los buses.
5. Analizar los datos de la nueva demanda.

1.2.3 Evaluación

Cada objetivo se verificaría de la siguiente manera:

1. Datos actualizados: Se espera contar con datos de validación de la tarjeta Bip! y registros de movilidad provistos por Entel, así como información censal sobre residencia y lugar de trabajo.
2. Modelado de la red: Se espera contar con un modelo de la red de transporte público que permita representar recorridos, paradas y transbordos. Para ello, se compara con trabajos previos que han utilizado modelos similares de modelado de las redes.
3. Modelo de ML para predicción: Se espera contar con un modelo de aprendizaje automático que simule el comportamiento de los usuarios en función de múltiples factores. Este modelo se validará comparando sus predicciones con datos reales de uso de transporte público, como los proporcionados por la tarjeta Bip!.
4. Al modificar la red, se espera que el modelo de ML pueda predecir cambios en la demanda y la distribución de usuarios en la red. Esto se validará instanciando diferentes escenarios y comparando los resultados con datos reales de uso. (Por ejemplo, red pre/post línea 6)
5. Análisis de resultados: Se espera realizar un análisis exhaustivo de los resultados obtenidos a partir de la simulación, identificando patrones y tendencias que puedan informar futuras decisiones en la red de transporte.

1.3 Solución propuesta

La solución propuesta se basa en la creación de un sistema de simulación del transporte público que combine estructuras de grafos y técnicas de aprendizaje automático. El enfoque contempla los siguientes componentes:

1. Modelado de la red como grafo: La red de transporte será representada como un grafo, donde los nodos corresponden a paradas o estaciones, y las aristas a tramos recorridos. Esta representación permitirá modelar recorridos compartidos (por ejemplo, buses distintos que recorren el mismo tramo), y considerar distintas características de cada servicio como atributos de las aristas: frecuencia, tiempo estimado, comodidad, etc.

También se va a explorar la creación del hipergrafo peatonal, ya que no todas las estaciones combinan (por ejemplo, caminar dos cuadras para ir de un lugar a otro).

2. GNN + RNN: Se implementará un modelo de aprendizaje automático para replicar la demanda de uso de transporte público en función de múltiples factores. Este modelo aprenderá a predecir el comportamiento de los usuarios en función de variables como la duración del viaje, el número de transbordos y el tiempo de espera. Se utilizarán técnicas de aprendizaje supervisado para ajustar los parámetros del modelo, utilizando datos históricos de validaciones Bip! y patrones de movilidad. Para ello se utilizará un modelo con GNN + RNN (por ejemplo, una LSTM) . Una GNN procesará la estructura espacial del grafo y la demanda histórica con una LSTM.
3. Entrenamiento y ajuste del modelo: Utilizando datos históricos (validaciones Bip!, patrones de movilidad, datos censales), se ajustarán los parámetros del modelo de ML para que el comportamiento simulado refleje lo más fielmente posible la realidad. Esto puede abordarse como un problema de optimización o incluso como un sistema de aprendizaje supervisado.
4. Ajustes a la oferta: Con el modelo calibrado, se podrán introducir cambios en la red (nuevas líneas, suspensión de servicios, variaciones de frecuencia) y observar cómo cambia la distribución de la demanda. Esto permitirá anticipar efectos como saturación de recorridos, desplazamiento de flujos o desuso de servicios.
5. Análisis de resultados: Finalmente, se realizará un análisis exhaustivo de los resultados obtenidos: se evaluarán métricas como tiempos promedio de viaje, número de transbordos, uso por línea y comparativas entre escenarios. El objetivo es que este análisis brinde insumos para decisiones estratégicas en la planificación del sistema de transporte.

En la figura 1.1 se presenta un diagrama de la solución propuesta, que ilustra los componentes y flujos de información del sistema.

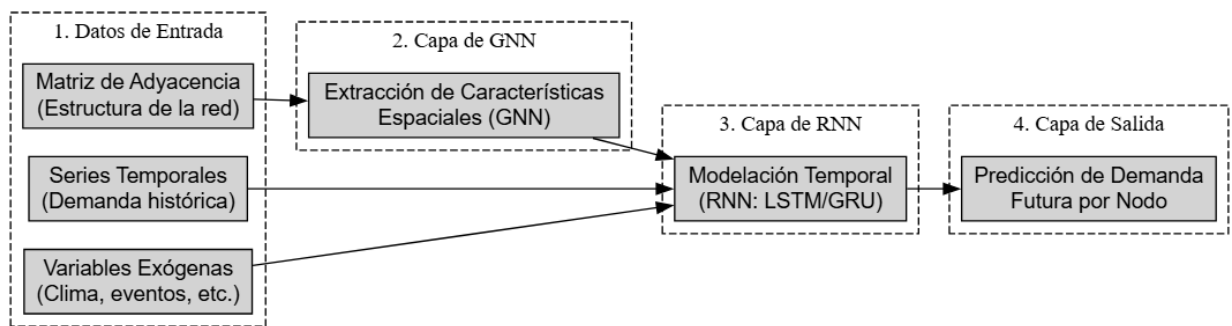


Figura 1.1: Diagrama de solución

1.4 Plan de trabajo

Tabla 1.1: Carta Gantt.

Tarea	Mes 1	Mes 2	Mes 3	Mes 4
Obtención y limpieza de datos	X__			
Análisis exploratorio de los datos	_X_			
Parseo de datos a grafo		_XX		
Validación de la estructura de datos.		_XX		
Crear y optimizar modelo de ML para uso de red			XXX_	
Comparar modelo de ML de uso con los reales			_XXX	
Con la red hecha y el modelo de ML validado, experimentar con cambios en la oferta modificando la red				XX__
Analizar los cambios de la demanda y ajustar el modelo según resultados				_XX
Redactar memoria y preparar defensa.				XXXX

1.5 Trabajo adelantado

%%

Bibliografía

- [1] Cayul, L.H.C. 2017. *Desarrollo y aplicación de modelo de simulación basada en agentes a gran escala para la ciudad de santiago*. Universidad de Chile.
- [2] Dirección de Transporte Público Metropolitano 2024. Modelos de demanda. <https://dtpm.cl/index.php/documentos/modelos-de-demanda>.
- [3] Ramírez, Á.E.T. 2020. *Análisis espacial de los impactos en la demanda de transporte público producto de una nueva línea de metro utilizando datos masivos*. Universidad de Concepción.
- [4] Soto, F.J.M. 2023. *Estimación y análisis de modelos de demanda agregada para el transporte público en santiago de chile*. Universidad de Chile.
- [5] Torrepadula Franca, R. di et al. 2024. Machine learning for public transportation demand prediction: A systematic literature review. *Engineering Applications of Artificial Intelligence*. 126, (2024), 107234. DOI:<https://doi.org/10.1016/j.engappai.2023.107234>.