



**UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL**

**MACHINE LEARNING PARA PREDECIR VOLÚMENES
OPERACIONALES DE LAS LÍNEAS DE NEGOCIO DE
CENABAST**

**TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN GESTIÓN Y POLÍTICAS
PÚBLICAS**

SEBASTIÁN MAURICIO MADARIAGA TORRES

PROFESOR GUIA:

ERWIN FISCHER ANGULO

PROFESOR CO-GUÍA:

JAVIER FARIAS SOTO

COMISIÓN:

EZEQUIEL MUÑOZ KRSULOVIC

**SANTIAGO DE CHILE
2021**

RESUMEN DE LA TESIS PARA OPTAR AL GRADO DE: Magíster en Gestión y Políticas Públicas.
POR: Sebastián Mauricio Madariaga Torres
FECHA: 31/08/2021
PROFESOR GUIA: Erwin Fischer Angulo

MACHINE LEARNING PARA PREDECIR VOLÚMENES OPERACIONALES DE LAS LÍNEAS DE NEGOCIO DE CENABAST

La presente tesis, tiene el objetivo de utilizar modelos de Machine Learning para predecir volúmenes operacionales de las líneas de negocio de Cenabast. Las líneas de negocio son conformadas por medicamentos que poseen tres características su alta dispersión en el precio, la inelasticidad del precio y la regresividad del precio. Dada las características del mercado del medicamento la producción de parte de los proveedores y demanda de los productos por parte de las personas, tienen un impacto directo en la salud de la población. Por lo anterior, es de gran importancia predecir los volúmenes operacionales de las líneas de negocio de Cenabast, ya que, a través de las líneas de negocio se busca mejorar las condiciones sanitarias de la población.

Para el desarrollo de la tesis se utilizan datos históricos de abastecimiento entregados por Cenabast de las compras desde el año 2011 al 2021. Donde, en primera instancia los medicamentos son clasificados según sus similitudes en su coeficiente de variación y cantidad promedio de datos, como resultado del total de la base de datos se utilizó el 21% equivalentes a 441 medicamentos, debido que al obtener gran cantidad de datos y bajo coeficiente de variación de los productos, se presentan datos más homogéneos para utilizar modelos de Machine Learning.

En los 441 medicamentos, se utilizan seis modelos Machine Learning los cuales son Lineal Regression, Decision Tree Regression, Random Forest Regression, Ridge Regression, Lasso Regression y Elastic Net Regression, con el objetivo de predecir la demanda de los medicamentos. Se logra concluir, que el mejor modelo utilizando parámetros por defecto Lineal Regression con un coeficiente de determinación de 72,76%, y en caso de utilizar el método de hiperparámetros Decisión Tree Regressión con un coeficiente de determinación de 73,14%.

Dedicado a mi hija, por ser mi alegría, mi luz y felicidad.

Agradecimientos

Agradezco a mi familia, en especial a mi abuela por el apoyo en mis años de dedicación al magíster.

A mi madre por enseñarme la vocación de servicio y hermano, por recordarme que con humildad y esfuerzo uno puede lograr los sueños.

Tabla de contenido

Tabla de contenido	iv
Índice de figuras	vii
Índice de Tablas	viii
1. Antecedentes generales	1
1.1 Descripción del marco normativo institucional.....	3
1.2 Caracterización de Cenabast.	4
1.3 Programas ministeriales de Cenabast.....	8
2. Pregunta de investigación	9
3. Objetivos	9
3.1 Objetivo general	9
3.2 Objetivos específicos.....	9
4. Resultados esperados y alcances	10
5. Marco Conceptual	10
5.1 El mercado de los medicamentos en Chile	10
5.2 Gasto en salud en Chile.....	12
5.3 Valor público en el uso de datos en el sector público	14
6. Metodología de investigación	18
6.1 Metodología de estudio (cuantitativa).....	18
6.1.2 Comprensión de los datos	19

6.1.3 Preparación de datos	19
6.1.4 Modelado	20
6.1.5 Evaluación.....	20
6.2 Tipo de estudio (predictivo).....	20
6.2.1 Lineal Regression.....	21
6.2.2 Decision Tree Regression	22
6.2.3 Random Forest Regression	22
6.2.4 Regularization	23
6.2.5 Elastic Net Regression	24
6.3. Fuentes de datos a utilizar.....	24
6.4 Plan de análisis.....	25
6.4.1 Análisis de base de datos de productos	25
6.4.2 Análisis de base de datos de abastecimiento.....	26
6.4.3 Análisis de base de datos de contratos vigentes.....	31
6.4.4 Análisis de base de datos de código internos de Cenabast.....	32
7. Preparación de datos para el modelamiento.....	33
7.1 Análisis descriptivo de los productos de Cenabast	33
7.2 Clustering de los productos de Cenabast	34
8. Modelamiento Machine Learning de datos.....	35
8.1 Datos de entrenamiento y testeo	35

8.2 Resultados de los modelos Machine Learning	36
8.2.1 Resultados de los modelos utilizando parámetros por defecto	36
8.2.2 Resultados de los modelos utilizando hiperparámetros	36
8.2.3 Comparación de resultados	37
9. Conclusión	38
10. Bibliografía	39

Índice de figuras

Figura 1: Variación Mensual (%) del precio de "Productos, artefactos y equipos médicos".....	2
Figura 2: Distribución de volumen de venta total según mecanismo de compra.	6
Figura 3: Distribución de los clientes por región para el año 2019.	7
Figura 4: Cantidad de clientes por categoría para el año 2019.	7
Figura 5: Estructura funcional del Sistema de Salud Pública.	12
Figura 6: Comparación de la variación anual entre el gasto de salud pública y asistencia, y PIB para Chile, entre los años 2017 al 2019.	13
Figura 7: Resultados del Índice de Gobierno Digital OCDE, para el año 2019.	17
Figura 8: Etapa del modelo de proceso CRISP-DM.	18
Figura 9: Ejemplo de árbol de decisiones.	22
Figura 10: Cantidad de productos registrados de Cenabast.	25
Figura 11: Comparación entre precio de mercado y Cenabast del producto "Paracetamol 500 mg cm/cm rec".	26
Figura 12: Proporción de cada mecanismo de compra del total de ventas netas CLP, por año.	29
Figura 13: Porcentaje de productos de Cenabast por Cluster.	34
Figura 14: Representación visual de datos de entrenamiento y testeo.	35

Índice de Tablas

Tabla 1: Volumen de Venta de cada línea de negocio de Cenabast, en MM\$.....	5
Tabla 2: Cantidad de documentos emitidos por año y variación con respecto al año anterior.	26
Tabla 3: Cantidad y porcentaje de documentos por tipo de compra.....	27
Tabla 4: Ventas netas CLP por tipo de compra.	28
Tabla 5: Ventas netas CLP por mecanismo de compra y año.....	28
Tabla 6: Ventas netas CLP para cada línea de negocio por año.	29
Tabla 7: Ranking de productos más comprados por Cenabast por cantidad unitaria, para el año 2019.	30
Tabla 8: Ranking de productos más comprados por Cenabast por Ventas netas CLP para el año 2019.....	30
Tabla 9: Cantidad de documentos emitidos de los contratos vigentes de Cenabast, por año.	31
Tabla 10: Ventas netas CLP total por año de los contratos vigentes de Cenabast.....	31
Tabla 11: Ranking de los principales proveedores de Cenabast por Ventas netas CLP para el año 2019.	32
Tabla 12: Análisis descriptivo por producto de Cenabast.....	33
Tabla 13: Tabla de resultados de la caracterización de Clustering de los productos de Cenabast.	34
Tabla 14: Resultados de los modelos de Machine Learning.....	36

Tabla 15: Resultados de los modelos de Machine Learning utilizando hiperparámetros.
.....36

Tabla 16: Comparación porcentual entre los modelos con parámetros por defecto e hiperparámetros.....37

1. Antecedentes generales

Los medicamentos son bienes de consumo. Pero no son normales. Desde el punto de vista económico registran particularidades que los diferencia de cualquier otro bien. Si se le considera como un bien de salud constituye el recurso médico y terapéutico usado con mayor frecuencia. Esto significa que las limitaciones en su producción, circulación y consumo tiene un altísimo impacto sobre la salud de la población y fundamentalmente sobre la distribución de los stocks de salud dentro de la sociedad. La consecuencia más relevante de tales limitaciones (o “fallas”) es el desajuste entre los recursos ofertados y las necesidades de salud de la población, esto configura lo que ha sido denominado el problema de acceso (Tobar y Sánchez, 2005, p. 24).

Tobar (2008), explica que el mercado del medicamento presenta 3 principales características:

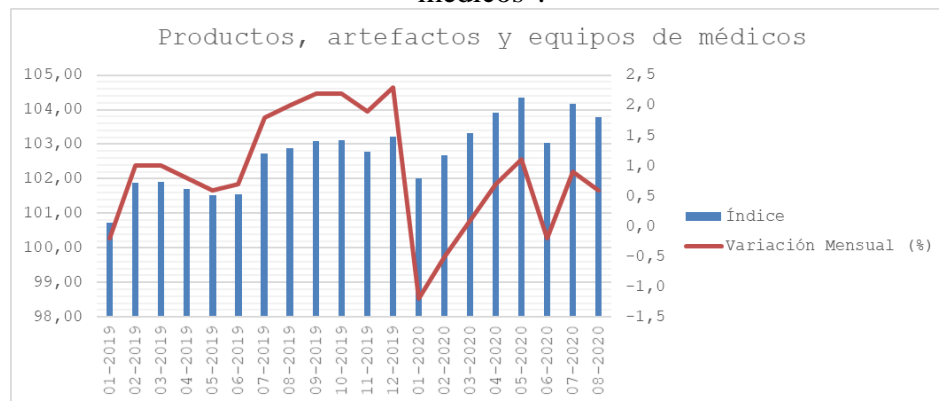
1. La alta dispersión de precios: no existe ningún otro bien que presente tanta diferencia en precio entre productos que responden a la misma especificación técnica. En este caso, no es un problema local o regional. No discrimina el tipo de economía. La volatilidad de sus precios puede llegar a duplicar la de alimentos y otros bienes industriales. Se trata de la consecuencia de toda una compleja lógica de producción, comercialización (incluyendo, aquí, prescripción y dispensación) así como de utilización de los productos que propician su diferenciación.
2. La inelasticidad precio: las cantidades de medicamento demandadas varían menos que la variación de su precio, proporcionalmente. Lo anterior genera una externalidad negativa en los consumidores. La baja concentración de empresas por el lado de la oferta (oligopolios) hace que las mismas tengan cierto control sobre el precio, con lo cual éstas podrían elevar sus precios sin que ello genere grandes disminuciones en la demanda.
3. La regresividad de su financiamiento: los pobres gastan, proporcionalmente, mucho más que los ricos en medicamentos, porque registran mayores necesidades relativas y disponen de menores condiciones para adquirirlos. Esta situación es mucho más grave en los países en desarrollo, ya que en ellos una porción mayor

de la financiación de los medicamentos depende de los ingresos de los hogares. En otras palabras, en los países desarrollados el gasto en medicamentos es predominantemente financiado con recursos públicos mientras que en los países en desarrollo se financia con recursos privados.

El sector público representa aproximadamente un 30% del total de compras de medicamentos del país. Un análisis econométrico de las compras realizadas a través del sistema de compras públicas (Chilecompras) confirma que el Estado obtiene mejores precios cuando compra volúmenes altos y esto ocurre principalmente cuando compra Cenabast (FNE, 2019, p.23).

Según la información del Índice de Precios al Consumidor (IPC)¹ publicado por el Instituto Nacional de Estadísticas, los productos, artefactos y equipos médicos presentan una variación anual del 3,04% entre enero 2019 y agosto 2020. En la Figura 1 se puede observar el comportamiento de la variación mensual del precio de productos, artefactos y equipos médicos, donde se puede identificar el alto grado de volatilidad entre periodos del precio.

Figura 1: Variación Mensual (%) del precio de "Productos, artefactos y equipos médicos".



Fuente: INE Chile.

¹ El Índice de Precios al Consumidor, es un indicador económico que mide mes a mes la variación de los precios de una canasta de bienes y servicios representativa del consumo del país.

Otro aspecto relevante, según los resultados de la VIII Encuesta de Presupuesto Familiar (EPF)², informado por el Instituto Nacional de Estadísticas, es que el gasto de productos, artefactos y equipos médicos corresponden al 36,32% del gasto promedio mensual por hogar utilizado en la división de salud, y comparando el gasto promedio mensual entre la VIII y VII EPF, aumentó un 51,72% en el mismo ítem.

Los antecedentes presentados nos muestran que en Chile se cumplen las tres características del mercado de medicamentos, su alta dispersión de los precios, la inelasticidad de los precios y la regresividad de su financiamiento, según lo argumentado por Tobar (2008).

1.1 Descripción del marco normativo institucional

La Central de Abastecimiento del Sistema Nacional de Servicio de Salud (Cenabast), surge con el Decreto N° 78 de 1980, publicado en el diario oficial el 04 de junio de 1980, donde su objetivo es proveer de medicamentos, equipos, instrumentales y demás elementos o insumos necesarios para la ejecución de acciones de salud a los organismos y personas que integran ese sistema.

La Central dependerá del Ministerio de Salud para los efectos de someterse a la supervigilancia de éste en su funcionamiento y a cuyas políticas, normas y planes generales deberá sujetarse en el ejercicio de sus actividades, en la forma y condiciones que determine la presente ley, donde las principales funciones son:

- a) Proveer de medicamentos, artículos farmacéuticos y de laboratorio, material quirúrgico, instrumental y demás elementos e insumos que se requieran para el ejercicio de las acciones de salud a que se refiere el artículo 46. Para estos fines, la Central podrá adquirir, almacenar, distribuir, transportar, arrendar y vender esos elementos a los organismos, entidades, establecimientos y personas que formen

² La Encuesta de Presupuestos Familiares (EPF) es una encuesta socioeconómica aplicada a hogares, cuyo propósito es recopilar información sobre gastos en los que estos incurren y los ingresos que perciben en un período de tiempo determinado.

parte del Sistema, con el solo objeto de cumplir los planes y programas del Ministerio y a los demás organismos públicos, entre cuyos fines institucionales esté la realización de acciones de salud en favor de sus beneficiarios; de conformidad al reglamento.

- b) Mantener en existencia una cantidad adecuada de elementos de la misma naturaleza, determinados por el Ministerio de Salud, necesarios para el eficiente cumplimiento de sus programas;
- c) Atender las necesidades que en las materias de su competencia le encomiende satisfacer el Supremo Gobierno, en caso de emergencias nacionales o internacionales, y
- d) Prestar servicios de asesoría técnica a otros organismos y entidades del Sistema.

1.2 Caracterización de Cenabast.

Cenabast es una institución pública y descentralizada dependiente del Ministerio de Salud. Su misión es “contribuir al bienestar de la población, asegurando la disponibilidad de medicamentos, alimentos, insumos y equipos a la Red de Salud, mediante la gestión de un servicio de abastecimiento de excelencia, eficiente y de calidad, para mejorar la salud, y su visión es liderar el mercado de suministros de la salud, siendo reconocidos a nivel nacional e internacional como un referente de la gestión del servicio de abastecimiento, a través de procesos innovadores y estrategias de alto impacto, con un equipo de expertos y especialistas que trabajan de forma armónica y coordinada, comprometidos por la equidad y el bienestar de la población (Cenabast, 2020).

Cenabast (2020) declara que sus tres objetivos estratégicos son:

1. Generar ahorro al Sistema Público de Salud mediante la agregación de demanda, aumentando el poder de negociación para obtener mejores precios.
2. De la adquisición de productos no presentes en el mercado, debido a su baja demanda a través de mecanismos de compra como importaciones o vía OPS, para asegurar su abastecimiento. Aumentar el número de productos bioequivalentes que no tengan bioequivalencia en Chile.

3. Conseguir fuentes alternativas de abastecimiento para medicamentos de alto costo, para garantizar el acceso de quienes lo requieran.

También Cenabast, mediante sus memorias anuales, informa su gestión institucional, donde se puede extraer la siguiente información financiera y operacional de la institución.

Existen dos líneas de operaciones, Intermediación y Programas Ministeriales. La primera consiste en consolidar la demanda de dispositivos médicos de los establecimientos de la red pública de salud, con el objetivo de obtener un volumen de compra que permita acceder a descuentos en el precio. La segunda consiste en la ejecución de Programas Ministeriales mandatados por el Ministerio de Salud, Subsecretaría de Redes Asistenciales, Subsecretaría de Salud Pública y el Fondo Nacional de Salud.

En la Tabla 1 se puede identificar la evolución entre los años 2016 al 2019 de los volúmenes de operación para cada línea de operación, informados en la Memoria Anual de Cenabast del año 2019.

Tabla 1: Volumen de Venta de cada línea de negocio de Cenabast, en MM\$.

Línea de operación	Monto 2016⁴	Monto 2017⁴	Monto 2018⁴	Monto 2019⁴
Intermediación	\$114.082	\$136.480	\$149.993	\$202.648
Programa Ministeriales	\$311.687	\$338.467	\$343.444	\$474.065
Ley Ricarte Soto	\$38.830	\$49.910	\$66.729	\$109.358
VIH/SIDA	\$79.826	\$78.458	\$55.252	\$93.607
PNI ¹	\$30.541	\$32.360	\$30.951	\$44.087
PNAC ²	\$40.075	\$39.637	\$36.763	\$39.539
PACAM ³	\$18.577	\$22.555	\$21.766	\$23.962
Otros Programas Ministeriales	\$103.839	\$115.546	\$131.983	\$163.511
Total	\$425.769	\$474.947	\$493.437	\$676.712
Variación Anual	-	11,6%	3,9%	37,1%

¹ Programa Nacional de Inmunizaciones.

² Programa Nacional de Alimentación Complementaria.

³ Programa de Alimentación Complementaria del Adulto Mayor.

⁴ En millones de pesos (MM\$)

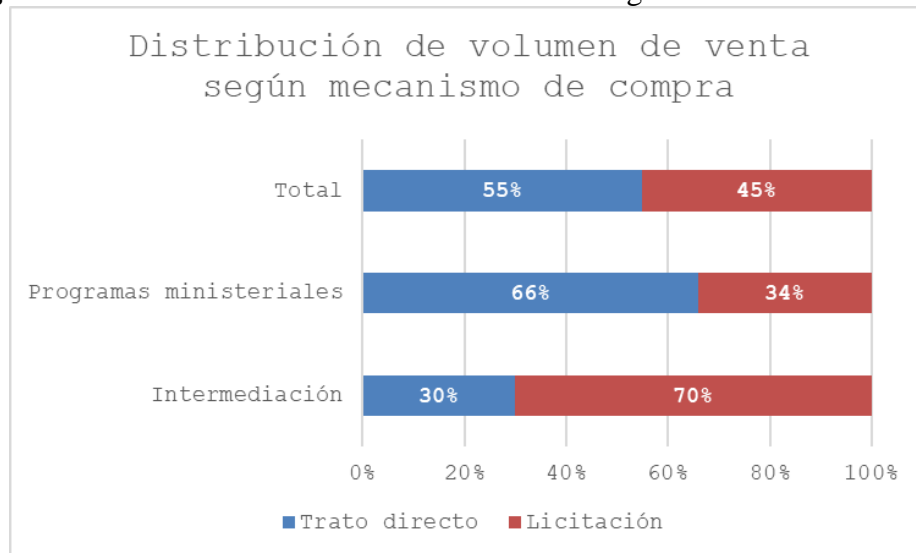
Fuente: Memoria Anual de Cenabast, año 2019.

La institución se rige por la Ley de compra N° 19.886 donde, según las cifras del portal web datos abiertos (Chilecompras, 2020), desde el año 2017 al año 2019, se han generado 24.763.537 órdenes de compras, y el sector Salud corresponde el 29,90% de las

transacciones realizadas, principalmente para servicios de salud, hospitales y al Ministerio de Salud.

En la Figura 2 se puede identificar la distribución del volumen de venta por la línea de negocio, según el mecanismo de compra. Se puede analizar que, para la línea de operación Programas Ministeriales, el principal mecanismo de compra es de Trato directo con un 66%, caso contrario de la línea de operación Intermediación donde prevalece el mecanismo de Licitación.

Figura 2: Distribución de volumen de venta total según mecanismo de compra.

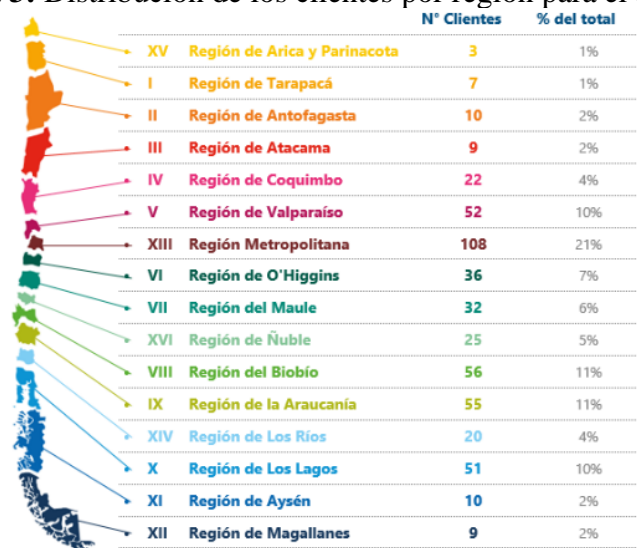


Fuente: Memoria Anual de Cenabast, año 2019.

Los clientes de Cenabast comprenden a toda la red pública del país. Según Cenabast (2020), 505 clientes realizaron solicitudes de productos a través de su línea de operación Intermediación, distribuidos en todo el territorio nacional, tal como se puede observar en la Figura 3.

Los principales clientes se concentran en la Región Metropolitana, con un 21%, seguido de la Región del Biobío y La Araucanía, con un 11%.

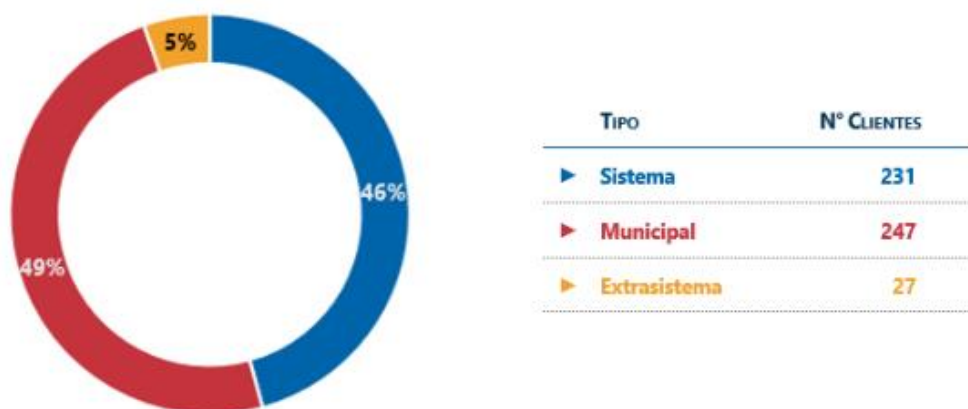
Figura 3: Distribución de los clientes por región para el año 2019.



Fuente: Memoria Anual de Cenabast, año 2019.

Los clientes se pueden segmentar en tres categorías: Sistema, con dependencia del Servicio de Salud; Municipal, con dependencia municipal; y Extrasistema, otros organismos. En la Figura 4, se puede identificar la cantidad de clientes por categoría.

Figura 4: Cantidad de clientes por categoría para el año 2019.



Fuente: Memoria Anual de Cenabast, año 2019.

Durante el 2019, un total de 226 proveedores presentaron 2.955 contratos con Cenabast para el abastecimiento de 1.487 productos, es importante destacar que un producto puede tener más de un proveedor. En comparación con el año 2018, Cenabast aumento en 20 la cantidad de proveedores, 518 los contratos, impactando en 175 productos adicionales, comparando los resultados de Cenabast (2019).

1.3 Programas ministeriales de Cenabast

Ejemplos de políticas públicas del sector es la Ley 20.850 o Ley Ricardo Soto, publicada en el mes de junio de 2015, con el objetivo de crear un sistema de protección financiera para diagnósticos y tratamiento de alto costos. Según datos de la Superintendencia de Salud (2019), han sido beneficiados 21.030 ciudadanos, principalmente con el sistema de salud Fonasa (76,1%), Isapre (22,0%) y FF.AA. y de Orden (1,9%).

De la Ley Ricardo Soto se puede concluir que las agrupaciones sociales son capaces de generar conciencia y presionar al Estado; las agrupaciones pueden averiguar la información que los médicos no tienen: suplencia. Este es un indicativo de que la ciudadanía en Chile ha cambiado, los enfermos han logrado presionar, hacer lobby. Se puede argumentar incluso que la Ley Ricarte Soto no se entiende si no hay una ciudadanía, una sociedad civil que esté activa y propositiva hacia sus condiciones de salud. No obstante, parece también relevante el hecho de que el cálculo utilizado en la medicina social moderna, el del costo/efectividad, demuestra en estas circunstancias un claro desajuste, dado que lo que está en juego en estas enfermedades son las condiciones específicas del afectado, la capacidad tecnológica y de investigación médica para intervenir en la enfermedad, pero sobre todo los cambios en la calidad de vida. Es decir, lo que está en cuestión es el sistema basado en la efectividad, su soporte y su definición. Como nos señaló un experto: ¿Cómo priorizar con arreglo a una identidad, una salud, una población, el dolor o la vida de un niño que requiere una atención de alto costo para una sobrevivida de 30 años? (Seguel y Zabala, 2019, p.336).

También durante el año 2019, a través del Instituto de Salud Pública de Chile, el Ministerio de Salud presentó su política nacional de medicamentos, con tres objetivos principales: aumentar la disponibilidad de los productos farmacéuticos; disminuir el gasto de bolsillo de las familias; y asegurar la calidad de los medicamentos que se comercializan en el país, a través de 30 medidas, donde Cenabast tiene un rol fundamental en el establecimiento de convenios de importaciones de medicamentos, buscando que el sector público adquiera medicamento a un menor precio y aumentar el número de medicamentos que intermedia (Ministerio de Salud, 2020).

2. Pregunta de investigación

Las instituciones gubernamentales manejan gran cantidad de información (o datos), que no siempre son utilizados para la toma de decisiones. Por lo anterior, en el siguiente documento se busca responder las siguientes preguntas, ¿Cuál será la demanda de productos de las líneas de negocio de Cenabast para los siguientes periodos?, y ¿Qué modelo de Machine Learning se debe utilizar para estimar la demanda y disminuir lo más posible el error?

Según Orellana (2017), las tecnologías en sí constituyen valor público, ellas permiten a los gobiernos una mejor comunicación para el cumplimiento de sus objetivos de bien común, y a los ciudadanos participar de lo público, con herramientas digitales que faciliten el diseño e implementación de políticas públicas.

La pregunta de investigación busca comprobar que es posible realizar estimaciones de las líneas de negocio de Cenabast, utilizando modelos Machine Learning, a través de la demanda de producto, según las principales variables identificadas que impactan en la demanda de productos, con el fin de poder obtener contratos más atractivos al negociar por mayores volúmenes de productos y generar un ahorro en el sector público.

3. Objetivos

A continuación, se definen los objetivos que permiten responder las preguntas de investigación definida en el capítulo 2.

3.1 Objetivo general

Desarrollar modelos de Machine Learning que permitan estimar la demanda de las líneas de negocio de Cenabast.

3.2 Objetivos específicos

1. Realizar un análisis descriptivo y exploratorio de los datos de Cenabast.

2. Generar modelos de Machine Learning que permitan predecir los volúmenes operacionales de las líneas de negocio de Cenabast.
3. Generar recomendaciones de estrategia de optimización para Cenabast, según los resultados obtenidos para la toma de decisiones.

4. Resultados esperados y alcances

El resultado esperado es identificar el o los mejores modelos de Machine Learning que permitan estimar la demanda de las líneas de negocio de Cenabast.

La tesis desarrollada busca generar una herramienta alternativa para la toma de decisiones por parte de Cenabast, utilizando la metodología de investigación CRISP-DM hasta la etapa 5 del proceso de minería de datos.

5. Marco Conceptual

En el capítulo marco conceptual se encuentra estructurado en base a tres temáticas, las primeras dos buscan caracterizar el mercado y estructura de gasto de los medicamentos en Chile, y el tercero el valor público en el uso de los datos en el sector público.

5.1 El mercado de los medicamentos en Chile

La dispersión de precios se produce cuando distintos vendedores ofrecen precios distintos por el mismo bien en un mercado determinado. Una simple explicación de la dispersión de precios es que surge de información imperfecta de parte de los consumidores que no todos compran a los más bajos precio vendedor porque, algunos al menos, no saben quién es el vendedor de precio más bajo. Es un tema importante en el campo de la economía de la información, ya que existe una considerable evidencia empírica de que la dispersión de precios es generalizada y significativa (Hopkins, 2006, p.1).

Otro aspecto importante, según el análisis de la Fiscalía Nacional Económica de Chile (2019), es que la demanda por medicamento clínico es fuertemente inelástica, es decir, los pacientes no tienen capacidad de sustituir los medicamentos clínicos que necesitan de

una receta. De esta forma, se estima que la discriminación de precios no redundará ni en un aumento ni en una caída en la demanda por medicamentos, no habiendo efectos perjudiciales en términos de bienestar o eficiencia económica, implicando únicamente una transferencia de excedentes desde el sector de alta disposición a pagar hacia los oferentes.

Desde el punto de vista de la demanda, la información sobre los productos del mercado de medicamentos es por naturaleza asimétrica: para conocer la forma de medicación y la calidad de los fármacos e insumos médicos se requiere de un conjunto de conocimientos técnicos que normalmente están fuera del alcance del consumidor medio. La elección de mercado del demandante depende en gran parte de terceros: del médico prescriptor o del farmacéutico. Pero si dicha elección depende del agente de venta, como es recurrente en Chile, y éste es impulsado primordialmente por el afán de lucro, se distorsiona gravemente las condiciones de racionalidad en la libre elección del demandante. Racionalidad que, se supone, es una de las bases de la economía de mercado (Alvear, 2013, p.140).

Vasallo (2010) determina que el funcionamiento y caracterización del mercado del medicamento, tiene tres características particulares respecto a otros mercados:

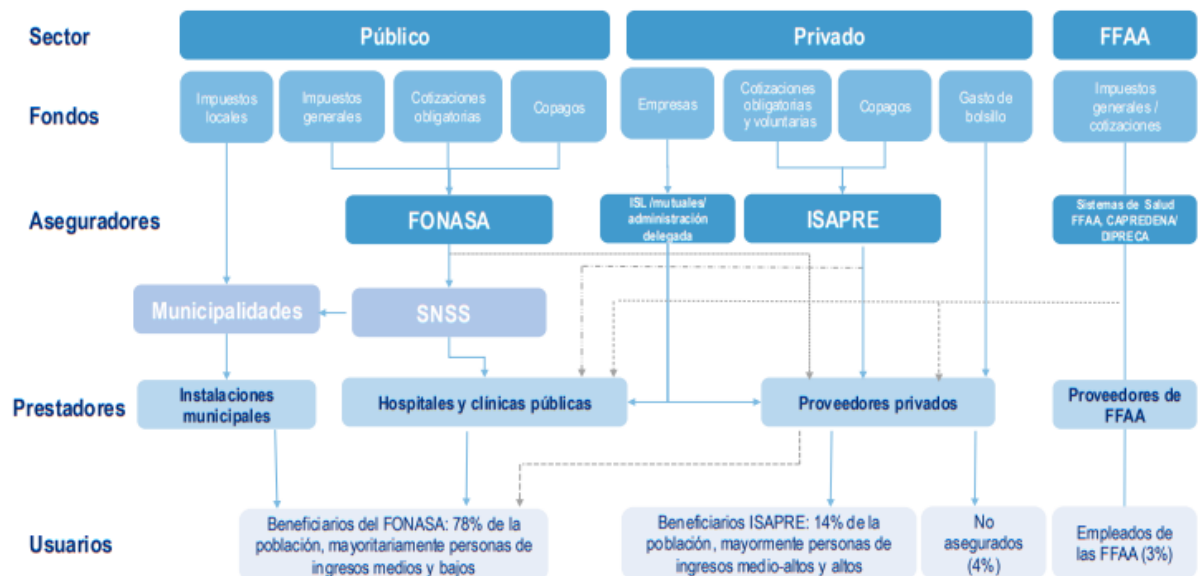
1. En primer lugar, debemos señalar que se trata de un bien y no de un servicio, razón por la cual se distingue del resto de los componentes de la función de producción de la salud.
2. Asimismo, las características del bien hacen que las relaciones de intercambio que se verifican en este mercado no se den en condiciones de igualdad de información. El mercado de medicamentos se caracteriza, entonces, porque se establecen relaciones entre los agentes donde, en la mayoría de los casos, impera la información asimétrica. Esta característica hace que se viole uno de los supuestos principales asumidos por el modelo competitivo, como es la información perfecta y accesible.
3. Por último, existe una multitud de agentes que, ocupando diferentes roles, participan en las distintas etapas del intercambio. El precio pagado por un

medicamento será la retribución al bien en sí mismo, el cual incorpora a los costos emanados del proceso de producción a los beneficios de los agentes encargados del mismo, así como a los servicios vinculados a la distribución y dispensa.

5.2 Gasto en salud en Chile

En Chile, el sistema de salud es un sistema mixto (público y privado), donde los ciudadanos de acuerdo con sus recursos económicos y motivaciones se adscriben a alguno de los subsistemas. Toda persona puede acceder a alguno de los seguros, pero la orientación del seguro público (Fonasa) es colectiva e independiente de los riesgos de salud que pueda presentar una persona, mientras que los seguros privados (Isapres) funcionan en base a planes que se tarifican en base a los riesgos o preexistencias (Sanhueza et al., 2019, p.41-42). En la Figura 5 se puede observar la estructura funcional del Sistema de Salud Pública elaborado por el autor.

Figura 5: Estructura funcional del Sistema de Salud Pública.



Fuente: Facultad de Medicina Clínica Alemana, Universidad del Desarrollo.

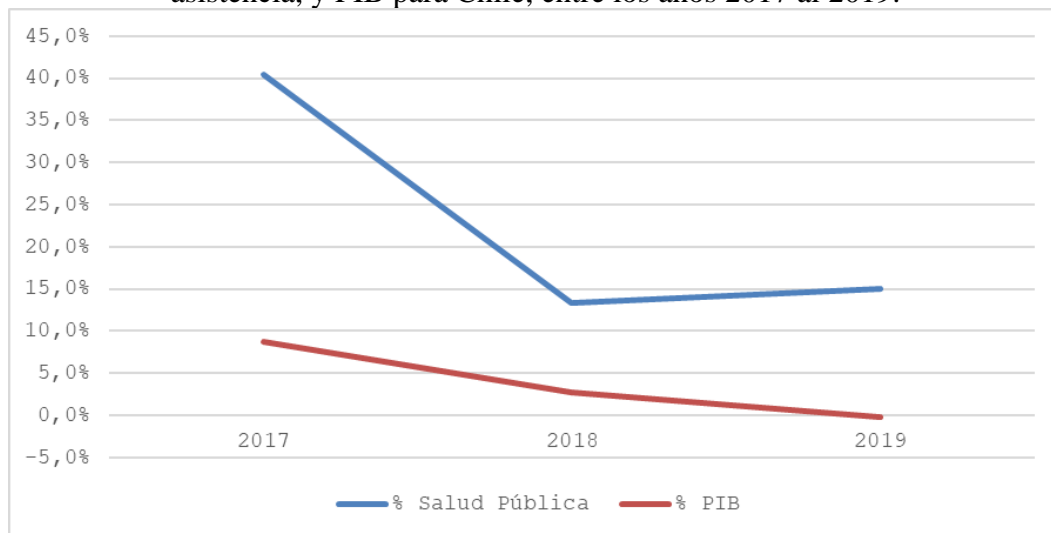
Con un condicionamiento adicional: el sistema de salud privado chileno se ubica entre los más caros del mundo. En consecuencia, se puede afirmar grosso modo, que el sistema de salud mixto comporta una dualidad: en lo público es defectivo y en lo privado inequitativo y encarecido. Esta estructura ha afectado el acceso a los medicamentos. Hay

diferencias notables en la distribución de medicamentos al usuario final, según se provean en el sector público o en el privado. El paciente que por diversas razones no accede a los medicamentos en el marco del Régimen General de Garantías de Salud en los establecimientos asistenciales del Sistema Público, debe procurarlos en el mercado (Alvear, 2013, p.132-133).

Según datos de Health Statistics (2018), el gasto en salud significa un 8,1% del PIB de Chile en el año 2017, aumentando en 1,3% desde el año 2010. Otro aspecto relevante, es que se comienza a disminuir la brecha entre el promedio de los países de la Organización para la Cooperación y el Desarrollo Económico (desde ahora OCDE), en un 0,8%. De lo anterior se puede concluir que las políticas públicas sobre salud son prioridad fiscal para Chile en los últimos años.

En la Figura 6, se puede observar la comparación de la variación anual del gasto público de salud y asistencia, y el PIB para Chile entre los años 2017 y 2019. Al analizar los datos, de gasto de salud y asistencia, entre los años 2017 al 2019, presentan una variación de 30,4%, en comparación con el PIB que presenta una variación del 2,5%.

Figura 6: Comparación de la variación anual entre el gasto de salud pública y asistencia, y PIB para Chile, entre los años 2017 al 2019.



Fuente: Elaboración propia basada en datos de Superintendencia de Salud y Banco Central de Chile.

Mizrahi (2012), concluyó lo siguiente:

En el caso de América Latina, dada la existencia de barreras regulatorias a la entrada de productos farmacéuticos y la homogeneidad de estos últimos, las prácticas anticompetitivas en el mercado farmacéutico se presentan tanto en el mercado mayorista y minorista del sector privado, como en las adquisiciones del sector público. La revisión de los casos atendidos por las autoridades de competencia de la región indica que son frecuentes los acuerdos entre laboratorios para proveer medicamentos e insumos médicos a los hospitales privados y públicos [...] en lo referente a las licitaciones públicas, se observa con frecuencia que los precios fijados responden a acuerdos entre competidores y también a la división artificial de los mercados en términos geográficos o por unidades médicas u otros. Lamentablemente, el diseño de las licitaciones con asiduidad permite la división de mercado y facilita así los acuerdos colusivos (p.10).

5.3 Valor público en el uso de datos en el sector público

Según Buenadicha, Galdon, Hermosilla, Loewe, y Pombo (2019):

Un número creciente de gobiernos está publicando datos en sus portales de internet con el objeto de mejorar su transparencia, pero también con la esperanza de que esa disponibilidad estimule el desarrollo de aplicaciones útiles para la ciudadanía, ya sea por parte de las entidades gubernamentales o de los propios ciudadanos [...] lo que aquí está en juego no es otra cosa que sentar las bases para un nuevo contrato social que permita una utilización masiva y responsable de los datos por parte de las entidades gubernamentales para proporcionar mejores servicios sociales, al tiempo que se mantiene la confianza de los ciudadanos en que los gobiernos gestionen sus datos de manera responsable con ciertos resultados (p. 8-9).

Otros autores Goldsmith, Crawford y Grohsgal (2016):

Al igual que lo vienen haciendo desde hace algún tiempo empresas de Internet como Google y Facebook, los gobiernos pueden ahora recurrir al análisis predictivo para orientar mejor sus actividades. La idea es sencilla: a partir de datos históricos sobre características de un tema determinado (por ejemplo, una red de transporte, un restaurante, niños obesos, edificios que podrían sufrir incendios), los estadísticos aíslan las características que parecen correlacionarse de forma clara. Los problemas que deben resolver las empresas y los gobiernos son similares y el análisis predictivo puede resultar de ayuda. Ciertas tareas que podrían parecer imposibles se tornan posibles (p. 6)

Si bien los procedimientos técnicos y de análisis son relevantes, se hará énfasis en las características, capacidades y potencialidades de las instituciones analizadas dado que son requisitos para el éxito de implementación de iniciativas de inteligencia de valor público dentro de los gobiernos. En términos generales, se reconocen las siguientes dimensiones para dicho éxito (Rodríguez, Palomino, y Mondaca, 2017, p.14):

1. La construcción de una institucionalidad que permita generar, administrar y dar continuidad a los recursos para la infraestructura y el personal dedicado a la inteligencia de valor público.
2. Lograr una comunicación transparente y fluida con otras entidades externas (públicas o privadas) y con la ciudadanía. En etapas tempranas de un proyecto de inteligencia de valor público, la capacidad de adaptar la información a las necesidades de los ciudadanos —y no de las entidades que los procesan— es fundamental para cimentar lazos con la ciudadanía.
3. La disponibilidad del capital humano necesario, ya sean profesionales que trabajen directamente en los análisis, tomadores de decisiones o actores claves que definan preguntas y objetivos. Por su parte, los científicos de datos tienen los conocimientos y competencias necesarios para mejorar la infraestructura y el ecosistema de los datos abiertos, así como también son capaces de proponer la analítica pertinente para su análisis.

El gobierno de Chile, a través del Gobierno Digital, tiene el objetivo de coordinar y asesorar intersectorialmente a los órganos de la administración del Estado en el uso estratégico de las tecnologías digitales, apoyando su uso, datos e información pública para mejorar la gestión y la entrega de servicios cercanos y de calidad a las personas. Donde se han creado diferentes soluciones tecnológicas como, por ejemplo, Clave Única, DocDigital, DatosGov, entre otros (Ministerio de la Secretaría General de la Presidencia, 2020).

Según la Agenda Digital 2020, el despegue del gobierno digital de Chile descansa sobre un desafío principal: alcanzar un Estado que responda a las demandas ciudadanas en forma oportuna y eficiente; en igualdad de condiciones, independiente de su ubicación geográfica (Barros, Campero, y Cabello, 2016).

Según el estudio para una gobernanza digital en Chile (2016), se identificaron 4 líneas de acción:

1. Masificar el uso de los trámites en línea del Estado en todo el territorio nacional, especialmente en los municipios, y mejorar la calidad de servicio.
2. Apoyar las políticas sectoriales para mejorar la oferta de servicios públicos digitales por parte de aquellas instituciones estatales más demandadas por la ciudadanía, comenzando por los ministerios de Salud, Desarrollo Social e Interior.
3. Lograr un Estado abierto en su información y proactivo en la disposición de servicios y aplicaciones útiles para la ciudadanía.
4. Mejorar la gestión pública para obtener resultados pertinentes con las necesidades ciudadanas, introduciendo la innovación como práctica efectiva y cotidiana.

Un aspecto identificado, comparando las líneas de acción del gobierno de Chile con las recomendaciones del BID, se comparten los mismo valores y ejes estratégicos de acción con foco en generar valor público para la ciudadanía, buscando, a través de herramientas tecnológicas, innovar y fortalecer la comunicación entre los ciudadanos y el gobierno.

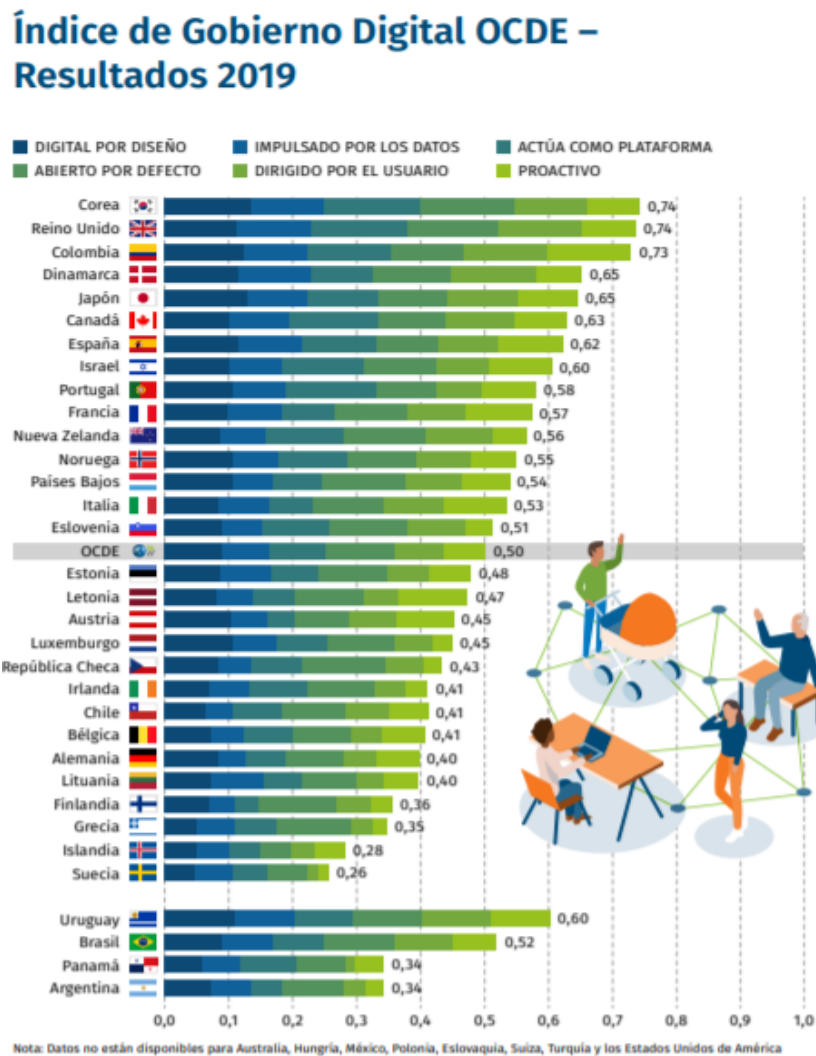
La OCDE en el año 2019 publica por primera vez el Índice de Gobierno Digital (Digital Government Index, DGI), el que tiene el objetivo de medir la transición desde el gobierno

electrónico al gobierno digital de acuerdo con la recomendación de la OCDE sobre estrategias de gobierno digital. Los resultados se pueden observar en la Figura 7.

En el caso de Chile, se encuentra 0,9 bajo el promedio de los países de la OCDE, y en cuarto lugar en los países de América Latina que se les realizó la medición.

Por lo anterior, en la presente tesis se busca generar valor público a través de la utilización de los datos en la toma de decisiones, con el fin de estimar los volúmenes operacionales de las líneas de negocio con algoritmos Machine Learning, que permitirán tener una herramienta alternativa para negociar mejores contratos de medicamentos para Cenabast.

Figura 7: Resultados del Índice de Gobierno Digital OCDE, para el año 2019



Fuente: OECD.

6. Metodología de investigación

Para el desarrollo de la metodología de investigación, se utiliza la metodología de estudio cuantitativa la metodología CRISP-DM, el tipo de estudio es predictivo donde se definen seis modelos de Machine Learning para el desarrollo de la tesis.

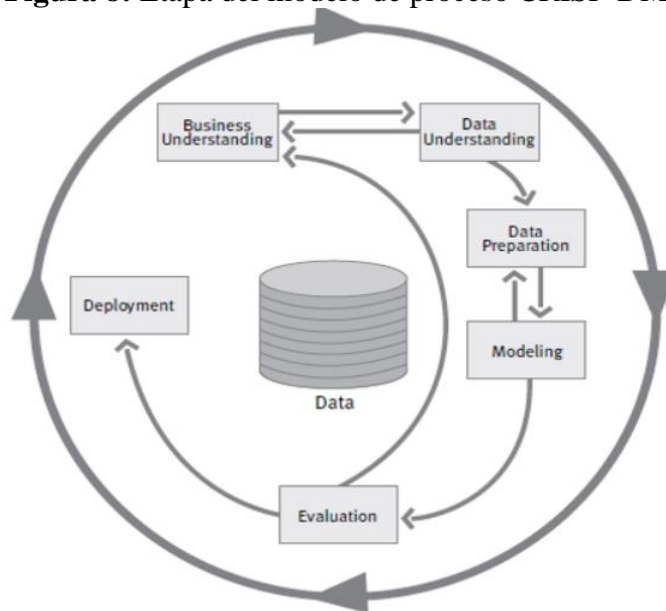
Las fuentes de datos de abastecimiento son entregadas por Cenabast y se determina el plan de análisis de la información, según los hallazgos de su revisión.

6.1 Metodología de estudio (cuantitativa)

Para el desarrollo de la tesis se utilizará la metodología CRISP-DM, es un modelo de proceso de minería de datos que describe una manera en la que los expertos en esta materia abordan el problema (Galán, 2015).

El modelo de proceso CRISP-DM, se puede estructurar en seis etapas identificadas en la Figura 8, donde muchas de las tareas se pueden realizar en un orden diferente y, a menudo, será necesario volver a hacer tareas anteriores repetidamente y repetir ciertas acciones (Chapman et al., 2000).

Figura 8: Etapa del modelo de proceso CRISP-DM.



Fuente: Chapman et al., 2000.

A continuación, se describen cada etapa del modelo definido con anterioridad y sus principales actividades.

6.1.2 Comprensión de los datos

La etapa de entendimiento de datos comienza con la recolección de datos inicial y continua con las actividades que le permiten familiarizarse primero con los datos, identificar los problemas de calidad de datos, descubrir los primeros conocimientos en los datos, y/o descubrir subconjuntos interesantes para formar hipótesis en cuanto a la información oculta.

Las principales actividades son:

- a) Recolectar datos iniciales
- b) Describir y explorar datos
- c) Verificar la calidad de los datos

6.1.3 Preparación de datos

La etapa de preparación de datos cubre todas las actividades necesarias para construir el conjunto de datos final (los datos que serán provistos en las herramientas de modelado de los datos en brutos iniciales). Las tareas de preparación de datos probablemente van a ser realizadas muchas veces y no en cualquier orden prescripto.

Las tareas incluyen la selección de tablas, registros, y atributos, así como la transformación y la limpieza de datos para las herramientas que modelan.

Las principales actividades son:

- a) Selección de datos
- b) Limpieza de datos
- c) Construir datos
- d) Integrar datos
- e) Formatear datos

6.1.4 Modelado

En esta etapa varias técnicas de modelado son seleccionadas y aplicadas, y sus parámetros son calibrados a valores óptimos. Típicamente hay varias técnicas para el mismo tipo de problema de minería de datos. Algunas técnicas tienen requerimientos específicos sobre la forma de datos. Por lo tanto, volver a la fase de preparación de datos es a menudo necesario.

Las principales actividades son:

- a) Seleccionar técnica de modelamiento
- b) Generación de prueba de diseño
- c) Construcción de modelo

6.1.5 Evaluación

En esta etapa del proyecto se ha construido un modelo (o modelos) que parece tener la alta calidad de una perspectiva de análisis de datos.

Antes del proceder al despliegue final del modelo, es importante evaluar a fondo ello y la revisión de los pasos ejecutados para crearlo, para comparar el modelo correctamente obtenido con los objetivos de negocio. Un objetivo clave es determinar si hay alguna cuestión importante de negocio que no ha sido suficientemente considerada. En el final de esta fase, una decisión en el uso de los resultados de minería de datos debería ser obtenida.

Las principales actividades son:

- a) Evaluación de resultados
- b) Conclusiones de resultados

6.2 Tipo de estudio (predictivo)

Para realizar la estimación de la demanda de los medicamentos comercializados por Cenabast, se utilizará una base de datos de información sobre los contratos vigentes con

los proveedores, donde se pueden obtener datos sobre cantidad, precio, fecha, tipo de línea de negocio, entre otros.

Los modelos que se utilizaran para predecir volúmenes operacionales de las líneas de negocio de Cenabast son:

- Lineal Regression
- Decision Tree Regression
- Random Forest Regression
- Regularization
 - Ridge Regression.
 - Lasso Regression.
- Elastic Net Regression

6.2.1 Lineal Regression

La regresión lineal, se puede identificar con la siguiente fórmula:

$$Y = mx + n$$

Donde,

Y: es la variable dependiente.
x: es la variable independiente.
m: es la pendiente de la recta.
n: es el intercepto con el eje y.

Según Newbold y Thorne (2008), la regresión lineal da dos importantes resultados:

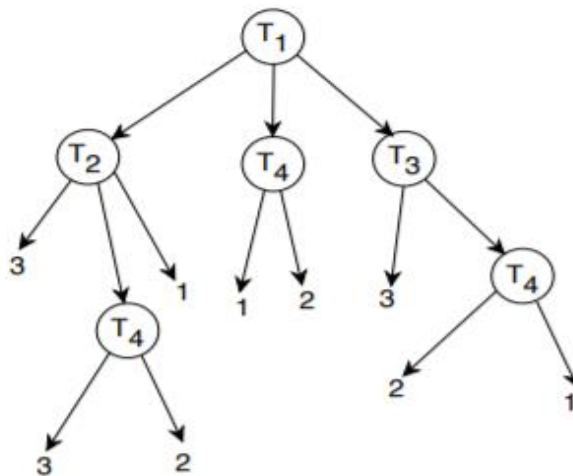
1. Los valores predichos de la variable dependiente o endógena en función de la variable independiente o exógena.
2. La variación marginal estimada de la variable endógena provocada por una variación unitaria de la variable independiente o exógena (Newbold et al., 2008, p.441).

6.2.2 Decision Tree Regression

“Un árbol de decisión (generalmente definido) es un árbol cuyos nodos internos son pruebas (en patrones de entrada) y cuyos nodos hoja son categorías (de patrones)” (Nilsson, 2005, p.73). Se debe utilizar los datos de entrenamiento para construir un modelo de generador de árbol que determinará qué variable dividir en un nodo y el valor de la división. La decisión de detener o dividir nuevamente asigna nodos hoja a una clase. Una ventaja de un árbol de decisión es que no hay necesidad de la creación exclusiva de variables Dummy (Swamynathan, 2017, p.176).

En la Figura 9 se puede observar un ejemplo de la construcción de un árbol de decisiones.

Figura 9: Ejemplo de árbol de decisiones.



Fuente: Swamynathan, 2017.

6.2.3 Random Forest Regression

Un subconjunto de observaciones y un subconjunto de variables se seleccionan al azar para construir múltiples modelos independientes basados en árboles de decisiones. Los árboles de decisiones están menos correlacionados ya que solo se usa un subconjunto de variables durante la división del árbol, en lugar de elegir con afección el mejor punto de división en la construcción del árbol (Swamynathan, 2017, p. 225).

6.2.4 Regularization

La regularización es una técnica para evitar el problema de sobreajuste. El sobreajuste ocurre cuando el modelo se ajusta demasiado bien a los datos. En este caso, podemos notar una alta precisión en el conjunto de datos de entrenamiento, mientras que el mismo modelo resultará en una baja precisión en el conjunto de datos de prueba.

Esto significa que el modelo ha ajustado la línea tan bien al conjunto de datos de entrenamiento que no pudo generalizarlo para que se ajuste bien al conjunto de datos de prueba. Con un aumento en la complejidad del modelo, el tamaño de los coeficientes aumenta exponencialmente, por lo que Ridge y Lasso Regression aplican una penalización a la magnitud del coeficiente para manejar el problema (Swamynathan, 2017, p.156).

Swamynathan (2017), identifica las siguientes modelos de sobreajuste:

Lasso Regression

También conocida como regularización L1. Orienta el valor del parámetro para que sea cero, es decir, los coeficientes de las variables que agregan valor menor al modelo serán cero, y agrega una penalización equivalente al valor absoluto de la magnitud de los coeficientes.

Ridge Regression

Conocida como regularización de Tikhonov (L2), guía los parámetros para que tiendan a cero, pero no de cero. Puede usar esto cuando tenga muchas variables que agreguen un valor menor a la precisión del modelo individualmente; sin embargo, mejora en general la precisión del modelo y no se puede excluir del modelo. Ridge regression otorga una penalización para reducir la magnitud del coeficiente de todas las variables que agregan un valor menor a la precisión del modelo, y que agrega una penalización equivalente al cuadrado de la magnitud de los coeficientes. Alfa es la fuerza de regularización y debe ser un valor positivo.

6.2.5 Elastic Net Regression

Es la combinación de Ridge y Lasso Regression busca combinar ambas penalizaciones. Cuando $\alpha=1$ se aplica Lasso Regression, y $\alpha=0$ se aplica Ridge Regression (Zou y Hastie, 2005).

6.3. Fuentes de datos a utilizar

Para el desarrollo de la tesis, los datos para aplicar los modelos de Machine Learning provienen de Cenabast entre los años 2010 al 2019.

La base de dato principal cuenta con 130.105 filas de datos por 50 columnas donde los principales campos se pueden agrupar en cinco grupos:

- Datos de documento: número de documento, fecha de documento, tipo de compra, mecanismo de compra, entre otros.
- Datos de medicamento: código de medicamento, nombre de medicamento, entre otros.
- Datos de proveedores: rut de proveedor, nombre de proveedor, entre otros
- Datos de compra: cantidad de compra, precio de compra, datos de facturación, entre otros.
- Seguimiento de compra: cantidad pedida, cantidad entregada, fecha de entrega, entre otros.

Las bases de datos complementarios permiten obtener datos sobre la evolución de los precios de los productos en el tiempo, identificar los contratos vigentes de Cenabast y maestro con los códigos internos de la institución.

Para el análisis de la base de datos, se presenta un análisis descriptivo y exploratorio de los datos entregados por Cenabast de forma agregada.

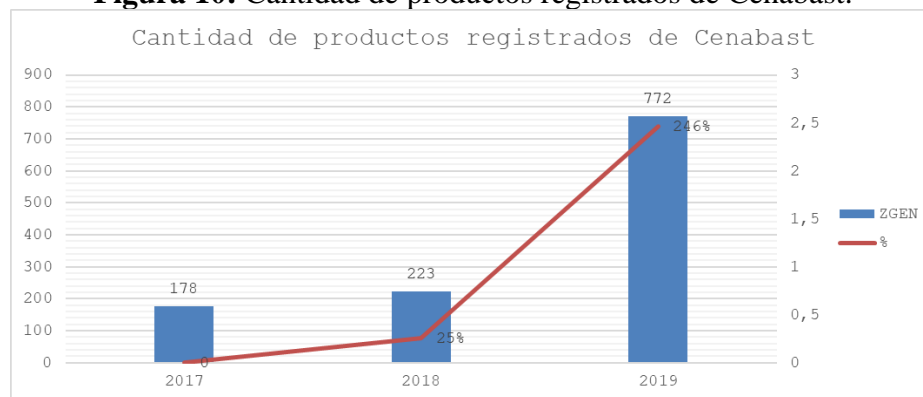
6.4 Plan de análisis

El plan de análisis se desarrolla analizando las bases de datos entregadas por Cenabast, a continuación.

6.4.1 Análisis de base de datos de productos

Cenabast, se abastece de productos que clasifica según un código interno, un texto breve sobre el producto y una de las siguientes clasificaciones fármacos, psicotrópicos, refrigerados y sueros. En la Figura 10 se puede observar la cantidad de productos registrados por año.

Figura 10: Cantidad de productos registrados de Cenabast.

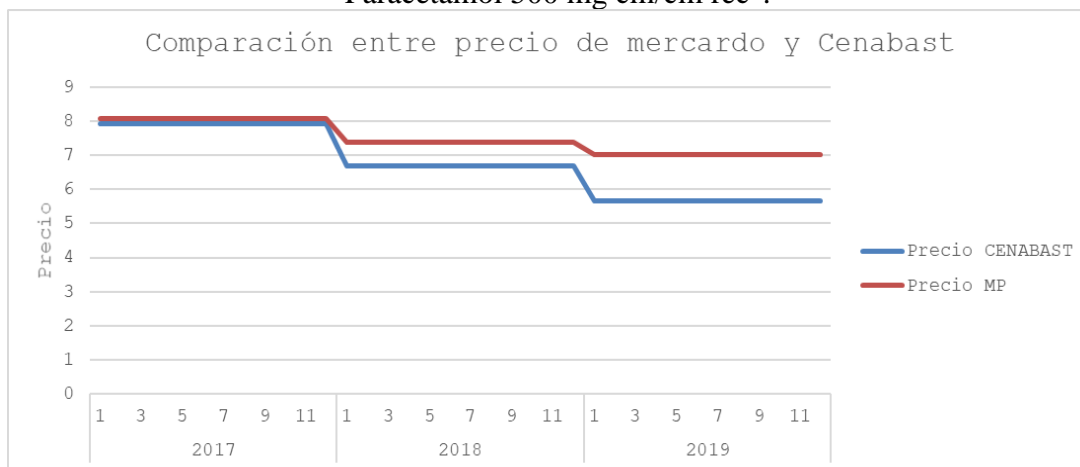


Fuente: Elaboración propia a partir de datos de Cenabast.

Desde el año 2017 al año 2020 se pueden identificar 784 productos en los registros de la institución, obteniendo un máximo de productos registrados durante el año 2019 y un alza del 246% comparado con el año anterior.

Otro aspecto relevante es que Cenabast realiza seguimiento al precio de mercado de los productos. En la Figura 11 se puede apreciar la comparación entre el precio de mercado y el precio Cenabast del producto “Paracetamol 500 mg cm/cm rec”.

Figura 11: Comparación entre precio de mercado y Cenabast del producto "Paracetamol 500 mg cm/cm rec".



Fuente: Elaboración propia a partir de datos de Cenabast.

La diferencia entre el precio de mercado y el precio Cenabast significa una diferencia del 9% para el año 2018 y un 19% para el año 2019.

6.4.2 Análisis de base de datos de abastecimiento

La base de datos de abastecimiento cuenta con datos desde el año 2010 al 2019, y entre las fechas se han emitido en total 13.156 documentos. En la Tabla 2 se puede identificar la cantidad de documentos emitidos por año y su variación anual.

Tabla 2: Cantidad de documentos emitidos por año y variación con respecto al año anterior.

Año	Cantidad de documentos	Variación anual
2010	927	-
2011	1.140	23,0%
2012	920	-19,3%
2013	821	-10,8%
2014	1.172	42,8%
2015	2.245	91,6%
2016	1.331	-40,7%
2017	1.378	3,5%
2018	1.510	9,6%
2019	1.712	13,4%
Total	13.156	-

Fuente: Elaboración propia a partir de datos de Cenabast.

De la Tabla 2 se puede observar que la máxima cantidad de documentos emitidos se encuentra en el año 2015 con 2.245 documentos y la menor cantidad en el año 2013, con 821 documentos. También se puede observar que la variación entre el año 2010 y 2019 es de 84,7%.

Los documentos se pueden categorizar en tipo de compra nacional o importación. En la Tabla 3 se puede observar la cantidad de documentos por tipo de compra y su porcentaje del total de documentos por año.

De la Tabla 3, se puede observar que el año 2019 se emitieron el máximo de documentos por tipo de compra importaciones, con 50 documentos, y para el tipo de compra nacional fue el año 2015 con 2.211 documentos. También se puede observar que los documentos emitidos por tipo de compra nacional son por sobre el 97% entre los años 2010 y 2019.

Tabla 3: Cantidad y porcentaje de documentos por tipo de compra.

Año	Importación	Nacional	% Importación	% Nacional
2010	0	927	0,0%	100,0%
2011	10	1.130	0,9%	99,1%
2012	9	911	1,0%	99,0%
2013	19	802	2,3%	97,7%
2014	22	1.150	1,9%	98,1%
2015	34	2.211	1,5%	98,5%
2016	16	1.315	1,2%	98,8%
2017	13	1.365	0,9%	99,1%
2018	17	1.493	1,1%	98,9%
2019	50	1.662	2,9%	97,1%
Total	190	12.966	-	-

Fuente: Elaboración propia a partir de datos de Cenabast.

Otra dimensión relevante para el análisis exploratorio son las ventas netas CLP por tipo de compra. En la Tabla 4 se puede observar el total de ventas netas CLP por tipo de compra.

Tabla 4: Ventas netas CLP por tipo de compra.

Año	Ventas netas CLP importación	Ventas netas CLP nacional	Ventas netas CLP total
2010	0	51.618.271.045	51.618.271.045
2011	321.296.786	212.565.801.534	212.887.098.320
2012	2.088.302.716	183.629.375.166	185.717.677.882
2013	2.602.057.414	262.411.392.194	265.013.449.608
2014	717.626.839	284.066.585.853	284.784.212.692
2015	17.527.938.567	368.007.318.294	385.535.256.861
2016	6.740.981.450	247.871.384.894	254.612.366.344
2017	8.216.333.664	442.112.150.475	450.328.484.139
2018	9.565.943.094	490.932.370.830	500.498.313.924
2019	18.429.565.548	616.518.505.801	634.948.071.349
Total	66.210.046.078	3.159.733.156.086	3.225.943.202.164

Fuente: Elaboración propia a partir de datos de Cenabast.

En la Tabla 4 se puede observar que las ventas netas CLP son mayores en el tipo de compra nacional en comparación a importaciones, y durante el año 2019 Cenabast presenta el máximo de ventas netas tanto en importaciones como nacional. También, entre el año 2010 al 2019, las ventas netas CLP totales han variado un 1.130,1%.

Cenabast compró los productos entre los años 2010 y 2019 a través de cuatro tipos de mecanismos de compra: convenio marco, licitación, trato directo y otros. En la Tabla 5 se puede identificar las ventas netas CLP para cada tipo de mecanismo de compra.

Tabla 5: Ventas netas CLP por mecanismo de compra y año.

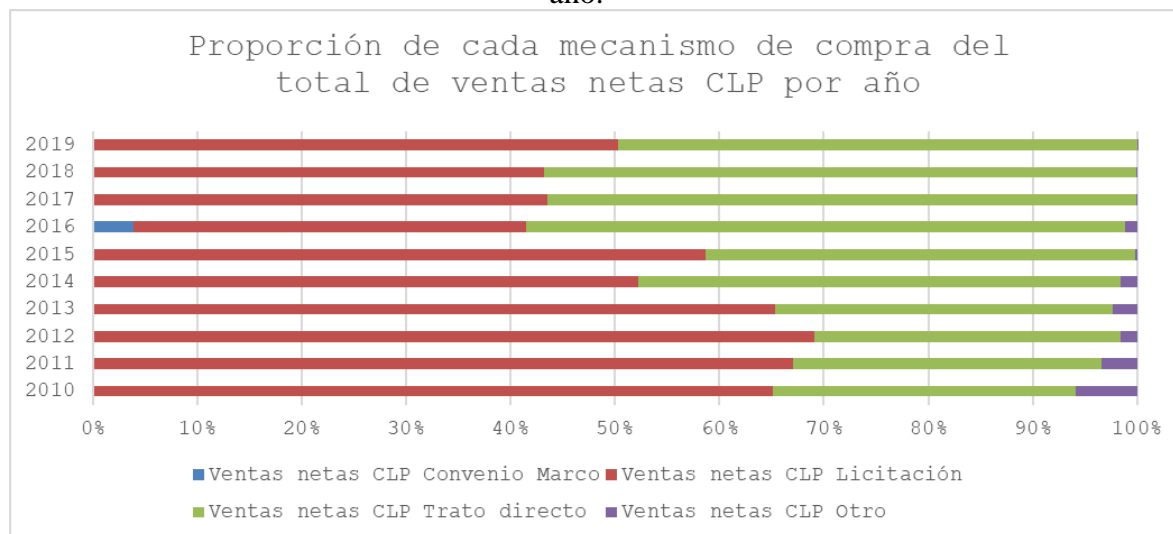
Año	Ventas netas CLP Convenio Marco	Ventas netas CLP Licitación	Ventas netas CLP Trato directo	Ventas netas CLP Otro	Ventas netas CLP total
2010	0	33.639.243.656	14.925.398.772	3.053.628.617	51.618.271.045
2011	0	142.685.135.409	63.045.316.855	7.156.646.055	212.887.098.319
2012	0	128.368.645.344	54.435.953.825	2.913.078.713	185.717.677.882
2013	0	173.191.626.868	85.713.227.450	6.108.595.290	265.013.449.608
2014	127.112.845	148.701.870.167	131.324.401.267	4.630.828.413	284.784.212.692
2015	109.534.881	226.139.692.500	158.758.984.608	527.044.872	385.535.256.861
2016	9.784.391.154	95.853.254.003	145.943.872.475	3.030.848.712	254.612.366.344
2017	102.155.625	196.021.777.086	253.849.755.381	354.796.047	450.328.484.139
2018	0	216.250.988.234	284.080.537.814	166.787.876	500.498.313.924
2019	0	319.187.325.996	315.664.628.496	96.116.857	634.948.071.349
Total	10.123.194.505	1.680.039.559.263	1.507.742.076.943	28.038.371.452	3.225.943.202.163

Fuente: Elaboración propia a partir de datos de Cenabast.

Se puede observar en la Tabla 5 que el mecanismo de compra convenio marco sólo se utilizó entre los años 2014 y 2017.

En la Figura 12 se puede observar la proporción de cada mecanismo de compra del total de ventas netas CLP para cada año.

Figura 12: Proporción de cada mecanismo de compra del total de ventas netas CLP, por año.



Fuente: Elaboración propia a partir de datos de Cenabast.

Los mecanismos de compra más utilizados por Cenabast son licitaciones y trato directo. También se puede observar en la Figura 12 el aumento del mecanismo de compra trato directo desde el año 2015.

La institución cuenta con tres líneas de negocios: intermediación, programas alimentarios y programas ministeriales. En la Tabla 6, se pueden identificar las ventas netas CLP para cada línea de negocio por año.

Tabla 6: Ventas netas CLP para cada línea de negocio por año.

Año	Ventas netas CLP Intermediación	Ventas netas CLP Programas Alimentarios	Ventas netas CLP Programas Ministeriales	Ventas netas CLP total
2010	22.774.189.577	21.674.040.708	7.170.040.760	51.618.271.045
2011	50.458.863.971	36.773.595.396	125.654.638.952	212.887.098.319
2012	54.971.833.716	49.583.104.306	81.162.739.860	185.717.677.882
2013	46.039.604.379	47.110.275.632	171.863.569.597	265.013.449.608
2014	67.231.574.090	68.650.838.016	148.901.800.585	284.784.212.691
2015	109.866.272.590	62.068.096.040	213.600.888.231	385.535.256.861
2016	94.305.054.216	7.698.937.512	152.608.374.616	254.612.366.344
2017	103.895.292.337	83.934.087.037	262.499.104.765	450.328.484.139
2018	156.335.773.143	39.138.380.422	305.024.160.358	500.498.313.923
2019	263.036.415.124	61.964.818.460	309.946.837.764	634.948.071.348
Total	968.914.873.143	478.596.173.529	1.778.432.155.488	3.225.943.202.160

Fuente: Elaboración propia a partir de datos de Cenabast.

En la Tabla 6 se puede observar que la principal línea de negocio de Cenabast es Programas Ministeriales, representando para el año 2019 el 48,8% del total de ventas netas CLP, seguido por la línea de negocio Intermediación con un 41,4% del total de ventas CLP para el año 2019.

Para el año 2019, el ranking de los productos más comprados por Cenabast por cantidad unitaria se pueden observar en la Tabla 7.

Tabla 7: Ranking de productos más comprados por Cenabast por cantidad unitaria, para el año 2019.

Producto	Cantidad Unitaria
PARACETAMOL 500 MG CAJ 1000 CM	401.002.000
ATORVASTATINA 20 MG CAJ 1000 CM REC	308.915.000
THROMBO A.A.S. 100 MG CAJ 100 CM REC ENT	290.000.000
LOSARTAN POTASICO 50 MG CAJ 1000 CM	245.191.000
ENALAPRIL 10 MG CAJ 1000 CM	244.254.000
OMEPRAZOL 20 MG CAJ 250 CP	185.753.250
JER INS 1ML 100UI 31G X 6MM CAJ 100 UN	178.929.100
EUTIROX 100 MCG CAJ 100 CM	135.408.900
HIDRORONOL 50 MG CAJ 1000 CM	134.067.000
NIDAL 1 800G X 12TA CAJ 9.600 G	122.995.200

Fuente: Elaboración propia a partir de datos de Cenabast.

En la Tabla 7, se puede observar que el producto más comprado por Cenabast por cantidad unitaria es el Paracetamol 500 MG, con 401.002.000 unidades y el segundo producto más comprado es la Atorvastatina con 308.915.000 unidades.

También se puede observar el ranking de los productos más comprados por Cenabast por Ventas netas CLP para el año 2019 en la Tabla 8, donde se puede observar que el producto “Hexaxim Vac Hexav Susp InY 1do Caj10Fam” es el producto más comprado por ventas netas CLP, con 26.208.000.000 para el año 2019.

Tabla 8: Ranking de productos más comprados por Cenabast por Ventas netas CLP para el año 2019.

Producto	Ventas Netas CLP
HEXAXIM VAC HEXAV SUSP INY 1DO CAJ10FAM	26.208.000.000
HERCEPTIN 600 MG/5ML SOL INY CAJ 1 FAM	13.122.000.000
INSULINA INSULATARD 100UI/ML 10ML FAM	12.593.667.576
TRIUMEQ 600/300/50 MG FRA 30 CM	11.998.800.000
VIMIZIM 5MG/5ML CAJ 1 FAM	11.596.101.100
SPRYCEL 100 MG FRA 30 CM	10.853.772.669
HEXAXIM VAC HEXAV SUSP INY 1DO CAJ10FAM	10.199.180.000
ORENCIA 125 MG SOL. INY. CAJ 4 JRP	8.332.194.000
BISINTEX 440 MG LIOF P/SOL INY CAJ 1 FAM	7.000.000.000
ELAPRASE 2MG/ML SOL. P/INFUSION 3ML FAM	6.664.269.950

Fuente: Elaboración propia a partir de datos de Cenabast.

6.4.3 Análisis de base de datos de contratos vigentes

La base de datos tiene un total de contratos vigentes de 12.138 documentos, entre los años 2010 y 2019. En la tabla 9 se pueden identificar la cantidad de documentos emitidos de los contratos vigentes de Cenabast por año.

De la Tabla 9, se puede observar que la cantidad de documentos vigentes entre los años 2010 y 2019 varían un 82,7%, donde la cantidad máxima de documentos emitidos fue el año 2019 y la menor cantidad el año 2013, con 1.679 y 793 respectivamente.

Tabla 9: Cantidad de documentos emitidos de los contratos vigentes de Cenabast, por año.

Año	Cantidad de documentos	Variación anual
2010	919	-
2011	1.019	10,9%
2012	859	-15,7%
2013	793	-7,7%
2014	1.157	45,9%
2015	1.603	38,5%
2016	1.319	-17,7%
2017	1.305	-1,1%
2018	1.485	13,8%
2019	1.679	13,1%
Total	12.138	-

Fuente: Elaboración propia a partir de datos de Cenabast.

Las ventas netas de los contratos vigentes de Cenabast entre los años 2010 y 2019, se pueden observar en la Tabla 10, donde se puede observar el aumento de las ventas netas entre los años 2010 y 2019, con una variación del 1.172,0%.

Tabla 10: Ventas netas CLP total por año de los contratos vigentes de Cenabast.

Año	Ventas netas CLP total	Variación anual
2010	49.658.136.185	-
2011	205.626.375.902	314,1%
2012	182.164.049.697	-11,4%
2013	262.139.053.582	43,9%
2014	283.618.123.622	8,2%
2015	380.325.102.213	34,1%
2016	251.638.892.582	-33,8%
2017	449.167.143.433	78,5%
2018	496.733.923.165	10,6%
2019	631.653.045.138	27,2%
Total	3.192.723.845.519	-

Fuente: Elaboración propia a partir de datos de Cenabast.

Otra dimensión para analizar son los proveedores de Cenabast. En la Tabla 11 se puede observar el ranking de los diez principales proveedores de la institución.

Tabla 11: Ranking de los principales proveedores de Cenabast por Ventas netas CLP para el año 2019.

Nombre proveedor	Ventas netas CLP total
SANOFI PASTEUR S.A.	36.423.430.000
SOC PROCESADORA DE LECHE DEL SUR SA	36.050.451.400
ROCHE CHILE LIMITADA	28.507.453.561
BRISTOL - MYERS SQUIBB DE CHILE	26.459.219.175
GLAXOSMITHKLINE CHILE FARMAC. LTDA	19.821.469.795
NOVARTIS CHILE S.A.	19.121.955.120
JOHNSON AND JOHNSON DE CHILE S.A.	18.426.916.432
NOVO NORDISK FARMACEUTICAL LTDA	16.902.152.522
GADOR LIMITADA	14.952.713.940
GRIFOLS CHILE S.A.	14.731.795.309

Fuente: Elaboración propia a partir de datos de Cenabast.

Se puede observar que los tres principales proveedores de Cenabast son: Sanofi Pasteur S.A., Soc. Procesadora de Leche del Sur S.A. y Roche Chile Limitada, con ventas netas CLP de \$36.423.430.000, \$36.050.451.400 y \$28.507.453.561 respectivamente.

6.4.4 Análisis de base de datos de código internos de Cenabast

En la base de datos de códigos internos de Cenabast se puede identificar los siguientes datos:

- Los clientes de cada Servicio de Salud.
- Vigencia de los productos de Cenabast.
- Identificación de los proveedores que comercializan cada producto que provee la institución.
- Los códigos de producto, cliente, sector, canal, entre otros.

Los datos entregados en la base de datos permiten realizar la preparación de los datos antes del modelamiento.

7. Preparación de datos para el modelamiento

Para la preparación de los datos para el modelamiento se utilizará la base de datos de abastecimiento, los criterios definidos son:

- Se agrupan los productos por año.
- Se consideran las compras realizadas entre los años 2010 y 2019.
- Se consideran las variables:
 - o Datos de producto.
 - o Cantidad por producto.
 - o Año por producto.

7.1 Análisis descriptivo de los productos de Cenabast

Se realiza el análisis descriptivo de los productos de Cenabast obteniendo un total de 12.674 datos representando 2.080 productos. En la Tabla 12 se puede identificar la cantidad, el promedio, varianza y coeficiente de variación de datos por año para 10 productos.

Tabla 12: Análisis descriptivo por producto de Cenabast.

Producto	Cantidad	Cantidad mean	Cantidad std	Cantidad cv
Producto 1	4	3.707.250.000	1.757.014.015	0,47
Producto 2	11	706.765.090.909	836.700.361.296	1,18
Producto 3	2	183.500.000	146.371.104	0,80
Producto 4	10	5.838.300.000	2.851.389.254	0,49
Producto 5	10	97.478.700.000	57.953.670.050	0,59
Producto 6	11	1.332.181.818	943.139.101	0,71
Producto 7	11	165.949.272.727	222.871.770.387	1,34
Producto 8	7	30.568.285.714	30.050.152.721	0,98
Producto 9	5	188.000.000	113.410.758	0,60
Producto 10	7	14.880.571.429	20.903.483.943	1,40

Fuente: Elaboración propia a partir de datos de Cenabast.

De la Tabla 12, se puede identificar que cada producto de Cenabast poseen sus propias características y propiedades. Por lo anterior, se decide realizar un proceso de categorización antes de realizar el modelamiento de los datos.

7.2 Clustering de los productos de Cenabast

Según Swamynathan (2017):

Los Clustering tienen el objetivo clave de identificar grupos distintos (denominados conglomerados) en función de alguna noción de similitud dentro de un conjunto de datos determinado [...] para realizar Clustering en los productos de Cenabast, se utilizará el algoritmo K-means, que tiene el objetivo de agrupar los datos en grupos con alta similitud entre los datos pertenecientes al grupo y baja similitud entre datos que no pertenecen al mismo grupo (p.195).

Al aplicar el algoritmo K-means, se logran identificar 4 Cluster, donde las características se pueden observar en la Tabla 13.

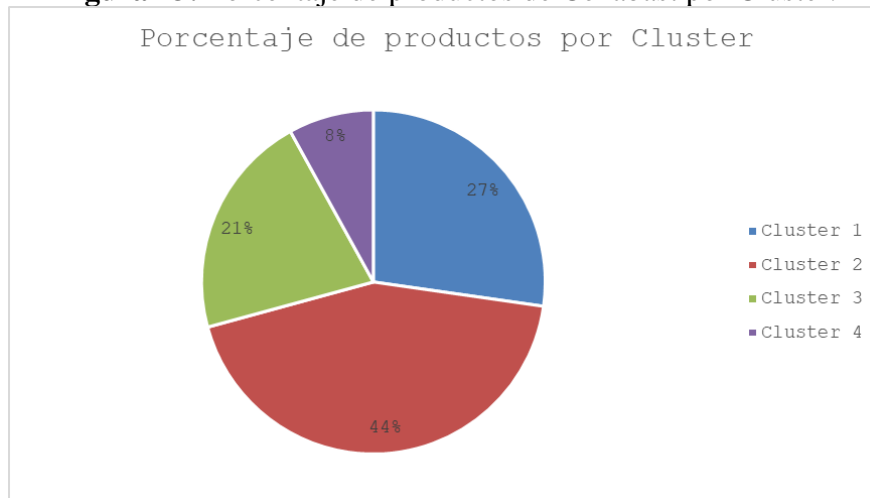
Tabla 13: Tabla de resultados de la caracterización de Clustering de los productos de Cenabast.

Cluster	Cantidad de productos	Cantidad count	Cantidad cv
Cluster 1	567	6.440.917	0,71
Cluster 2	905	2.782.320	0,74
Cluster 3	441	10.163.265	0,57
Cluster 4	167	9.167.665	1,53

Fuente: Elaboración propia a partir de datos de Cenabast.

En la Figura 13, se puede identificar el comportamiento de los Cluster del porcentaje de productos de Cenabast.

Figura 13: Porcentaje de productos de Cenabast por Cluster.



Fuente: Elaboración propia a partir de datos de Cenabast.

Para la aplicación de los modelos Machine Learning se utilizará el Cluster 3, porque presenta una gran cantidad de datos por producto y bajo coeficiente de variación, obteniendo un Cluster de datos más homogéneos en comparación a los otros. Por lo anterior, se realiza el modelamiento para el 21% de los productos, que se caracteriza por tener 441 productos, con un promedio de datos de 10,16 y un coeficiente de variación promedio de 0,57.

8. Modelamiento Machine Learning de datos

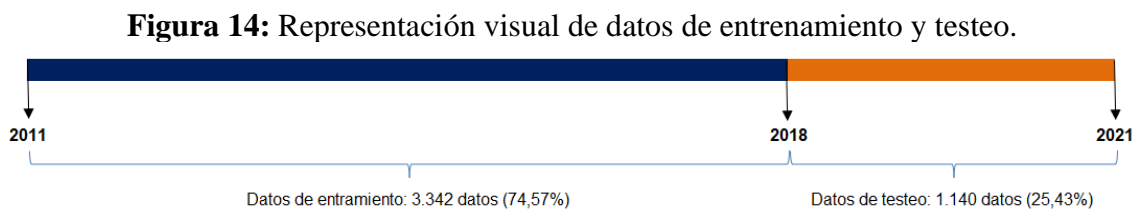
Para el modelamiento Machine Learning de los datos, se utilizan los modelos descritos en el capítulo 6.2, donde se definen seis modelos para estimar la demanda. La variable independiente es el año y producto, y la variable dependiente es la cantidad de producto.

En primera instancia, se utilizan los modelos con los parámetros dados por defecto, luego se utiliza el algoritmo Grind Search para obtener los mejores hiperparámetros dentro de cada modelo, dado una combinación de parámetros preestablecidos.

8.1 Datos de entrenamiento y testeo

Se considera como datos de entrenamiento a la información ubicada entre el año 2011 y 2018, equivalente a 3.342 datos (74,57%), mientras que los datos de testeo son los ubicados entre el 2018 y 2021, equivalente a 1.140 datos (25,43%).

En la Figura 14 se detalla la representación visual de los datos de modelamiento.



Fuente: Elaboración propia a partir de datos de Cenabast.

8.2 Resultados de los modelos Machine Learning

A continuación, se detallan los resultados de los modelos utilizando los parámetros por defecto e hiperparámetros de los modelos de Machine Learning.

8.2.1 Resultados de los modelos utilizando parámetros por defecto

A continuación, en la Tabla 14, se observan los resultados de los seis modelos con parámetros por defecto.

Tabla 14: Resultados de los modelos de Machine Learning.

Modelo	ECM	RECM	EAM	R²
Lineal Regression	21.560.773.972	146.836	65.971	72,76
Decision Tree Regression	28.438.521.470	168.637	35.962	64,07
Random Forest Regression	23.741.388.020	154.082	35.503	70,00
Ridge Regression	23.353.734.349	152.819	70.985	70,49
Lasso Regression	21.574.245.878	146.882	66.066	72,74
ElasticNet Regression	79.148.472.024	281.333	119.647	0,00

ECM: Error cuadrático medio.

RECM: Raíz del error cuadrático medio.

EAM: Error absoluto medio.

R²: Coeficiente de determinación.

(*) -0,002 al decimal más cercano

Fuente: Elaboración propia.

Analizando las métricas de desempeño, el mejor modelo es Lineal Regression al minimizar el error entre la cantidad real y predicho. También es el modelo que presenta un mayor coeficiente de determinación de los seis modelos de Machine Learning, con 72,76%.

El modelo que presenta menor desempeño al estimar la demanda de los productos de Cenabast es Elastic Net Regression, presentando un coeficiente de variación de -0.002%

8.2.2 Resultados de los modelos utilizando hiperparámetros

A continuación, en la Tabla 15 se pueden observar los resultados de los modelos utilizando hiperparámetros.

Tabla 15: Resultados de los modelos de Machine Learning utilizando hiperparámetros.

Modelo	ECM	RECM	EAM	R²
Lineal Regression	21.262.205.900	145.816	42.719	73,14
Decision Tree Regression	21.256.405.134	145.796	42.507	73,14
Random Forest Regression	21.556.302.233	146.821	42.515	72,76
Ridge Regression	21.597.539.387	146.961	66.054	72,71
Lasso Regression	21.566.025.935	146.854	65.980	72,75
ElasticNet Regression	21.566.761.372	146.856	65.980	72,75

ECM: Error cuadrático medio.

RECM: Raíz del error cuadrático medio.

EAM: Error absoluto medio.

R²: Coeficiente de determinación.

Fuente: Elaboración propia.

Analizando las métricas de desempeño utilizando hiperparámetros, el mejor modelo es Decision Tree Regression, al minimizar el error entre la cantidad real y predicho. También es el modelo que presenta un mayor coeficiente de determinación de los seis modelos de Machine Learning con 73,14%.

El modelo que presenta menor desempeño al estimar la demanda de los productos de Cenabast es Ridge Regression, presentando un coeficiente de variación de 72,71%

8.2.3 Comparación de resultados

A continuación, en la Tabla 16 se puede observar la comparación de los resultados utilizando los parámetros por defecto y lo hiperparámetros.

Tabla 16: Comparación porcentual entre los modelos con parámetros por defecto e hiperparámetros.

Modelo	ECM	RECM	EAM	R²
Lineal Regression	-1,38	-0,69	-35,25	0,52
Decision Tree Regression	-25,25	-13,54	18,20	14,16
Random Forest Regression	-9,20	-4,71	19,75	3,94
Ridge Regression	-7,52	-3,83	-6,95	3,15
Lasso Regression	-0,04	-0,02	-0,13	0,01
ElasticNet Regression	-73	-48	-45	3.637.450

ECM: Error cuadrático medio.

RECM: Raíz del error cuadrático medio.

EAM: Error absoluto medio.

R²: Coeficiente de determinación.

Fuente: Elaboración propia.

Al comparar los resultados utilizando parámetros por defecto e hiperparámetros, se puede identificar que en todos los modelos se logra mejorar las métricas de desempeño. El modelo Machine Learning que presenta mayor cambio es Elastic Net Regression y el modelo que presenta menor cambio es Lasso Regression.

9. Conclusión

La utilización de modelos de Machine Learning para la estimación de la demanda busca promover dentro del sector público la utilización de los datos para la toma de decisiones con la finalidad de aumentar el valor público que Cenabast entrega a la ciudadanía.

Para el desarrollo del trabajo se utilizó la metodología CRISP-DM, donde se pudo seleccionar el mejor modelo de Machine Learning para predecir la demanda, utilizando los modelos con parámetros por defecto e hiperparámetros.

En conclusión, este trabajo pretende contribuir en dar un primer paso para la utilización de herramientas de Machine Learning para estimar la demanda de los productos de Cenabast, con la finalidad de reducir la merma y obtener contratos más atractivos con los proveedores de los productos.

Se sugiere como extensión de este trabajo la utilización de información de inventarios, con la finalidad de poder realizar una estimación de la demanda de forma mensual.

10. Bibliografía

- Alvear, J. (2013). Protección de la salud, acceso a los medicamentos y deficiencias del mercado. Hacia un derecho social plurivalente en el marco de una economía social de mercado.
- Anuario Cenabast 2018. (2019).
- Barros, A., Campero, T., & Cabello, P. (2016). Estudio para la gobernanza digital en Chile. Chile.
- Buenadicha, C., Galdon, G., Hermosilla, M. P., Loewe, D., & Pombo, C. (2019). La gestión ética de los datos.
- CAECID, & CEPAL. (2017). Valor público, gobierno abierto y tecnologías digitales “Reflexión colectiva.” 1–5.
- Cenabast. (2020). Anuario Cenabast 2019. Chile.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0.
- Chilecompras. (2020). Datos abiertos. Retrieved from <http://datosabiertos.chilecompra.cl/>
- FNE. (2019). Estudio de Mercado sobre Medicamentos (EM03-2018).
- Galán, V. (2015). Aplicación de la metodología CRISP-DM a un proyecto de minería de datos en el entorno universitario.
- Goldsmith, S., Crawford, S., & Grohsgal, W. (2016). El análisis predictivo: impulsando la mejora a partir de los datos.
- Hopkins, E. (2006). Price Dispersion.
- Ministerio de la Secretaría General de la Presidencia. (2020). Gobierno Digital. Retrieved from <https://digital.gob.cl/>
- Ministerio de Salud. (2020). Ley N° 21.198. Retrieved from https://www.bcn.cl/historiadelaley/nc/historia-de-la-ley/vista-expandida/7713/#h2_5_1
- Mizrahi, E. (2012). Regulación y competencia en el mercado de medicamentos: experiencias relevantes para América Latina.
- Newbold, P., Carlson, W. L., & Thorne, B. (2008). Estadística para Administración y Economía (Ed. Sexta).

- Nilsson, N. J. (1998). Introduction to machine learning an early draft of a proposed textbook.
- Rodríguez, P., Palomino, N., & Mondaca, J. (2017). El uso de datos masivos y sus técnicas analíticas para el diseño e implementación de políticas públicas en Latinoamérica y el Caribe.
- Sanhueza, X., Laborde, C., Correa, T., Becerra, I., Bravo, R., Gomez, M., ... Santiago, M. (2019). Estructura y funcionamiento del sistema de salud Chileno.
- Seguel, A., & Zabala, X. (2019). Enfermedades de alto costo, crisis del costo/efectividad y biociudadanías emergentes en Chile.
- Swamynathan, M. (2017). Mastering machine learning with python in six steps.
- Tobar, F. (2008). Economía de los medicamentos genéricos en América Latina.
- Tobar, F., & Sánchez, D. (2005). El impacto de las políticas de medicamentos genéricos sobre el mercado de medicamentos en tres países del MERCOSUR.
- Vasallo, C. (2010). El mercado de medicamentos en Chile: caracterización y recomendaciones para la regulación económica.
- World Health Organization. (2018). World health statistics 2018: monitoring health for the SDGs, sustainable development goals.
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net.