

Universidad de Buenos Aires
Facultad de Ciencias Económicas
Escuela de Estudios de Posgrado

**CARRERA DE ESPECIALIZACIÓN EN MÉTODOS
CUANTITATIVOS PARA LA GESTIÓN Y ANÁLISIS DE
DATOS EN ORGANIZACIONES**

TRABAJO FINAL DE ESPECIALIZACIÓN

**VOLUMEN DE VENTAS DE MEDICAMENTOS Y
EL MANEJO DE STOCK EN LABORATORIOS**

El diseño de una predicción a través de modelos de
machine learning.

AUTOR: LORENA ZAMBRANO

DICIEMBRE 2019



Resumen

Para optimizar las estrategias de Volumen de ventas, se busca crear un modelo predictivo del stock de medicamentos a través de la prueba de diferentes algoritmos y aprovechando grandes volúmenes de datos relacionados con los que cuentan las cadenas de farmacias con sus laboratorios, para predecir mejores proyecciones que contribuyan en la toma de decisiones de la organización. Del mismo modo, a través de un modelo de clasificación se pretende anticipar cuáles medicamentos están con stock en el laboratorio farmacéutico de manera de poder actuar comercialmente, ya sea buscando alternativas para evitar que estos medicamentos puedan estar listos para las cadenas de farmacias. Esto abre a consideración las siguientes preguntas: ¿Cuáles son los métodos de aprendizaje automático que ayudan a predecir con mayor precisión el volumen de ventas? ¿Cuáles son los factores o variables que más influyen en el stock de medicamentos de un laboratorio?

Durante el desarrollo del presente trabajo se busca mostrar que los métodos de clasificación de aprendizaje automático permiten predecir el stock de ventas de un laboratorio en base a la planificación, control y gestión de inventarios del volumen de ventas, manejándolos adecuadamente con mayor precisión controlando procesos de ventas de cada medicamento



Estructura

INTRODUCCIÓN	5
APARTADO 1: ORGANIZACIONES FARMACÉUTICAS EN GRANDES VOLÚMENES DE DATOS	9
1.1 LA TECNOLOGÍA DE LA INFORMACIÓN Y COMUNICACIÓN DEL BIG DATA EN ORGANIZACIONES E INDUSTRIAS FARMACÉUTICAS	14
1.2 PLANIFICAR Y MANEJAR GRANDES VOLÚMENES DE DATOS	16
1.3 COMO LOS GRANDES VOLÚMENES DE DATOS CONTRIBUYEN PARA LA TOMA DE DECISIONES.	19
APARTADO 2: PROCESAMIENTO DE LA BASE DE DATOS UTILIZADA PARA LA TOMA DE DECISIONES.	22
2.1 PROCESO DE RECOPIACIÓN DE LOS DATOS	23
2.2 PROBLEMÁTICAS VINCULADAS CON LA BASE DE DATOS.....	24
2.3 PROCESO DE RECOPIACIÓN DE DATOS FINAL	25
2.4 BASE DE DATOS CONSOLIDADA PARA EL MODELO DE MACHINE LEARNING	27
APARTADO 3: MODELOS DE PREDICCIÓN A TRAVÉS DE MACHINE LEARNING	28
3.1 MODELOS DE PREDICCIÓN	28
3.2 MODELO DE REGRESIÓN LINEAL	29
3.3 ESPECIFICACIÓN DEL MÉTODO.....	30
3.4 MODELOS APLICADOS INDIVIDUALMENTE.....	31
CONCLUSIÓN.....	32
REFERENCIAS BIBLIOGRÁFICAS.....	33
APÉNDICES	34



Introducción

Los volúmenes de ventas, se pueden realizar por medio del manejo del stock de los medicamentos en las cadenas de farmacias representan para el inventario entre el 50% hasta el 70% de los activos de las empresas (Industrias Farmacéuticas), razón por lo cual en oposición de a lo anterior (es decir, la falta de saber mensualmente el volumen de ventas), los altos volúmenes de ventas ocasionan dificultades financieras por generar costes a su empresa por inmovilización de capital lo que afecta fundamentalmente el flujo de efectivo, se han registrado incluso casos recientes de despidos masivos de empresas distribuidoras por la ocasión que si no tenemos una predicción de nuestro volumen de ventas en cada una de las cadenas de farmacias donde el producto puede tener más accesibilidad podemos llegar a no manejar el stock de ventas de lo que podemos tener en el tiempo que se necesita.¹

La implementación de la predicción obtenida del Volumen de ventas en el stock será de gran alcance y utilidad para frenar el declive empresarial y dar impulso al crecimiento sostenido ya que la satisfacción de los clientes crea lealtad y se convierte en un arma estratégica para las organizaciones; el impacto que tendrá la propuesta se verá beneficiada en muchas direcciones y economías como en la Industria Farmacéutica, cadenas de farmacias, trabajadores de dichas empresas y en la sociedad misma ya que cada unidad de negocio contribuye con el producto interno bruto del país.²

El método de Machine Learning tiene grandes beneficios para esta predicción por un enfoque innovador, los algoritmos inteligentes y el análisis de volúmenes masivos de datos ofrecen valiosas herramientas para lograr el problema y objetivo a realizar, porque en el podemos encontrar crear sistemas capaces de aprender de forma automática, revisa los datos y es capaz de predecir comportamientos futuros, tiene ventajas principales entre ellas optimizar los

¹ Bender, A., & Mazza, N. (2017). Análisis predictivo: difusión e impacto en áreas de la sociedad. In *Simposio Argentino sobre Tecnología y Sociedad (STS)-JAIIO 46* (Córdoba, 2017).

² Reyes, G. V., Thompson, E. B., Vanoye, J. A. R., & Penna, A. F. (2018). MODELOS DE TECNOLOGIAS DEL BIG DATA ANALYTICS Y SU APLICACIÓN EN SALUD. *Pistas Educativas*, 39(128).



resultados de forma eficiente y a tiempo real, permitiendo optimizar las decisiones y predicción de resultados.³

La aplicación de estrategias para propiciar la toma de soluciones en la industria de la salud podría producir cientos de millones en valor anual en el sistema de salud, optimizando la innovación, mejorando la eficacia de la investigación y ensayos clínicos; así como diseñando nuevas herramientas a fin de que las industrias farmacéuticas, cumplan la promesa de un enfoque más personalizado. La factibilidad de poner en marcha el proyecto es muy viable ya que la Industria Farmacéutica cuenta con los recursos necesarios, para poder tener una predicción y poder mantener un nivel de stock óptimo por medio del volumen de ventas.⁴

La problemática que aqueja a muchas Industrias Farmacéuticas es que actualmente en la distribución del sector de consumo masivo registran disminuciones en sus utilidades como consecuencia de ventas pérdidas por falta de un stock de inventario de ventas, por el control de donde se puede generar reportes de los medicamentos para poder predecir el volumen de ventas, de acuerdo con los medicamentos que se registran en cada una de las cadenas de farmacias que se hace entrega para la comercialización de dichos medicamentos, la inexistencia de la predicción de los volúmenes de ventas y su incidencia en el stock de las cadenas farmacéuticas,⁵ desprendiéndose las implicaciones que tendría una mala toma de decisiones al momento de un inventario de ventas, la Automatización de los pedidos para el volumen de ventas se puede obtener las Estrategias de Comercialización y Ventas.

Los volúmenes de datos en su incidencia pueden manifestar amplias aplicaciones por medio de la ciencia de datos sea para ventas cruzadas, para los clientes (cadenas de farmacias), gestión de relaciones para analizar el comportamiento de las ventas, del stock de acuerdo con cada cadena de farmacia, con el fin de gestionar la deserción y maximizar el valor

³ Lúa-Lemus, H., & Estrada-Ramírez, M. (2019). Optimización de inventario de división de salud animal en comercializadora de ingredientes activos farmacéuticos.

⁴ Briones Bozza, D. M. (2019). Análisis de los factores que inciden en el comportamiento de compra de medicamentos genéricos desde la perspectiva del canal de distribución en el cantón Balzar.

⁵ Mogollón Espinoza, J. L. (2015). *Influencia en las ventas de la logística de inventario para dos distribuidoras farmacéutica del sector centro de la ciudad Guayaquil* (Master's thesis, Universidad de Guayaquil Facultad de Ciencias Administrativas)



esperado del mercado de medicamentos. En los volúmenes de ventas, stock en laboratorios de las Industrias Farmacéuticas se puede implementar diferentes maneras de tener mas eficiente y óptimo las estrategias de comercialización para implementar métodos de aprendizaje automático “Machine Learning” dejando paso para una implementación de la ciencia de datos que es donde están los principios fundamentales que apoyan y guían la extracción basada en principios de información y conocimiento de los datos donde se puede relacionar con Minería de Datos que muchas veces en las empresas u organizaciones no se ven por que no esta bien manejada la extracción real del conocimiento de los datos a través de tecnologías que incorporan estos principios que tienen manera a manejarse en todas las áreas funcionales del negocio para poder tener mejores predicciones.

Una problemática es tener datos, pero no poder manejarlos, ni darles la utilización que requiere al momento de tener una buena predicción, en la era del Big Data la interpretación de los volúmenes masivo de datos es fundamental para realizar cambios al predecir el número de bajas y altas de clientes, anticiparse a la demanda o a la morosidad o lanzar productos específicos ajustados a las necesidades reales de los usuarios, entre otras muchas aplicaciones.

El machine learning, los algoritmos inteligentes y el análisis de volúmenes masivos de datos ofrecen valiosas herramientas para las organizaciones, empresas, porque se trata de una disciplina científica del ámbito de la inteligencia artificial que crea sistemas capaces de aprender de forma automática.

Como objetivo General tenemos en Diseñar un Sistema Logístico de Planificación de volúmenes de ventas para aprovisionamiento que permita la precisión y manejo de stock de medicamentos de un laboratorio, en distintas cadenas de farmacias; sabiendo que para cumplir este objetivo general, debemos de cumplir los objetivos específicos que nos hemos planteado en nuestro trabajo que serán: Analizar cuáles son las características que mejor explican el volumen de ventas en el manejo de stock de medicamentos. Recopilar y preparar los datos para el armado del modelo. Definir un Plan por las ventas por cadena de farmacias para el manejo de stock de medicamentos, que se adecue a cada medicamento. Analizar las



Universidad de Buenos Aires
Facultad de Ciencias Económicas
Escuela de Estudios de Posgrado



características del modelo que predice con mayor AUC la rentabilidad por volumen de ventas. Desarrollar diferentes modelos con el método de machine learning.

Teniendo en cuenta estos objetivos planteados podemos lograr un buen resultado en nuestro trabajo propuesto.



Apartado 1: Organizaciones farmacéuticas en grandes volúmenes de datos

Pablo Iturralde en el informe “El negocio invisible de la salud” sostiene que el mercado de medicamentos tiene tal concentración de capital que puede ser considerado un oligopolio.⁶ Los fármacos en la sociedad entera son de vital importancia, las saludes de muchos en el mundo dependen de ellos. Las empresas cuya actividad se vinculen a ellos tanto como productores, distribuidores o consumidores tienen la obligación de administrarlos adecuadamente debido a que aportan enormemente al bienestar de los seres vivos, además de su propio bienestar como empresa.

En las empresas comerciales una de las partidas con mayor valor son los inventarios; puesto que son ellos los que permiten trabajar. Lo cual implica que debemos tener sumo cuidado sobre su tratamiento ya que puede surgir una infinidad de dificultades que los afecten. Como pueden ser obsolescencia, robo, caducidad, entre otros, llevando incluso a la quiebra de una empresa y generando un efecto negativo en la economía, por ejemplo, aumento en el desempleo, por su deficiente administración. Debido al auge de la industria farmacéutica surge la imperiosa necesidad de crear laboratorios nacionales y casas de representación de los laboratorios internacionales que fueron los encargados de surtir el stock de medicamentos comercializados en las farmacias.

“Casi toda empresa tiene un sistema contable, es decir, un medio para recolectar, sumariar y presentar, en signos monetarios, información relativa al negocio.” (Anthony, 1980)

En las Industrias Farmacéuticas los estados financieros, es precisamente el cumulo de información que tras varios procesos durante el ciclo contable se convierte en información financiera que determina en ella el estado situacional de la empresa, sus productos, sus ventas, su situación en distribución de los fármacos, si tienen stock de ventas, estos informes

⁶ Campoverde Velasco, A. E., & Panjón Barreto, E. C. (2016). *Análisis financiero y administrativo a la partida inventarios de la distribuidora de productos farmacéuticos y medicinales Silvio Abad (DISA) de la ciudad de Cuenca, durante los periodos 2013 y 2014* (Bachelor's thesis).



que se detallan van a permitir cubrir las necesidades de usuarios y de la empresa dicho sea de paso, es resumida y plasmada en los estados financieros que son instrumentos que proporcionan información para ayudar a la industria farmacéutica a determinar la situación de la empresa en la fecha de su elaboración, es aquí donde con predicciones de volúmenes de ventas se puede obtener un informe en base al stock de las cadenas de farmacias y como se puede llevar de una manera mas óptima eficiente y eficaz la distribución de los medicamentos a cada ciudad, donde se comercializan.⁷

El objetivo es presentar información útil sobre el rendimiento financiero y las ventas en cada una de sus cadenas de distribución para la toma de decisiones de una diversa gama de usuarios; además de reflejar los resultados de la gestión realizada por la administración. Es por eso por lo que se amplía la información aduciendo que el juego completo de estados financieros ayuda en la predicción del volumen de ventas, su distribución temporal y su grado de certidumbre, para el manejo de stock del laboratorio. En una empresa en donde el rubro de inventarios constituye la mayoría de sus activos genera la necesidad de tener un control sobre los mismos, el poder mantener un nivel equilibrado de stock el mismo que cubra los niveles de ventas y que no genere costos adicionales al poseer niveles altos.

Es así que, según Alberto Ortiz, en su libro *Gerencia Financiera*, un enfoque estratégico plantea que: “La política financiera de inventarios está encaminada al alcance del equilibrio, mediante la evaluación de los siguientes temas: establecimiento de la política, prácticas de compras, niveles de inventarios, decisiones financieras, instrumentos matemáticos empleados para optimizar el capital invertido en inventarios y objetivos de la política”.⁸ La política de inventarios nos sirve de pauta para poder efectuar el control requerido en los inventarios, tomando en cuenta lo volúmenes de ventas y compras, los costos implícitos en

⁷ Montes Chino, A. A. (2005). *La administración del departamento de crédito y cobranzas en empresas farmacéuticas* (Master's thesis, Universidad Iberoamericana Ciudad de México. Departamento de Administración y Contaduría Pública).

⁸ Alberto, O. G. (1997). *GERENCIA FINANCIERA: UN ENFOQUE ESTRATEGICO*.



los inventarios y los niveles de rentabilidad requeridos por la empresa, de tal forma que se debe considerar al momento de plantear una política dentro de una empresa.

“Alcanzar el equilibrio depende de la estimación que se efectuó de todos los costos derivados de mantener un nivel determinado de inventarios y de tener una eficiencia y eficacia en su control”.⁹

Las empresas tratan de optimizar sus recursos, por un lado, necesitan abastecer (comprar) para mantener un stock que cumpla satisfactoriamente con sus clientes, sin embargo, esas adquisiciones deben ser eficientes, pues, debemos mantener un surtido adecuado y al menor costo. Mientras que las ventas deben ser cuidadas para conseguir cantidad suficiente para obtener beneficios. Para esto diseñar una predicción del volumen de ventas en cada una de las distintas cadenas de farmacias, con ello podemos determinar el punto de equilibrio, cuya finalidad es conocer el volumen mínimo de ventas que nos permitirá cubrir al menos los costos y gastos. Lo cual no será útil para comparar con las ventas actuales y conocer la situación. De esta manera crear estrategias que permitan mejorar la rentabilidad; para consiguiente necesitaremos una lista de productos con el número de unidades vendidas en un período tal, su costo y precio de venta.

La administración de los grandes volúmenes de ventas es un punto determinante en el manejo estratégico de toda industria farmacéutica para el manejo de stock, por lo tanto, constituye también un punto de discrepancia entre los administradores por determinar cual es la manera más eficiente de gestionarlos, por tal motivo se debe recalcar que esta no es una decisión que se puede tomar a la ligera, por lo contrario, se necesita de un análisis previo que constituya una guía sobre el manejo más apropiado según el tipo de negocio.

Uno de los indicadores que nos ayudan a evaluar los volúmenes de ventas sin duda alguna es el de rotación, toda empresa busca que sus inventarios salgan de sus bodegas de la manera más rápida que el mercado lo permita, es decir buscan que sus activos sean líquidos a la brevedad posible, y al realizar un estudio sobre la rotación de inventarios se puede obtener

⁹ Campoverde Velasco, A. E., & Panjón Barreto, E. C. (2016). *Análisis financiero y administrativo a la partida inventarios de la distribuidora de productos farmacéuticos y medicinales Silvio Abad (DISA) de la ciudad de Cuenca, durante los períodos 2013 y 2014* (Bachelor's thesis).



una visión general de como se encuentran los mismo dentro de la empresa, si existe el stock necesario o por el contrario se ha generado un estancamiento, resultados que al final afectan a la utilidad del negocio.

Para el Economista Humberto Moscoso en su libro “Investigación Operativa” Tomo II, los motivos que llevan a una empresa a la creación de un stock son primordialmente, el motivo de transacción, dentro de inventarios existe una entrada (compras) y salida (ventas-uso) de bienes de manera constante, motivo que genera la necesidad de contar con un stock necesario que ayude a cubrir la demanda en el momento requerido, en caso de empresas manufactureras un stock de materia prima que permita avanzar con el proceso productivo sin mayor inconvenientes, un inventario que se encuentre siempre disponible de acuerdo al monto de producción. En caso de una Industria Farmacéutica comercializadora, el stock de inventarios debe estar siempre acorde al nivel de ventas manejado por la Compañía.

En una empresa en donde el rubro de inventarios constituye la mayoría de sus activos genera la necesidad de tener un control sobre los mismos, el poder mantener un nivel equilibrado de stock el mismo que cubra los niveles de ventas y que no genere costos adicionales al poseer niveles altos, lo podemos determinar mediante la predicción de volumen de ventas, en cada una de las cadenas de farmacias. El tener datos, pero no poder manejarlos, ni darles la utilidad necesaria es donde las empresas pueden encontrar problemas que se pueden llevar a pérdidas grandes, implementar el diseño de Big data la interpretación de los volúmenes masivos de datos es un favor fundamental para realizar cambios al predecir en tiempo para la industria.

Es así que, según Alberto Ortiz, en su libro Gerencia Financiera, un enfoque estratégico plantea que: “La política financiera de Volúmenes de ventas está encaminada al alcance del equilibrio, mediante la evaluación de los siguientes temas: establecimiento de la política, prácticas de compras, niveles de inventarios, decisiones financieras, instrumentos matemáticos empleados para optimizar el capital invertido en inventarios de ventas y stock en objetivos de la política”.¹⁰

¹⁰ López Parra, M. F. (1997). *Contrato de gestión: viabilidad en la administración pública del Ecuador: un estudio estratégico para su implementación* (Doctoral dissertation).



Con todas las definiciones financieras se puede entender que se encuentran datos, tablas, pero deben de ser manejados y llevados a ser utilizados de una manera oportuna para la empresa, entonces necesitamos herramientas que nos permitan identificar los problemas y poderlos corregirlos a tiempo. La predicción mediante el método de Machine Learning es una herramienta muy útil en este proceso, que nos llevará a buscar y plantear el diseño que buscamos con sistemas capaces de aprender de manera automática, con volúmenes de datos que nos permitan tener una mejor predicción.

Contar con un modelo a entrenar con conjuntos de datos antes de ser implementados, requiere que el correcto conjunto de datos que se aplique a un proceso de aprendizaje. Una organización no necesita de Big data para utilizar las técnicas del machine learning; sin embargo, el Big data puede ayudar a mejorar la precisión de los modelos de machine learning. Con el Big data, ahora es posible tener los datos de manera virtual para que se puedan almacenar de manera más eficiente y rentable, ya sea ON PREMISES o en la nube. Además, las mejoras en la velocidad y confiabilidad en la red han eliminado otras limitaciones físicas asociadas con la gestión de cantidades masivas de datos a una velocidad aceptable. Agregue a esto el impacto de los cambios en el precio y en la satisfacción de la memoria de las computadoras y ahora es posible imaginar cómo las compañías pueden aprovechar los datos de formas que serían inconcebibles hasta sólo cinco años atrás.

Las predicciones son una herramienta indispensable para asegurar un buen desenvolvimiento de la gestión de la demanda y estas se ven afectadas cuando no tenemos un stock de venta en el laboratorio, los líderes de negocios están empezando a descubrir que muchas cosas que están sucediendo dentro de sus organizaciones e industrias no pueden ser entendidas a través de una consulta. No son las preguntas que sabe, pero son los patrones ocultos y las anomalías escondidas en los datos lo que pueden ayudarlo o dañarlo.

Por otro lado, la utilización de un sistema logístico de planificación de Volúmenes de Ventas en inventarios para el stock les permitiría a las empresas de este sector los siguientes beneficios: Aumentar los niveles de venta y a la vez aumentar la satisfacción y disminuir las



pérdidas de sus clientes. (cadenas de farmacias), Stock para poder reducir los costes de tener productos (medicamentos), Cuáles son los costes que conseguimos reducir mediante la venta de stock, Aumento del flujo de efectivo, disminuyendo los costos financieros de los inventarios, Reducir el costo de almacenamiento, No correr el riesgo si no vendemos los stocks que tenemos almacenados para que no queden obsoletos en el mercado, Alinear las operaciones de aprovisionamiento y distribución con los objetivos estratégicos del negocio.

En Línea con esto, las organizaciones cuentan cada vez más con grandes volúmenes de datos o con la capacidad de generarlos y recolectarlos, de gran variedad y que se generan a mucha velocidad. A través del procesamiento de estos datos mediante algoritmos, las empresas buscan generar información como factores de decisión para optimizar procesos, mejorar resultados. Los laboratorios farmacéuticos no se encuentran aislados de esta situación y existe un gran potencial. A partir del uso de Big Data, para producir información que sea pertinente, de calidad y oportuna para fundamentar y orientar decisiones, o para diagnosticar problemas que se conocían y entonces era imposible accionar.

1.1 La tecnología de la información y comunicación del big data en organizaciones e industrias farmacéuticas

La aplicación de la tecnología en la industria farmacéutica se canalizará a través de dos herramientas clave en 2019: 'big data' e inteligencia artificial. Así lo han puesto de manifiesto los participantes de The State of the Biopharmaceutical Industry – 2019, de la compañía de estudios y análisis GlobalData.

Los laboratorios farmacéuticos están compitiendo por obtener datos sobre la salud de los pacientes y alcanzar acuerdos con empresas de tecnología, ya que el análisis del big data empieza a desentrañar un tesoro de información sobre el efecto de las medicinas en el mundo real.

Estudiar este tipo de datos en el mundo real ofrece a los fabricantes una poderosa herramienta para demostrar el valor de sus fármacos, algo que se puede obtener por medio de las compras que se realicen a la industria farmacéutica. Los datos del mundo real implican recolectar



datos fuera de los tradicionales ensayos clínicos aleatorios, el actual estándar de oro para juzgar los medicamentos, y el interés en este campo está creciendo de forma exponencial; es por eso que las principales compañías farmacéuticas del mundo ahora tienen departamentos enfocados en el uso de datos del mundo real en múltiples ventas, volumen de ventas, entrada de medicina, y registro de enfermedades que se estean viendo de manera más repetitiva (diabetes, cardiología, hipertensión) y varios han completado estudios científicos utilizando la información para profundizar en áreas clave a las que dirigen sus medicamentos.

“Cada vez es más costoso realizar una investigación tradicional de ensayos clínicos, por lo que la industria está buscando maneras de lograr objetivos similares utilizando datos recopilados de forma rutinaria”, dijo Paul Taylor, un experto en informática de la salud del University College de Londres.

El big data será la tecnología ocn mayor impacto en las industrias Farmacéutica en este año en curso (2019), así se verá un mayor empleo del uso de datos en la investigación y diseño y volumen de medicamentos, donde servirá para monitorizar a los sujetos de las pruebas en tiempo real. Su empleo ayudará a crear registros electrónicos, que permitan detectar tendencias y, en última instancia, abaratar el coste de los medicamentos, en tiempo, stock para una mejor calidad. Se creará un nuevo tipo de inteligencia, asociada a la salud, que ayudará a revolucionar las estadísticas de usuarios, y seguimeintos de los medicamnetos.

En los sistemas de salud, centros médicos y empresas farmacéuticas, el uso de 'big data' ayudará a predecir resultados, diseñar protocolos de tratamiento, y poder predecir los volumenenes de medicamentos que pueden requerir para tratamientos que tienen medicinas que no son tan amplias la manera de conseguir y tenerlas en las farmacéuticas.

Una tecnología de impacto podemos denominarla a la inteligencia artificial en la industria para el próximo año; donde se puede predecir que su uso en salud crecerá rápidamente la gran cantidad de datos del paciente que se pueden extraer de sus registros, para una mayor utilizacieon de dichos registros que nos ayudan a determinar de manera oprotuna el coste, producción de medicamentos de acuerdo a necesidades que persisten de manera mas



recurrente. Las empresas farmacéuticas deben de implementar como una inversión de trabajo, tiempo, y coste de producción; estas nuevas fases de inteligencia artificial que ayudarán a tener una mejor adopción de sistemas para la eficiencia y eficacia del trabajo.

A la hora de adoptar este tipo de mecanismos hay una evolución respecto al modelo de trabajo clásico. Donde antes se optaba por fusiones y adquisiciones para integrar nuevas perspectivas o enfoques de trabajo, ahora se plantean las alianzas y asociaciones como la mejor opción. Para que las farmacéuticas puedan capitalizar todo el potencial de estas dos herramientas, la responsable global de esta industria en GlobalData, Bonnie Bain, considera que "las compañías necesitan comprender los beneficios específicos" que pueden aportar. "Invertir en IA y 'big data' no otorgará una cura para los problemas de negocios, así que no se debe creer el 'hype' y sí plantear expectativas realistas de éxito". Hablando en concreto sobre inteligencia artificial, Bain explica que "hay también un número de soluciones IA en el mercado y no todas son adecuadas para cada objetivo de negocio. Elegir la solución correcta para cada necesidad es clave", incide.

1.2 Planificar y manejar grandes volúmenes de datos

La industria farmacéutica siempre ha tenido que ir de la mano de la tecnología no solo por los avances que existe en la industria propiamente dicha, sino también por los avances que se presentan en el mundo actual en lo que refiere a tecnología moderna, la cual, sirve como apoyo para la industria. De allí, la importancia de aprovechar la Big Data farmacéutica como parte de esa esencia de la industria de fármacos que necesita de las comunicaciones o de la información para poder adaptarse a las necesidades industriales de hoy en día, pero más aún, para aprovechar al máximo la tecnología aplicada a la fabricación de los medicamentos.

La Big data farmacéutica se debe aplicar en todos los campos de la investigación de los fármacos, siendo de especial de ayuda en el campo clínico. Si bien, es importante conocer los efectos del fármaco directamente desde los pacientes, para una mejor elaboración del medicamento y poder ser entregado no dejando medicamentos sin consumir ni que pueda llegar hacer falta, hacer estos ensayos controlados y es a los que se le denomina los ensayos



clínicos, en el cual, se debe llevar a cabo un seguimiento del ensayo. Ofreciéndose en la actualidad, la Big Data farmacéutica como medio de almacenamiento de datos y de información relevante del fármaco en los ensayos clínicos.

Con la Big data no solo se puede manejar un ensayo aisladamente, sino que, se pueden evaluar y considerar diferentes ensayos clínicos que ocurran de forma global en otros laboratorios asociados en cualquier otro lugar del mundo. De hecho, la información se puede intercambiar y servir de valioso apoyo para los grupos farmacéuticos aliados que se encuentren realizando los estudios sobre el uso de algún fármaco determinado.

Así mismo, la Big Data farmacéutica es necesaria para la actualización de los sistemas regulatorios de la industria de los laboratorios farmacéuticos, ya que, con la información compilada sobre los avances y los resultados, se puede restablecer las acciones regulatorias internas de los medicamentos, como también, se puede solicitar la actualización de los sistemas de control externos en salida y entrada de medicamentos, producción de fármacos, volumen de ventas, y entrega de medicamentos en diferentes partes de distribución, donde se desarrolla la actividad industrial farmacéutica.

Siempre la información que se tiene es de manera valiosa, así que el buen uso de la Big Data farmacéutica conformada por los estudios de los laboratorios y las propias informaciones compiladas por cada una de los departamentos que se encargan de ver información de Hospitales de acuerdo de pacientes, de los usuarios que acceden a los medicamentos, será una compilación valiosa que debe ser valorada y considerada, pero lo más importante, debe ser bien administrada para ser usada de forma adecuada.

Los datos son el alma de la industria y en las industrias de la ciencia, como lo es la industria farmacéutica, es aún mucho más. Los datos siempre y cuando sean confiables, permiten una comprensión completa de las relaciones entre los proveedores de atención médica (farmacias), médicos y las redes integradas de entrega de productos farmacológicos que son cruciales para lograr la eficacia de las ventas y satisfacer las necesidades de los pacientes. La organización y análisis de datos es fundamental para que este tipo de industria pueda comercializarse con mayor facilidad.



A su vez, con la globalización, las organizaciones de ciencias de la vida buscan formas de establecer procesos de gobernanza de los datos que les ayuden a administrar la información de forma corporativa.

Hoy en día, las compañías farmacéuticas emplean decenas de miles de representantes de ventas, repartidos por todo el mundo, quienes son elementos de búsqueda de información ya que recolectan las percepciones de distintos pacientes en diferentes lugares. En el entorno comercial de hoy, la velocidad del mercado es extremadamente importante. Las organizaciones de ciencias biológicas deben asegurarse de que los representantes de ventas en todo el mundo tengan acceso a datos confiables y que conozcan como ofrecer los productos farmacéuticos para que generen el impacto debido y logren la efectividad deseada.

Por otra parte, es necesario planificar la coherencia de los datos y la personalización como la comprensión de cada mercado en cada sector. Seguidamente será necesario conectar todas las fuentes de datos internas, externas y de terceros y combinar los datos en tiempo real para crear una sola fuente de datos fiables.

Así mismo, los sistemas modernos de gestión de datos basados en la nube le permiten conectarse rápidamente a todas las fuentes de datos requeridos, incluidas las suscripciones de terceros y la incorporación de empresas confiables que se encarguen de procesar esta información mediante plataformas digitales.

Estas plataformas digitales son capaces de proporcionar la agilidad necesaria para agregar nuevas fuentes de datos y actualizar atributos de conformidad con las demandas del negocio.

Una vez que las reglas del anejo de los datos estén bien definidas y entendidas, prosigue la implementación de la transformación de datos según los mercados, funciones y sistemas. Este diseño de organización y análisis de datos garantiza que sólo exista un flujo de datos coherente que fluye a través de todos los sistemas y grupos funcionales de información de relevancia para la industria farmacéutica. La organización y análisis de datos es fundamental para que los productos de la industria farmacéutica puedan comercializarse con mayor



facilidad. Cuando tengamos la base de datos del laboratorio farmacéutico es fiable, se deben crear reglas de acceso, así los datos recogidos pueden ser ofrecidos a empleados mediante distintos sistemas. Toda esta información recolectada será la calve para los trabajos futuros y para el desarrollo de nuevos fármacos y formas de comercialización, logrando así facilitar la llegada del fármaco al paciente que lo necesita.

1.3 Como los grandes volúmenes de datos contribuyen para la toma de decisiones.

Como su nombre lo sugiere el Big Data es precisamente eso, "una gran cantidad de información". La aplicación adecuada del Big Data mejoraría la eficiencia de la empresa. Sin embargo, es necesario llevar a cabo un análisis detallado antes de incorporar de manera efectiva las soluciones que el Big Data ofrece.

En una oficina de farmacia, por ejemplo, con la información que proporciona esta herramienta te podrías informar de forma general de qué productos se están vendiendo más en tu farmacia o qué están buscando las personas y tú no tienes.

Este tipo de información relevante para un negocio se puede recoger y almacenar gracias a los recientes avances en tecnología. Los beneficios que puedes obtener con el uso adecuado de este medio es incalculable. Con este tipo de información segmentarías la publicidad y sus resultados serían más positivos. Ya no solo se visualizaría a un público general, sino segmentado por género, edad, condición física, entre otros. Así como también tanto los farmacéuticos como suministradores podrían recopilar datos importantes como la cantidad de productos disponibles, precio, informes de venta, visitas diarias y analizar más de cerca las necesidades de los pacientes.

Es esta realmente la verdadera magia del Big Data, introducimos un gran volumen, y gracias a complejos algoritmos matemáticos estos programas relacionan o correlacionan lo que la mente humana es incapaz de realizar.



Sin embargo, a pesar de todas estas virtudes, al sector farmacéutico le ha costado reaccionar y establecer estrategias de marketing online. Pese a la resistencia, laboratorios y grupos farmacéuticos importantes están generando a diario gran cantidad de información. Por lo tanto, es absolutamente necesario desarrollar técnicas adecuadas de análisis de grandes volúmenes de datos para que las farmacias y laboratorios ofrezcan un mejor servicio y sean más competitivos. Se puede estimar que, aplicando las estrategias de Big Data en la industria del cuidado de la salud se podrá mejorar la toma de decisiones y generar optimización de recursos que lleven a tener una mejor administración monetaria. Así como también se podría optimizar y diseñar nuevas herramientas que mejoren la eficiencia en la investigación y en la demanda de medicamentos, incrementando el volumen de ventas para favorecer una mejor gestión tanto para consumidores, farmacias y la industria. En el entorno empresarial la oportunidad de emplear el Big Data es bastante convincente. La utilización efectiva de estos datos ayudará a las compañías farmacéuticas a identificar mejor quienes son los candidatos potenciales a fármacos y desarrollarlos de manera más rápida y eficaz en medicamentos aprobados y reembolsados.

En el Big Data encontramos ventajas que nos ayudarán a la empresa en las actividades que se quieran hacer como: Ahorro de costes, Diagnóstico eficiente, Creación de patrones de enfermedad, Mejoras en la planificación e interpretación de los resultados, Mejor predicción de la seguridad y eficacia de los medicamentos logrando beneficios de Big Data para farmacias y proveedores: Revisar los datos en tiempo real ha sido un factor clave en el desarrollo de estrategias que requieren una toma de decisiones rápida.

El control de almacén en la farmacia a menudo se realiza con herramientas no inteligentes que requieren dedicación exclusiva del personal. Esto puede causar pérdidas importantes cuando se cometen errores en las compras, o cuando el producto no se compra a tiempo y hay una ruptura de stock. Esto se puede evitar si consideramos cada uno de los procesos de planificación en la cadena de suministro de la farmacia y así predecimos los eventos que cambiarían el stock.

Además, el Big Data le permite a las farmacias: Mejorar la calidad del servicio a tus clientes, Encontrar productos al instante, Obtener un informe de stock, rápida y fácilmente,



Universidad de Buenos Aires
Facultad de Ciencias Económicas
Escuela de Estudios de Posgrado



Verificación de ofertas, pedidos, facturación o contactos, Disponibilidad de una lista de ofertas de los productos que no están en stock, Mejorar la imagen de tu marca, Tener registros de entrada y salida de productos, Gestionar todos tus pedidos de forma integrada (online y offline), Manejo adecuado de la demanda anticipada en función de sus datos históricos, En definitiva, el uso de Big-Data para mejorar la efectividad de la administración de tu farmacia.



Apartado 2: Procesamiento de la base de datos utilizada para la toma de decisiones.

El presente análisis se realiza en base a los datos de una Industria Farmacéutica que tiene en funcionamiento medicamentos genéricos en el Mercado de las cadenas de farmacias Ubicado en Ecuador. La misma proviene en distintos formatos y con distintas maneras de generar la información dependiendo de cada región en donde están las sucursales de empresas, teniendo su matriz en una sola Región. La empresa maneja una gran cantidad de información de parte de ellos como Industria farmacéutica y recopilan base de datos de las dos cadenas de farmacias mas grandes donde se distribuyen los medicamentos a nivel de todo el país.

Es por eso por lo que, como primer paso, el proceso de recopilación de información que requiere unificar en una misma base de datos de cada de estas diferentes fuentes, una matriz en cada ciudad secundaria, de las cuales tienen una por región primaria todo eso se requiere unificar para contar con una matriz que tenga todos los datos de cada uno de los medicamentos.

Cabe señalar que la información utilizada fue provista con consentimiento de la Gerente de la Industria Farmacéutica, y que los datos se encuentran absolutamente anonimizados para preservar la identidad y privacidad de la empresa.

En la cual está registrado todo en una misma base de datos registran los nombres de los productos que se entregan a dos cadenas de farmacias, el precios de venta, precio por unidad, cantidad de cada producto, las ciudades donde se entregan los fármacos, se presenta las dos cadenas de farmacias, también encontramos el stock por punto de venta es decir por farmacia, la unidad de medida es caja completa pero en la data también podemos ver que está mes de caja completa y también unidades sueltas mezcladas, esto es preciso normalizar para evitar inconsistencias.



Este set de datos debe permitir tener una data cuando sea requerida mas precisa en buscar información, sea en las unidades de medicamentos que están disponibles en cada farmacia, como también ver el stock de ventas por cadena o productos de farmacias, Si yo quiero conocer las compras que realizaron en cada una de las cadenas de farmacias, como principal objetivo es evitar la redundancia.

2.1 Proceso de recopilación de los datos

La base de datos de la Industria Farmacéutica contiene un total de datos de 110,131 registros que contienen como información del stock en las cadenas de farmacias de productos de una de las líneas de productos de la empresa Farmacéutica. La cadena de farmacias en la cual encontramos los nombres de los productos genéricos (no contiene nombre comercial), las ciudades donde se distribuye el producto (zonas, ciudades), nos indica la base de datos que son registros del año (2017 y 2018), los nombres de la Unidad de Negocio, nombre de las categorías de cada medicamento (para que sirve o corresponde cada uno de ellos), los nombres de los locales y donde se entregan para ser comercializados, los códigos Ítem, los nombres de cada medicamentos (nombre del Ítem) que corresponden en esta columna si es en gotas, tabletas, cápsulas, una columna que nos muestra el estado del producto (continuo, discontinuo, próximo), la columna “BB” nos muestra la Unidad de empaque, tenemos también el Costo Real, Mínimo, Cantidad Venta, Venta Neta, Stock Precio de Venta, Stock USD.

Análisis de Datos. - Cada sistema de datos posee tipos de campos que pueden ser similares o diferentes, en esta Base con la que estamos trabajando podemos encontrar: Numéricos. - Entre los diferentes tipos de campos numéricos podemos encontrar enteros “sin decimales”, y reales “decimales”. Fechas. - Almacenan fechas facilitando posteriormente su explotación, posibilita ordenar los registros. Alfanuméricos. - contienen cifras y letras. Memos. - Estos campos son particularmente adecuados para dotar a cada registro de la tabla de un lugar para escribir todo tipo de comentarios.



Preparación de Datos. - Se busca predecir una variable de Precio de Stock en Ventas, para lo cual se tuvo que adaptar la base de datos para poder utilizarla y aplicar los algoritmos correspondientes para poder realizar las predicciones.

De la base de datos se desprende que la Industria Farmacéutica puede ser muy heterogéneo dado que las ventas pueden variar de acuerdo con las necesidades es decir varían mucho y sus cantidades tienen a variar de manera diaria; a partir de este trabajo se pretende ver que características de las columnas con sus diferentes aportes de departamentos que predicen mejor el Stock de ventas, puede ser por las categorías, por la cadena de farmacias (1,2), tenemos también los diferentes nombres de los medicamentos en genéricos, encontramos las ciudades pero por sectores (a,b,c).

2.2 Problemáticas vinculadas con la base de datos

Industria Farmacéutica “XX” es una empresa de medicamentos genéricos, ecuatoriana, que tiene como fin brindar una calidad de medicamentos para ser distribuidos en cada uno de los sectores de laboratorios y farmacias de dicho país.

Las diferentes fuentes de datos tenían información sobre el año del registro, la base de datos cuenta con un registro por medicamentos, en que ciudad o región es distribuido, cual es su presentación, el estado del producto (Continuo, discontinuo, próximo), desarrolla como descripción como esta la presentación del producto (cápsulas, comprimidos, jarabe, gotas, tabletas), también nos dice el stock de venta, Unidades, precio USD.

El archivo es una matriz que contiene muchas observaciones, datos no llenados en el campo, en la descripción constan todo lo del medicamento que eso fue llevado y analizado de una mejor manera al momento de comenzar a trabajar con la base.

Cada sistema de datos posee tipos de campos que pueden ser similares o diferentes, en esta Base con la que estamos trabajando se pueden encontrar: Numéricos.- Entre los diferentes tipos de campos numéricos podemos encontrar enteros “sin decimales”, y reales “decimales”. Fechas.- Almacenan fechas facilitando posteriormente su explotación, posibilita ordenar los registros.



A partir de este trabajo se pretende ver que características de las columnas con sus diferentes aportes de departamentos que predicen mejor el Stock de ventas, puede ser por las categorías, por la cadena de farmacias SanaSana – Fybeka, tenemos también los diferentes nombres de los medicamentos en genéricos, encontramos las regiones donde se distribuyen los medicamentos Costa, Sierra, Oriente.

Con el objeto de enriquecer los datos para esta Industria Farmacéutica, el proceso consistió en unificar en una misma base los datos de diferentes fuentes, de acuerdo con el siguiente detalle: A esa base de datos se le realizaron algunas transformaciones: la fecha de ingresado el medicamento se unifico a mes y año. Las ciudades fueron unificadas en regiones para delimitar el sector que es distribuido el medicamento. Las columnas del nombre de medicamento se lo dejo solo y se añadieron columnas que contengan la presentación de este. En Nombre del local que se distribuyen se separo en columnas las dos cadenas de farmacias con sus sectores. El estado Ítem actual Continuo, discontinuo, próximo se pudo sacar de la descripción mediante fórmulas y poder tener una base de datos mejor adaptada. En la base de datos también encontramos unidad de empaque, costo real, mínimo, stock PDV, cantidad Venta, Demanda Promedio, Stock USD, Días de inventario, estas columnas estaban con números puntos, comas, decimales, y se logro poder hacer un balance sobre esto para poder tener una base balanceada, de la cual se pueda trabajar con este set de datos, ya con atributos que aportaran con mi objetivo del trabajo.

2.3 Proceso de recopilación de datos final

Una base de datos nos permite guardar grandes cantidades de información de forma organizada para que luego podamos encontrar y utilizar fácilmente, son datos organizados y relacionados entre sí, como características en este trabajo podemos ver una base de datos en Redundancia Mínima, Integridad de los datos, Independencia Lógica y física de los datos.

Después de haber realizado de manera un Set de Datos Final, obteniendo una base de datos balanceada, en un nuevo archivo podemos formar ya un diagrama de entidad – relacional entre una base mayor que tendríamos original y las tablas que se crearon de forma que se codificaron

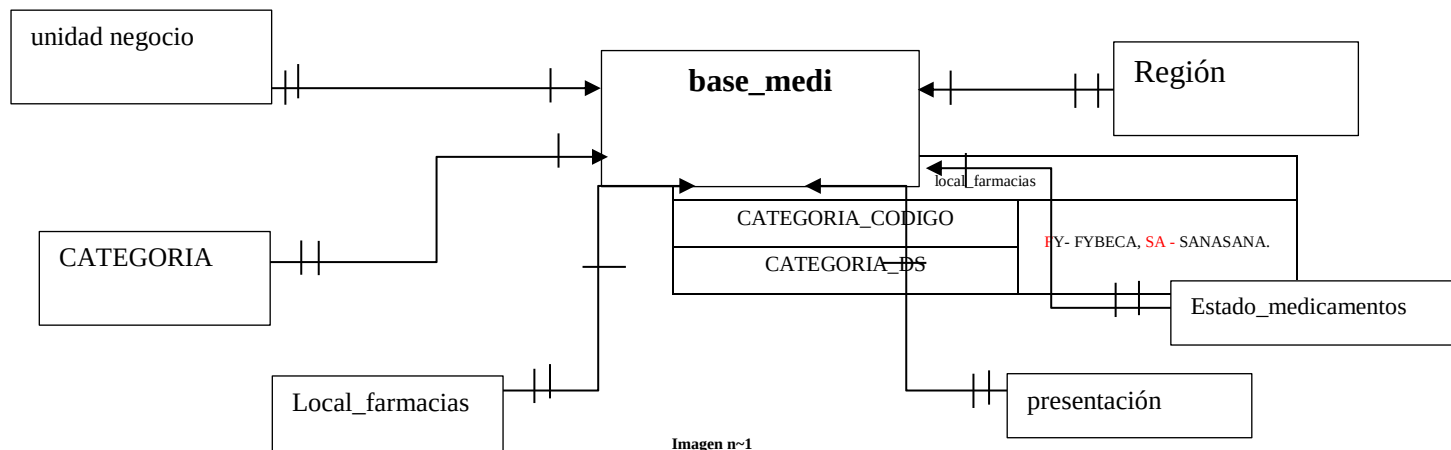


Imagen n~1

Imagen n~2

Esta base_medi, es la base en la que se carga toda la información de la base, pero para poderlo hacer se cargo esta base y a continuación se desglosa las tablas cargadas con su codificación en HeidiSQL.

CATEGORIA	
CATEGORIA_CODIGO	AD - ALIVIO DE DOLOR, AC - APARATO CARDIOVASCULAR,
CATEGORIA_DS	AL- APARATO LOCOMOTOR, AR APARATO RESPIRATORIO.

unidad_negocio	
CATEGORIA_CODIGO	EFC - ECONOFARM S.A, FRC - FARMOCED, FGT- FARMAGESTION.
CATEGORIA_DS	

estado_medicamento	
CATEGORIA_CODIGO	PRO - PROXIMO, CON - CONTINUO, DIS- DISCONTINUO.
CATEGORIA_DS	

REGION	
CATEGORIA_CODIGO	C- COSTA, S- SIERRA, O-ORIENTE.
CATEGORIA_DS	

CATEGORIA	
CATEGORIA_CODIGO	GT - GOTAS, JB - JARABE, TB- TABLETAS, TBR - TABLETAS RECUBIERTAS, CAP-CAPSULAS.
CATEGORIA_DS	

Imágenes n~3



Esta base permite manejar un conjunto de datos para convertirlos en información relevante para la organización, ya sea a nivel operativo o estratégico. Además, Los DBMS pueden trabajar con lenguajes de programación tradicionales o pueden incluir su propio lenguaje de programación.

2.4 Base de datos consolidada para el modelo de machine learning

Para la integración y preparación de los datos lo primero que se realizó es cargar la base de datos de medicamentos a Heidi SQL, denominada base_medi. Esta base de datos que se carga ya con la denominación que realizamos en el set de datos, y luego con la codificación que se realizó en cada tabla también se registra en Heidi SQL obteniendo nuestra base de datos inicial y seis tablas que contienen la codificación de información que simplifica el ingreso de datos a SQL.

Al cargar la base de datos se obtiene en la tabla que estará creada con contenido INT, DECIMAL, CHAR, de esta manera esta como tipo de dato que se van a utilizar en longitud variable e infinita, si son enteros y contienen decimales, si tienen caracteres alfanuméricos.

Nombre	Filas	Tamaño	Creado	Actualizado	Motor	Comentario	Tipo
base_medi	1,781	1,5 MiB	2019-09-28 12:02:55	2019-09-28 09:09:03	InnoDB		Table
categoria	4	16,0 KiB	2019-09-26 10:32:51	2019-09-26 11:21:27	InnoDB		Table
estado_medica...	3	16,0 KiB	2019-09-26 11:19:44	2019-09-26 11:21:28	InnoDB		Table
local_farmacias	2	16,0 KiB	2019-09-26 11:18:46	2019-09-26 11:21:28	InnoDB		Table
presentacion	5	16,0 KiB	2019-09-26 10:39:25	2019-09-26 11:21:29	InnoDB		Table
region	3	16,0 KiB	2019-09-26 10:35:45	2019-09-26 11:21:28	InnoDB		Table
unidad_de_neg...	6	16,0 KiB	2019-09-26 11:18:03	2019-09-26 11:21:26	InnoDB		Table

Imagen n~4

En las tablas también podemos ver que están creadas con un código y una descripción que refiere a cada columna de tabla que fue creada.

Mediante Heidi SQL se generan tablas dentro de la base, que refieren a cada una de las tablas creadas con código y descripción, en este caso por ejemplo unidad negocio que refiere a un código y una descripción. Y también podemos ver y verificar que la tabla se encuentra correctamente cargada. Esto se hace en las seis tablas cargadas y verificadas cada una de ellas. Incorporamos la base de datos donde se encuentra toda la información de cada una de las columnas que componen dicha base.



Apartado 3: Modelos de predicción a través de machine learning

Un set de datos bien analizado, con una base de datos con tablas, filas, columnas, códigos bien realizados, que nos lleven a tener una información eficiente y eficaz hacen que la empresa pueda tener mejores resultados en su toma de decisiones. Al trabajar esta Base de datos de la Industria farmacéutica nos vamos a dar cuenta cuanto tiempo ahorraríamos por obtener cada una de esta información, que sería relevante en todo lo que nos puede mostrar un set de datos bien distribuido a continuación observamos maneras en las que obtendríamos una información para entregar de manera rápida y oportuna en una organización.

En una organización que se trabaja con distribución de productos en este caso medicamentos a dos cadenas de farmacias es de total relevancia poder tener en el set de datos listar la cadena de farmacia que queramos en ese instante comentar u observar sus parámetros.

3.1 Modelos de predicción

La predicción en el stock de ventas para la Industria Farmacéutica, proponiendo como un enfoque operadores y algoritmos que nos den resultados de un estudio de la distribución en cuanto al Stock de las Cadenas de farmacias donde se distribuye el medicamento, Se aplica un balanceo a la base de datos porque estaba desbalanceada con un operador que distribuye el peso obtenido a la base balanceada, se aplicaron técnicas de regresión, bajo un procedimiento computacionalmente eficiente que realiza la selección simultánea de variables y modelos.

Se puede apreciar que los modelos que nos dan un conjunto de aprendizaje automático para mejorar la clasificación y los modelos de regresión en términos de estabilidad y predicción de clasificación que nos reduce la varianza y ayuda a evitar el sobreajuste que fue combinado con las redes neuronales que proveen mayor precisión predictiva que los algoritmos de aprendizaje simbólicos comúnmente utilizados (por ejemplos los árboles de decisión).

Aplicar un modelo de predicción por puntos de ventas, sobre un conjunto de datos del Stock de ventas en las cadenas de farmacias, con el fin de encontrar relaciones entre dos o más



productos, realizando la gestión necesaria para obtener una base de datos que permita la aplicación de alguna de las técnicas de Predicción. Determinar la técnica a utilizar sobre la base de datos de acuerdo con los datos que ésta contenga. Realizar las funciones de manipulación, selección y procesamiento de los datos. Dar una breve explicación de los resultados obtenidos y el por qué de los mismos.

3.2 Modelo de regresión lineal

Uno de los aspectos mas relevantes de la Estadística es el análisis de la relación o dependencia entre variables, Frecuentemente resulta de interés conocer el efecto que una o varias variables pueden causar sobre otra, e incluso predecir en mayor o menor grado valores en una variable a partir de otra. Se construye un modelo formal con el que se obtiene una expresión lógico-matemática que, a demás de resumir cómo es esa relación, va a permitir realizar predicciones de los valores que tomará la variable respuesta, Tabla que representa mediante el Modelo de Regresión Lineal, cuales son las variables que tienen mas efecto en la base de datos. (anexo1)

Como se puede observar en la tabla 1 (anexo1), las categorías que están sin marcar nos representan que son las que mejor explicarían el Precio Stock (porque influyen positivamente en el precio), Es decir que los precios más altos corresponderían a los que representan estos atributos, mientras que los atributos que están marcados influyen negativamente en el precio,

Pero con esta Predicción que nos daba de resultados se puede visualizar el desbalance de la base de datos y se usa para balancear datos el algoritmo Generate Weight (Stratification) que es un operador que distribuye el peso específico entre todos los ejemplos, de modo que los pesos se resumen por igual, divide el peso especificado a través del parámetro peso total entre todos los ejemplos. Mientras divide el peso, este operador se asegura de que la suma de los pesos de ejemplo de todos los valores de la etiqueta sea la misma. Esto nos mejorará la representatividad de los valores de la etiqueta. El parámetro será Rango = Real.



Al que también le conectamos el operador que crea una muestra Sample (Bootstrapped) que utiliza el muestreo con reemplazo, por lo que es posible que la muestra no tenga todos los ejemplos únicos de la muestra se puede especificar sobre una base absoluta y relativa, este operador utiliza el muestreo con reemplazo, todos los ejemplos tienen la misma probabilidad de ser seleccionados, una vez que se ha seleccionado la muestra sigue siendo candidato para la selección y puede seleccionarse nuevamente en cualquier otro paso. (Anexo2)

3.3 Especificación del método

Después de imputar los datos se logra obtener por medio del operador un Excel con los valores y variables preparados para realizar un proceso automático que queremos para poder predecir nuestro Objetivo planteado, por lo cual realizaremos una validación cruzada para estimar el rendimiento estadístico de un modelo de aprendizaje, con esto vamos a poder estimar la precisión, contaremos con dos subprocesos, un subproceso de capacitación y un subproceso de prueba. Y vamos a medir el rendimiento mediante una fase de prueba. Realizamos una validación cruzada para estimar el rendimiento estadístico de un modelo de aprendizaje, con este estimaremos la precisión con la que un modelo se desempeñará en la práctica.

La Evaluación del rendimiento de un modelo en conjuntos de pruebas independientes arroja una buena estimación del rendimiento en conjuntos de datos invisibles. También muestra si el modelo representa muy bien los datos de prueba, pero no se generaliza bien para los nuevos datos, Para la determinación de los modelos, en una primera instancia, se realizaron entrenamientos con algoritmos los que nos dan en comparación con ejemplos de los vecinos más cercanos al ejemplo desconocido donde la similitud se calcula mediante dada en el espacio de atributo. Un árbol de decisión es un árbol, como una colección de nodos destinados a crear una decisión sobre afiliación de valores a una clase o estimación de un valor objetivo numérico. Cada nodo representa regla de decisión para un atributo específico, la construcción de nodos se repite hasta que se cumplan los criterios. También logramos generar un modelo de bosque aleatorio, que se puede utilizar para la clasificación y regresión, por medio de un conjunto de un cierto número de árboles aleatorios, especificado por el parámetro número de árboles, estos árboles son creados y se entrenan en subconjuntos separando los valores de una manera óptima. Probamos un modelo que es determinista solo



si el parámetro reproducible se establece verdadero, se basa en una red neuronal artificial de alimentación avanzada de varias capas que entrena, y contribuye periódicamente al modelo global a través del promedio del modelo en toda la red. Por último, se realizó una técnica de regresión utilizada para la predicción numérica, siendo una medida estadística que intenta determinar la fuerza de la relación entre una variable dependiente y una serie de otras variables cambiantes conocidas como variables independientes, intentando moldear la relación entre una variable escalar y una o más variables explicativas ajustando una ecuación lineal a los datos observados de esta manera lo utilizamos en nuestra base relacionar los precios Stock con las cantidades de stock.

3.4 Modelos aplicados individualmente

Determinando todos estos modelos, se lograron combinar dichos modelos – ensambles; obteniendo mejores resultados con las predicciones de los ensambles que con los modelos aplicados individualmente. Se utiliza la técnica de Ensamble en la búsqueda de mejorar la performance de los modelos. Los Ensamblados fueron creados con modelos que tenían una mejor performance, con distintos algoritmos, parámetros y diferentes conjuntos de entrenamientos, dejando como referencia que los conjuntos de entrenamientos varían en la muestra de registros y muestras de atributo

A continuación, se presentan los resultados de algunos de los modelos realizados, que lograron dar resultados más relevantes, se debe considerar que la base de 110,131 en algunos modelos dio resultados 4.124 que no fueron considerados por el tiempo de procesamiento, y no arrojaban medidas significativas a modelos más sencillos.

Como punto relevante podemos decir que se pudo apreciar que con el operador Generalized Linear Model que se utilizó, se notó que Linear Regression sólo arrojaba mejores resultados. Dejando que en relación a que podrían mejorarse los resultados que a continuación se exponen en los casos donde Regression Lineal para un posterior análisis.



CONCLUSIÓN

Como puede observarse, se obtuvieron mejores resultados en los modelos que nos dan un conjunto de aprendizaje automático para mejorar la clasificación y los modelos de regresión en términos de estabilidad y predicción de clasificación que nos reduce la varianza y ayuda a evitar el sobreajuste que fue combinado con las redes neuronales que proveen mayor precisión predictiva que los algoritmos de aprendizaje simbólicos comúnmente utilizados (por ejemplos los árboles de decisión), El mínimo Error obtenido fue de **855,97**.

Sin embargo, se observa que debería mejorarse el RMSE. En ese sentido, para lograr mejores predicciones, y de acuerdo a los resultados obtenidos en este trabajo, podría incorporarse información más precisa sobre las cantidades distribuidas semanal o diarias, para tener una mejor precisión y sería importante tratar de clasificar lo mejor posible la entrega que se hace por sectores semanalmente (por ejemplo, se podría incorporar una ruta para la entrega de medicamentos de acuerdo a las cadenas de farmacias para poder descifrar que medicamentos en un mes pueden ser mas vendidos que otros en diferentes sectores y cadenas de farmacias que se hacen entregan, dando por ejemplo que un visitador lleva las promociones del medicamento y hacer seguimiento si tuvo una mejor aceptacion en cual de las dos cadenas de farmacias de los medicamento distribuidos por la red de cadena de farmacia teniendo la misma promoción en las dos cadenas de farmacias). Teniendo en cuenta también que existe información disponible para agregar sobre si se venden en cajas o sueltas los medicamentos, sería interesante incorporar detalles de cada medicamento clasificado de esa manera para mejorar la predicción.

Por último, para otra instancia del análisis, sería interesante graficar la predicción de los precios del stock de ventas en USD. del modelo con menor RMSE y los precios reales de ventas de cada una de las cadenas de farmacias para tener una mejor idea de la capacidad predictiva del mismo.



Referencias bibliográficas

Bender, A., & Mazza, N. (2017). Análisis predictivo: difusión e impacto en áreas de la sociedad. In *Simpósio Argentino sobre Tecnología y Sociedad (STS)-JAIIO 46* (Córdoba, 2017).

Reyes, G. V., Thompson, E. B., Vanoye, J. A. R., & Penna, A. F. (2018). MODELOS DE TECNOLOGIAS DEL BIG DATA ANALYTICS Y SU APLICACIÓN EN SALUD. *Pistas Educativas*, 39(128).

Lúa-Lemus, H., & Estrada-Ramírez, M. (2019). Optimización de inventario de división de salud animal en comercializadora de ingredientes activos farmacéuticos.

Briones Bozza, D. M. (2019). Análisis de los factores que inciden en el comportamiento de compra de medicamentos genéricos desde la perspectiva del canal de distribución en el cantón Balzar.

Mogollón Espinoza, J. L. (2015). *Influencia en las ventas de la logística de inventario para dos distribuidoras farmacéutica del sector centro de la ciudad Guayaquil* (Master's thesis, Universidad de Guayaquil Facultad de Ciencias Administrativas)

Campoverde Velasco, A. E., & Panjón Barreto, E. C. (2016). *Análisis financiero y administrativo a la partida inventarios de la distribuidora de productos farmacéuticos y medicinales Silvio Abad (DISA) de la ciudad de Cuenca, durante los períodos 2013 y 2014* (Bachelor's thesis).

Montes Chino, A. A. (2005). *La administración del departamento de crédito y cobranzas en empresas farmacéuticas* (Master's thesis, Universidad Iberoamericana Ciudad de México. Departamento de Administración y Contaduría Pública).

Alberto, O. G. (1997). *GERENCIA FINANCIERA: UN ENFOQUE ESTRATEGICO*.

Campoverde Velasco, A. E., & Panjón Barreto, E. C. (2016). *Análisis financiero y administrativo a la partida inventarios de la distribuidora de productos farmacéuticos y medicinales Silvio Abad (DISA) de la ciudad de Cuenca, durante los períodos 2013 y 2014* (Bachelor's thesis).

López Parra, M. F. (1997). *Contrato de gestión: viabilidad en la administración pública del Ecuador: un estudio estratégico para su implementación* (Doctoral dissertation).



Apéndices

Apéndice 1

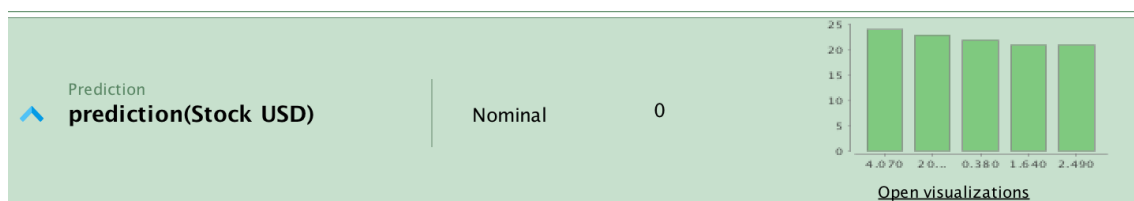
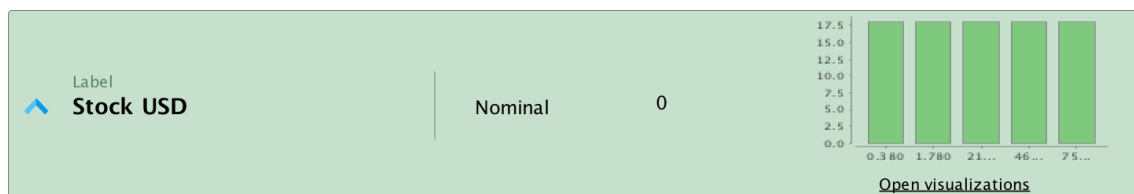
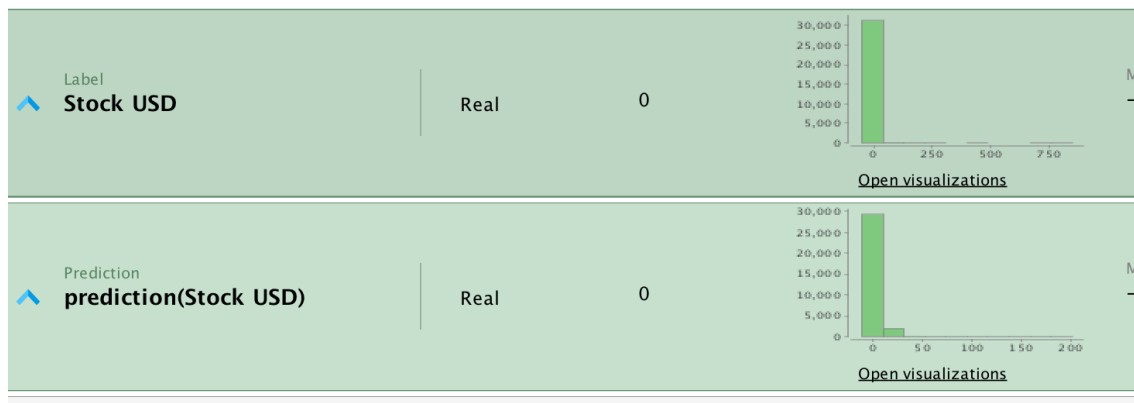
ATTRIBUTE	C OEFFICIENT	STD. ERROR	STD. COEFFICIENT	TOLERANC E	T-STAT	P-VALUE	CODE
AÑO 2017	0.062275656676211585	0.02081764285764266	0.07230373740021676	0.9939295166714098	2.991.484.535.596.620	0.0027766772926880456	***
AÑO 2018	0.06262698223042244	0.020817642857642663	0.01788465125437801	0.9939295166714093	30.083.608.724.909.200	0.0026270268817926157	***
ALIVIO DEL DOCLOR OTC	0.9424132939000209	0.02132610049937408	-0.013553159019590741	0.9470999032729391	44.190.605.494.318.100	0.0	****
ANTIIMFECCIOSOS VIA GENE	0.2528301394967493	0.020764468569609472	-0.055365701804690436	0.9990266051000756	12.176.094.882.909.100	0.0	****
ANTIPARASITARIOS	0.24560655191971506	0.020761293378841916	-0.02311466182426213	0.9993322067410662	1.183.002.173.506.270	0.0	****
APARATO CARDIOVASCULAR	0.808083943086179	0.02076365502943417	-0.08012306756488835	0.9991048923011564	3.891.819.344.622.390	0.0	****
APARATO DIGESTIVO METABOL	0.9735443794423889	0.020790273352639277	0.2021886424334303	0.9965481705419231	4.682.691.578.553.960	0.0	****
APARATO LOCOMOTOR	0.6633705322668796	0.02076349482323088	-0.03593336789503993	0.9991203100736009	31.948.886.154.014.800	0.0	****
APARATO RESPIRATORIO	0.6941399575076439	0.020792187897316503	0.09493752079661297	0.9963646546783677	3.338.465.201.140.430	0.0	****
DERMATOLOGICO	0.1716986079576542	0.020763588567737305	-0.09527038041714643	0.9991112883349101	8.269.216.440.959.600	1.11E+00	****
GRIPE OTC	0.1301148400374936	0.020813283708551036	-0.10021646175051935	0.9943458988902002	6.251.528.680.408.870	4.07E+06	****
PRO.GENITO URINARIOS	0.5315291751147111	0.020812244276850626	-0.06209899390181542	0.99444522315178109	255.392.531.456.076	0.0	****
SANGRE Y ORGANOS HEMALOP	0.22190845111772375	0.020772048915462357	-0.004460547971907923	0.9982975883919799	10.683.031.414.996.200	0.0	****
SISTEMA NERVIOSO	0.7692081310701803	0.0207581020840599	0.21627698234002216	0.999639499653448	37.055.802.498.478.500	0.0	****
ECONOFARM S.A	19.410.782.998.752.600	0.020767879900931267	0.0011734984827590413	0.9986984318989713	9.346.540.470.836.530	0.0	****
FARCOMED	0.3449722982603812	0.020774446667689732	0.01966714427833136	0.9980671579810736	1.660.560.706.037.540	0.0	****
FARMAGESTION	0.9114318155763751	0.02075504537078121	5.39E+11	0.9999339664911874	4.391.374.720.189.630	0.0	****
COSTA	0.914627378823286	0.020754868936403166	0.0017468699966358006	0.9999509671756205	4.406.808.742.492.750	0.0	****
SIERRA	0.9621114067189123	0.020755541135328202	0.011654914957411978	0.9998861984412835	46.354.436.169.399.200	0.0	****
ORIENTE	0.5961710215578531	0.0207547636235454	-9.37E+09	0.9999611150129404	28.724.539.212.844.700	0.0	****
FAR_SANA_SANA	0.042822745974320675	0.020773366589923344	0.03507486855849335	0.998170946472396	20.614.254.212.936.700	0.03926411553764064	**
FAR_FYBECA	20.763.310.546.533.700	0.02077268499834487	-0.005264517008140682	0.9982364513488228	9.995.487.125.611.390	0.0	****
CERRADO_FAR_SANA_SANA	0.011265976230933913	0.020756233443545036	-0.10066756678010252	0.9998194986817588	0.5427755600059273	0.5872851136397769	****
CERRADO_FYBECA	0.18881113459029322	0.02075665193516416	-0.08931187468010131	0.9997791827490293	909.641.570.230.435	0.0	****
OFICINA CENTRAL ECONOFARM	0.00517523640562869	0.0207543606036493	0.029230207364931923	0.9999999510090821	0.2493565812246084	0.8030852941551891	****
REGIONAL COSTA 2	0.01677053370736713	0.02075443552501458	0.009647267382748641	0.9999927312313256	0.8080457638635472	0.4190654631405063	****
REGIONAL COSTA APOYO SANA_SANA	0.11189106489092157	0.020758434163640347	0.008379378566465197	0.9996075167754275	5.390.149.565.659.700	7,05E+06	****
RES. EMPRESARIAL SANA_SANA	9,00E+10	0.020754367663663027	0.005425107909998593	0.9999992706695461	0.004334655453580276	0.996541461679658	****
RESP. EMPRESARIAL FYBECA	0.3367304187336022	0.020818418785304486	-0.037024923059054284	0.9938554281383639	16.174.639.496.218.400	0.0	****
MEDIGENER	0.05054111645851385	0.020754370341878964	-0.008024970105228733	0.9999990125804045	24.352.035.559.725.000	0.01488452075077562	**
NIFA	0.9664421652275487	0.021480657328621405	0.02377451176482499	0.9335198587644448	4.499.127.519.435.050	0.0	****
TABS_CAJA	0.8574237394160631	0.021253212226931797	-0.3356725230889491	0.9536072348860282	403.432.540.107.771	0.0	****
TABS_SUELTAS	0.2806197248963662	0.021732201409757002	-0.35278019477613626	0.912034469862289	12.912.623.052.093.400	0.0	****
GOTAS	0.09261697959084553	0.02076619275205643	0.02233986080102887	0.998860716956592	44.599.884.387.413.200	8,20E+07	****
GOTAS_PEDIATRICAS	0.08044482472422092	0.02076286647613831	-0.15734126508944715	0.9991807837885931	38.744.565.841.461.200	1,07E+12	****
JARABE	0.052082842596151556	0.020772889384902182	-0.11467373008232978	0.9982168079486952	250.725.075.511.188	0.012168438277363913	**



SUSPENSIÓN	0.3554515913426023	-	0.0207765882932740	4	0.04570692576996691	0.99786140947482	46	17.108.275.253.144.9	00	-	0.0	****
CREMA	0.0770423854719505	7	0.0207651965777301	7	-0.008024970105228733	0.99895655647865	44	3.710.168.848.320.73	0	2.07E+12	****	
CAPS	0.2282432302891962	4	0.0207901924773152	3	0.02377451176482499	0.99655592384357	75	10.978.408.715.467.1	00	0.0	****	
Estado Item Local_proximo	3.222.568.006.739.73	0	0.0208551716023393	75	-0.3356725230889491	0.99035559588053	07	15.452.128.940.421.9	00	0.0	****	
Estado Item Local_discontinuado	33.868.073.521.034.9	00	0.0208965146938454	7	-0.35278019477613626	0.98644069819865	89	16.207.522.650.181.4	00	0.0	****	
Estado Item Local_no discontinuado	0.2144701032715759		0.0213622841725114	62	0.02233986080102887	0.94389420392472	39	10.039.661.561.451.9	00	0.0	****	
Estado Item Local_inactivo	15.105.285.423.754.6	00	0.0207886625260519	34	-0.15734126508944715	0.99670261323262	44	7.266.117.002.392.53		0.0	****	
Unidad Empaque	11.009.059.972.363.1	00	0.0236396182099022	0	-0.11467373008232978	0.77079302993484	31	4.657.037.975.237.51	0	0.0	****	
Minimo	0.4388017086331000	3	0.0226443748300504	4	0.04570692576996691	0.84003624516430	18	19.377.956.420.805.3	00	0.0	****	
Cant Venta	1.590.374.059.736.95	0	0.0245223782875320	75	-0.165658217971573	0.71629763937634	44	6.485.398.932.719.12		0.0	****	
Venta Neta	17.940.621.464.400.3	00	0.0245701943100166	2	0.18687499100548727	0.71351238044673	02	730.178.249.224.771		0.0	****	
Stock PDV	8.238.779.898.433.14	0	0.0264159010181204	55	0.8581764697881513	0.61728800516388	43	3.118.871.430.045.71	0	0.0	****	
(Intercept)	2.959.678.091.744.17	0	Infinity		NaN	NaN		0.0		1.0		

Apéndice 2

De esta manera presentamos como estaba la base antes de utilizar estos operadores:





Apéndice 4

Se muestran los cuatro mejores resultados de mis predicciones, para que se puedan apreciar y ver el modelo en la tabla que se realiza de manera siguiente donde se explica que modelo se uso, y cuales fueron los algoritmos, parámetros utilizados.

MEJORES MODELES A PREDECIR	RESULTADOS
9	1.045,9
10	870,50
11	990,90
12	855,97

Apéndice 5



Detalle de los modelos realizados y sus resultados (medidos en RSME)

Modelo	Algoritmo / Parámetros	Root Mean Squared Error (RMSE)
1	KNN 5 Measure types = mixedMeasure Linear sampling	Cross Validation number of Folds. 10 2.402
2	KNN 100 Stratified sampling	Cross Validation number of folds 100 2.228
3	Random Forest 1000 Criterion least square Max. Depth 10	Cross Validation number of folds 100 2.120
4	Bagging Sample ratio 0.9 Iterations 10	Cross Validation 100 Deep Learning activation tanh epochs 15 adaptive rate si hidden layer sizes 200/50 1090.87
5	Bagging Sample ratio 0.9 Iterations 30	Cross Validation 100 Deep Learning activation rectifier epochs 20 adaptive rate si hidden layer sizes 200/100 990,90
6	Bagging Sample ratio 0.85 Iteraciones 10	Cross Validation 100 Random Forest Number of trees 1000 Criterion least square Max. Depth 1.643
7	Bagging Sample ratio 0,9 Iteraciones 10	Cross validation 100 Generalized linear model family Auto Solver Auto regularization si 1.045,9

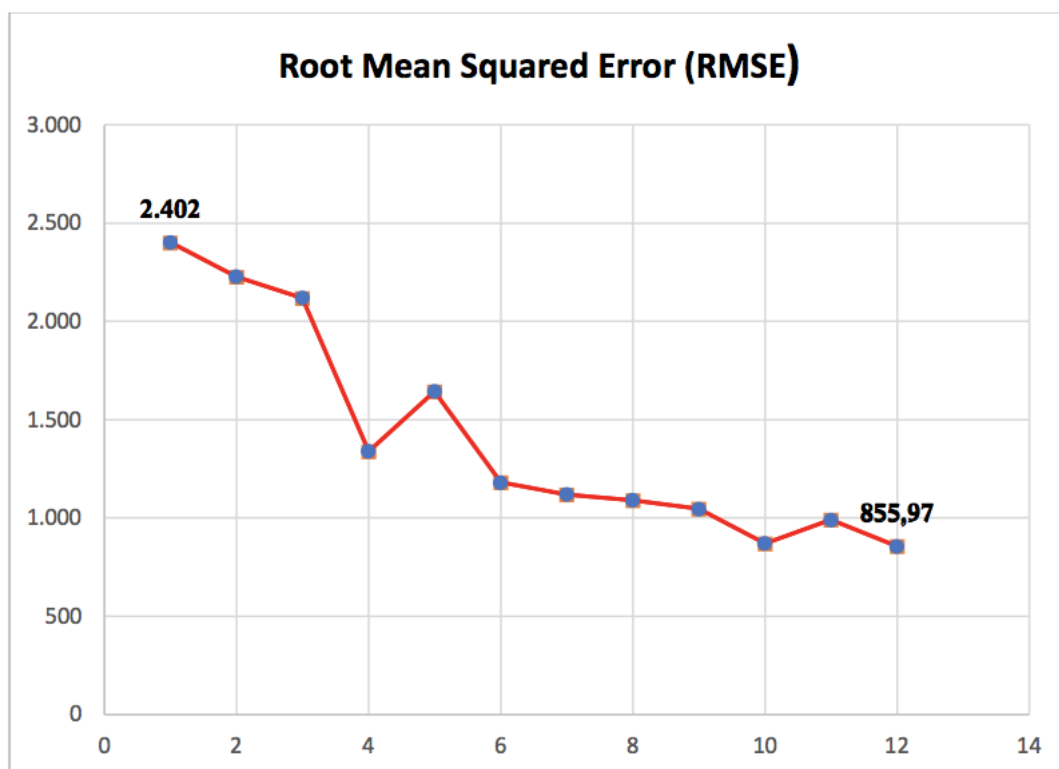


8	Stacking	Cross Validation	1.180
	Random Forest 100	stacking model learner	
	Criterio least square	Generalizaed linear model	
	max. Depth 10	family Auto	
		solver Auto	
	KNN K= 10	reproducible si	
		regularizacion si	
	Deep Learning	remove colinear columns	
	Activation Than		
	apochs 10		
	adaptive rate		
9	Stacking	Cross Validation	1.340
	Random Forest 100	stacking model learner	
	Criterio least square	Generalizaed linear model	
	max. Depth 10	family Auto	
		solver Auto	
	KNN K= 10	reproducible si	
		regularizacion si	
	Decision Tree 100	remove colinear columns	
10	Bagging	Cross Validation 100	870,50
	Sample ratio 0.9	Deep Learning	
	Iterations 20	activation rectifler	
		epochs 20	
		adaptive rate si	
		hidden layer sizes 300/50	
11	Bagging	Deep Learning	855,97
	sample ratio 0.9	activation rectifer	
	iteraciones 30	Epochs	
		adaptive rate	
		hidden layer sizes	
12	Stacking	Cross Validation	1120,76
	Random Forest 100	stacking model learner	
	Criterio least square	Generalizaed linear model	
	max. Depth 10	family Auto	
		solver Auto	
	Deep leraning	reproducible si	
		regularizacion si	
		renove colinear columns	



Apéndice 6

Se obtiene un gráfico por elaboración propia para poder analizar los resultados de los diferentes modelos obtenidos en RMSE que fueron realizados, estos se ordenaron de mayor a menor error (no de acuerdo con el orden que fueron realizados cada uno de los modelos).



Apéndice 7

Base de datos creada en HeidiSQL

base_medi	
NOMBRE	TIPO DE DATOS
Mes	INT
Nombre tipo de negocio	CHAR
Categoría	CHAR
Unidad de Negocio	CHAR
Región	CHAR
Código local	INT
Local de Farmacias	CHAR
Código ítem	INT



Presentación	CHAR
Estado de medicamentos	CHAR
Nombre ítem	CHAR
Unidad de empaque	INT
Costo real	DECIMAL
Mínimo	INT
Stock PDV	INT
Cant. Venta	INT
Venta Neta	DECIMAL
Demanda Promedio Unid	DECIMAL
Stock USD	DECIMAL
Días de Inventario Unid	DECIMAL

Apéndice 8