

Reconocimiento de patrones de demanda utilizando clustering en una empresa distribuidora de equipos de tecnología en Colombia

Ing. Juan Sebastián Herrera Cobo

Departamento de ingeniería civil e industrial

Pontificia Universidad Javeriana

Cali, Colombia

Abstract

Los pronósticos de demanda son quizá la tarea más difícil y crítica dentro del análisis de la cadena de abastecimiento ya que sobre estos valores se planea toda la operación, desde las compras y producción, hasta las ventas y servicio al cliente. Como un parte de los pronósticos de demanda, los patrones de demanda son comportamientos generales de la historia de la demanda que permiten tomar decisiones. En el siguiente trabajo se presenta la propuesta e implementación de un algoritmo que reconoce patrones de demanda en una empresa que distribuye equipos de tecnología en Colombia con el uso de clustering con el fin de mejorar la planeación de la logística de distribución en cada mes del año.

Palabras clave

Reconocimiento de patrones; patrones de demanda; pronósticos de demanda; clustering.

Introducción

Los pronósticos de demanda son quizá la tarea más difícil y crítica dentro del análisis de la cadena de abastecimiento ya que sobre estos valores se planea toda la operación, desde las compras y

producción, hasta las ventas y servicio al cliente, sea cual sea la empresa. Si ocurre un fallo en la predicción de la demanda, ocurre un fallo en toda la operación de la empresa, incluso hasta llevarla a la quiebra. Por ejemplo, para una compañía que fabrica cuadernos, una mala predicción de la demanda durante la época de inicio de períodos escolares puede hacerle perder millones de dólares, ya sea por sobrestock o faltantes.

Por esto, los pronósticos de demanda han sido siempre un campo de estudio riguroso en la ingeniería industrial, con ayuda de la estadística. Con el análisis de las series de tiempo se ha buscado predecir la demanda de las compañías para evitar problemas en sus operaciones para así disminuir costos por error y aumentar ventas y, consecuentemente, aumentar la utilidad.

Sin embargo, el análisis de las series de tiempo sólo fue el comienzo de una mejora en la continua búsqueda de precisión y exactitud. Primero fueron los pronósticos ingenuos, luego el promedio, el promedio ponderado, la regresión lineal y la suavización exponencial, para luego, debido a la acumulación de grandes cantidades de información, dar paso a las técnicas de Big data e inteligencia artificial

como las redes neuronales, los árboles de decisión y el clustering; siendo esta última utilizada no precisamente para predecir valores de demanda sino patrones de demanda.

En el siguiente trabajo se expone el caso de una empresa colombiana (llamada como ACME para mantener su confidencialidad) la cual comercializa productos tecnológicos en todo el país, más precisamente cámaras de seguridad y dispositivos de control de acceso, donde se aplica un algoritmo de clustering para el reconocimiento de patrones de demanda que faciliten la planeación de sus actividades en la logística de distribución.

La estructura del escrito será así: primero se hará una revisión de los patrones de demanda que se identifican en la actualidad; luego se expondrá con detalle el algoritmo de clustering; por último, se mostrará el caso de aplicación y sus resultados, así como sus conclusiones.

1. Patrones de demanda

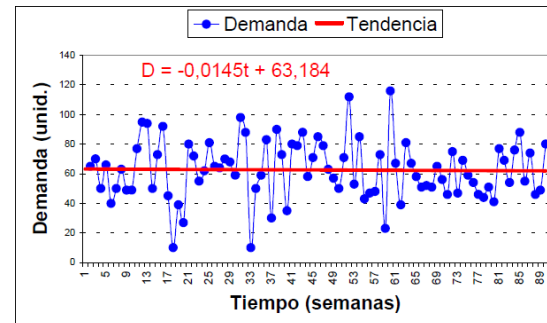
El análisis de datos históricos para los pronósticos de demanda es un proceso fundamental ya que, para cada comportamiento de demanda en específico, existe un método adecuado para predecir valores futuros de demanda.

En la siguiente sección se presentan los diferentes patrones de demanda que existen para calcular pronósticos de demanda según Vidal (2005) y el método adecuado para cada patrón. Con esto, el fin del algoritmo propuesto es el de reconocer estos patrones dentro de un mes de demanda para mejorar la planeación logística a corto plazo.

1.2. Demanda perpetua o uniforme

Demanda sin mucha variación en el tiempo y sin ningún tipo de tendencia.

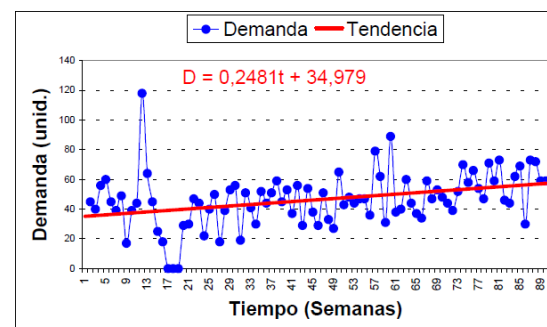
Ilustración 1 Demanda perpetua o uniforme. Fuente: Vidal (2005)



1.3. Demanda creciente

Demanda sin mucha variación, pero con una tendencia creciente.

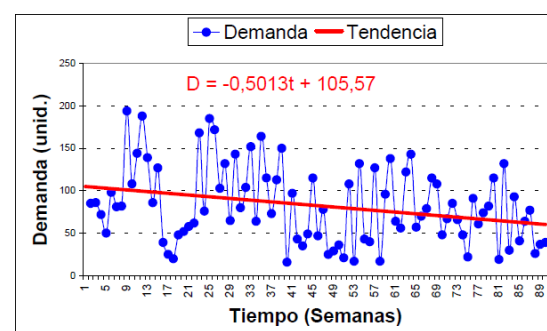
Ilustración 2 Demanda creciente. Fuente: Vidal (2005)



1.4. Demanda decreciente

Demanda sin mucha variación, pero con una tendencia decreciente.

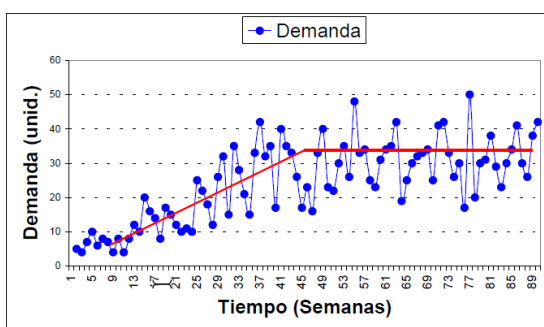
Ilustración 3 Demanda decreciente. Fuente: Vidal (2005)



1.5. Demanda creciente y luego uniforme

Demanda sin mucha variación con una tendencia creciente en un parte de su historia para luego convertirse en una demanda uniforme. Muy común en productos nuevos.

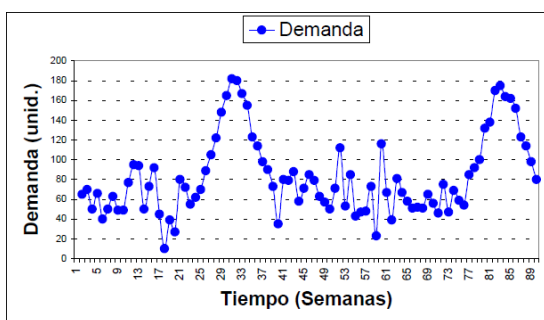
Ilustración 4 Demanda creciente y luego uniforme. Fuente: Vidal (2005)



1.6. Demanda estacional

Demanda con mucha variación ya que presenta picos cada cierto período en su historia. Muy común en productos de temporada como los útiles escolares o productos de tecnología como celulares.

Ilustración 5 Demanda estacional. Fuente: Vidal (2005)



Con esto patrones de demanda, Vidal (2005) resume los métodos indicados para cada patrón, como se ve en la Ilustración 6:

Ilustración 6 Sistemas de pronósticos recomendados por patrón de demanda

PATRÓN DE DEMANDA	SISTEMA DE PRONÓSTICO RECOMENDADO
Perpetua o uniforme	Promedio móvil o suavización exponencial simple
Con tendencia creciente o decreciente	Suavización exponencial doble
Estacional o periódica	Modelos periódicos de Winters
Demandas altamente correlacionadas	Métodos integrados de promedios móviles auto-regresivos (ARIMA)
Errática (Ítems clase A de bajo movimiento)	Pronóstico combinado de tiempo entre la ocurrencia de demandas consecutivas y la magnitud de las transacciones individuales

2. Algoritmo clustering K-means

Los algoritmos de clustering, o de agrupamiento, intentan agrupar individuos en una población por sus semejanzas, pero no con un propósito en específico (Provost y Fawcett, 2013), es decir, la agrupación está intrínseca en los datos; lo que lo convierte en un método de aprendizaje “no supervisado”, según la clasificación del aprendizaje automático o Machine Learning.

El más utilizado dentro de los algoritmos de clustering es el K-means. Según Cerreto et al. (2016), es un proceso iterativo de agrupamiento basado en la identificación de la media estadística de cada grupo. Cada grupo es representado por su centroide, calculado como la media de los elementos de cada grupo, y cada elemento es asignado al centroide más cercano. Dado un número k de clusters, el algoritmo se ejecuta en los siguientes pasos:

1. Asignar cada elemento al centroide más cercano.
2. Calcular los nuevos centroides de todos los clusters como la media de los elementos.
3. Repetir hasta que converja, la cual es alcanzada cuando ningún elemento cambia de clúster.

Cabe resaltar que el número de centroides, así como su posición inicial, son parámetros de entrada del algoritmo que pueden ser establecidos a partir de un conocimiento previo o por medio de ensayo y error.

3. Caso de aplicación

3.1. Descripción de la empresa

La empresa ACME tiene 7 años de trayectoria en el territorio colombiano y es reconocida como el distribuidor más grande de equipos de seguridad a nivel nacional con ventas aproximadas de 64 mil millones de pesos al año y 27 subdistribuidores. Actualmente posee más de 20 proveedores internacionales. Sin embargo, el 70% provienen de China y el 20% de Estados Unidos, lo que causa un Lead Time medio de 2 meses y medio. Además de estos proveedores tiene más de 20 proveedores nacionales.

Su red de distribución se compone de una bodega principal en Yumbo, Valle del Cauca donde se reciben las importaciones marítimas, las cuales llegan al puerto de Buenaventura, y tres puntos de venta en Bogotá (lugar donde se reciben las importaciones aéreas), Cali y Medellín. Llegan entre 8 y 10 órdenes en contenedores 40 High Cube mensuales a Yumbo y entre 4 y 8 órdenes aéreas a la ciudad Bogotá, con las cuales se abastece las sedes semanalmente: dos veces Cali y Bogotá y una vez Medellín, teniendo Bogotá el 43% de las ventas, Medellín el 29% y Cali el 28%.

En cuanto al cálculo de pronósticos y planeación de la demanda, la compañía utiliza un sistema, desarrollado por el departamento de cadena de abastecimiento, el cual utiliza cinco

métodos de pronóstico vistos anteriormente a sus más de 1000 referencias activas para calcular los próximo 180 días de demanda: promedio móvil, promedio móvil ponderado, suavización exponencial simple, suavización exponencial doble y regresión lineal, para las referencias con un coeficiente de correlación en su demanda mayor a 0,8. Luego de calcular los valores de estos métodos, se calculan sus métricas de rendimiento como el error absoluto, el error medio y la desviación de estos errores para luego escoger el método más exacto para cada referencia. Cabe resaltar que este procedimiento se ejecuta cada dos días para que la planeación sea sensible a cualquier cambio drástico en la demanda.

Con los pronósticos calculados, se clasifica el inventario en productos A, B y C, para los cuales se utiliza un modelo de pedido EOQ para los A y periódico para los B y los C, con uno y dos meses de período respectivamente.

A pesar de la tecnificación en los procesos de planeación de la empresa ACME, estos modelos de pronósticos son incapaces de reconocer patrones de la demanda dentro del mes calendario que pueden provocar faltantes o sobrestock por picos de demanda a inicio o final de mes. Por esto, considerando las ventajas que tiene el algoritmo de clustering en el reconocimiento de patrones, se decide probar que la implementación de esta técnica es la mejor solución para reconocer los patrones de demanda de la empresa ACME.

3.2. Descripción de la técnica

Como se ha mencionado, la técnica lo que busca es reconocer patrones de demanda para evitar faltantes en picos de demanda y sobrestock en valles de demanda. Así como el caso del alquiler de bicicletas urbanas en la ciudad de Nueva York de Chen et al. (2016) donde se utilizó el algoritmo de clustering para encontrar las zonas con más probabilidad de presentar un dato atípico en la demanda, dado un evento específico como un concierto, para planear el inventario y no tener faltantes. Otro ejemplo claro fue el trabajo de Cerreto et al. (2018) en el cual se implementó el algoritmo de clustering para agrupar los retrasos de las líneas de trenes, teniendo como datos de entrada los retrasos por estación por cada servicio, y así realizar minería de datos y mejorar el sistema. En la propuesta presentada en este escrito los valores de entrada son las ventas por día de cada referencia a lo largo de un mes para luego comparar con los comportamientos de otros meses y determinar patrones de demanda.

La implementación de este algoritmo se divide en tres partes: 1. Preprocesamiento de los datos, 2. La determinación del número óptimo de clusters y 3. Análisis de los clusters y reconocimiento de patrones.

3.2.1. Preprocesamiento de los datos

Teniendo en cuenta que dentro del dataset de la empresa se consideran ventas y devoluciones (siendo las devoluciones los valores negativos de venta), en el preprocesamiento se excluyen referencias que hayan tenido sólo devoluciones dentro de todo el mes o que, como consecuencia de la suma de ventas y devoluciones, su valor total de ventas en el mes sea menor o iguales a cero. Después se estandarizan

cada uno de los valores diarios con max-min, aplicando la siguiente fórmula:

$$r = \frac{\max - x}{\max - \min}$$

Siendo max el valor máximo del mes por referencia, min el valor mínimo del mes por referencia y x el valor diario a estandarizar. Luego de esto, se acumulan los valores desde el primer día hasta el último. La acumulación de las ventas es necesario para lo que el valor del último día siempre será uno ya que, si se toman los valores sin acumular, el algoritmo de clustering no hará un buen agrupamiento de los comportamientos de demanda.

3.2.2. Determinación de número óptimo de clusters

Con los datos ya preprocesados, se pasa a aplicar el algoritmo de clustering con un número de clusters entre 2 y 10, y teniendo la distancia euclidiana como la fórmula de distancia de cada punto hasta su centroide. En cada una de las ejecuciones del algoritmo se calcula el coeficiente Silhouette (Rousseeuw, 1987) y se grafican los clusters para cada uno de los meses con el fin de saber qué tan bien están agrupados los datos numérica y visualmente.

Por último, en base de los anteriores criterios, se escoge el número de clusters que agrupe mejor a los datos.

3.2.3. Análisis de los clusters y reconocimiento de patrones

Luego de ejecutar el algoritmo con el número de clusters indicado en el anterior paso, se pasa a identificar el patrón que representa cada clúster en cada mes visualmente, por ejemplo: productos que se venden al inicio del mes, producto que

se venden a final de mes o productos que se venden igual durante todo el mes.

Con esto, se procede a realizar una tabla resumen donde se consignan los clusters a los que perteneció cada producto en cada uno de los meses. De esta manera, y como paso final, se puede observar si la demanda de un producto se comporta con el mismo patrón durante todos los meses.

4. Resultados

Para la implementación del algoritmo de reconocimiento de patrones de demanda se utilizaron los datos de venta de los meses de enero a junio de 2016 de los productos con clasificación AAA, A y B ya que se observó que, si se incluyen los productos de la clasificación C, el algoritmo no arrojan buenos resultados porque existe demasiada variación en la venta, además de ser productos que sólo generan el 5% de los ingresos de la organización. El resultado de esto, luego del preprocesamiento, fue un dataset con información sobre 254 productos vendidos en cualquiera de los seis meses.

En relación con la herramienta utilizada para el preprocesamiento de los datos y la programación del algoritmo, se utilizó el lenguaje Python 3.6 con la ayuda de librerías como pandas y scikitlearn, y el entorno de desarrollo Spyder. Por otra parte, el algoritmo se ejecutó en un computador ACER Spin 5, Intel(R) Core(TM) i5-7200U CPU @2.50 GHz 2.71 GHz, con un sistema operativo Linux Ubuntu 16.0.4.

En la fase de determinación del número óptimo de clusters se tuvo dos momentos: el primero donde se redujeron las posibilidades de 9 a 3 y el segundo donde se estableció el número indicado de

clusters. En el primer momento (Ilustraciones 1, 2 y 3) se escogieron 3, 4 y 5 clusters. Ya para el segundo momento se determinó que tres clústeres era el número indicado, ya que en las otras dos opciones había grupos con características similares o con muy pocos productos. Además, generar tres clústeres, representa tres patrones demanda fundamentales para la planeación de la distribución logística: “se vende al inicio del mes” (Inicio), “se vende al final del mes” (Final) y “se vende durante todo el mes” (Normal).

Sin embargo, a la hora de realizar la tabla resumen para verificar el patrón de demanda por producto en cada mes, se encontró que menos del 50% de los productos tenían un mismo patrón en todos los seis meses de venta. Por esto, durante el análisis se crearon nuevos patrones desde los ya encontrados para un total de 7 patrones de demanda: “Inicio”, “Final”, “Normal”, “Inicio-Normal”, “Inicio-Final”, “Normal-Final” y “Errático”; donde “Inicio-Normal” son los productos que se venden normal durante todo el mes pero con más probabilidad de ser vendidos en el inicio, “Inicio-Final” son los productos que se venden al inicio o final de mes, “Normal-Final” son los productos que se venden normal durante todo el mes con más probabilidad de ser vendidos a final de mes y “Errático” son productos que no se pueden encasillar en ninguno de los seis patrones anteriores ya que su demanda ocurre en momento muy puntuales durante el año o mes.

A continuación, en la Tabla 1 se presenta un cuadro resumen de los patrones que se obtuvieron luego del análisis:

Ilustración 7 resultado del algoritmo con 3 clústeres. Fuente: propia

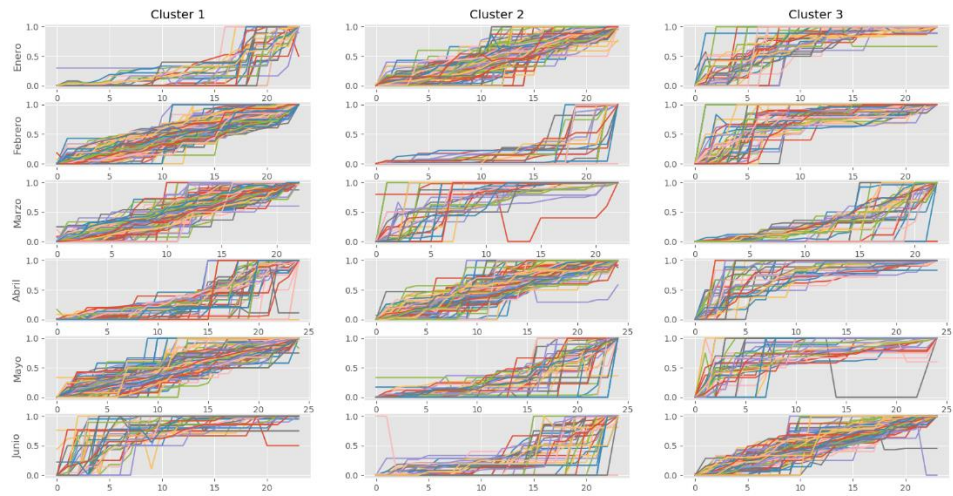


Ilustración 8 Resultado del algoritmo con 4 clústeres. Fuente: propia

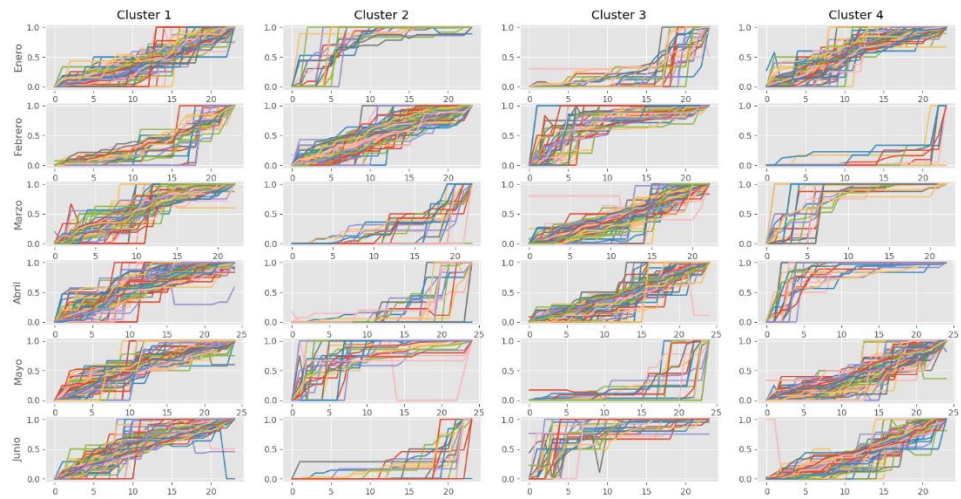
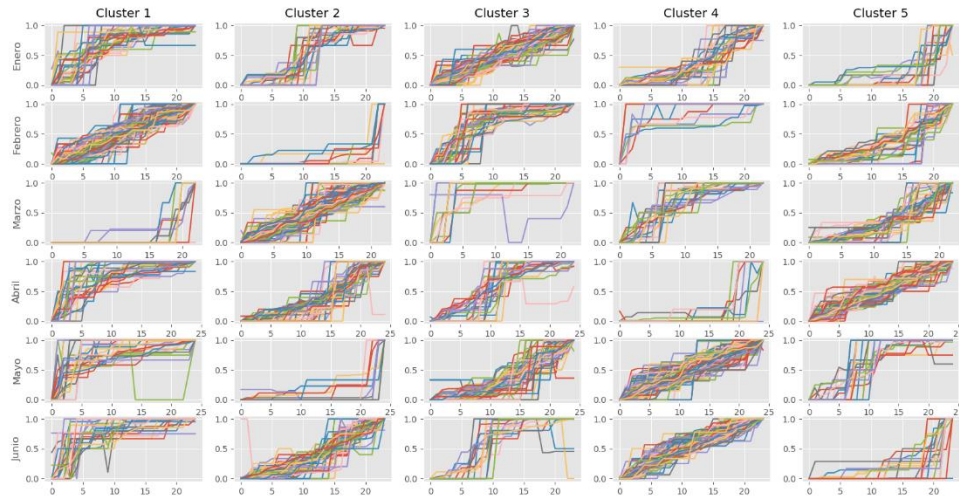


Ilustración 9 Resultado del algoritmo con 5 clústeres. Fuente: propia



Patrón de demanda	Frecuencia absoluta	Frecuencia relativa	Frecuencia relativa acumulada
Normal	97	38%	38%
Errático	70	28%	66%
Normal-Final	39	15%	81%
Inicio-Normal	32	13%	94%
Inicio-Final	8	3%	97%
Inicio	4	2%	98%
Final	4	2%	100%
Total general	254	100%	

Tabla 1 Tabla resumen de patrones de demanda. Fuente: propia

Con esto, se observa que el 66% de los datos corresponden a los patrones “Normal” (38%) y “Errático” (28%), los cuales se tienen cubiertos dentro de los modelos actuales de la empresa. El otro 34% representa los patrones que deben ser manejados de manera puntual y, según sea el caso, tener más inventario disponible a inicio o final de mes y, de esta manera, evitar faltantes.

Después de presentar los resultados al departamento de cadena de abastecimiento, se decidió utilizar esta

información para mejorar la operación logística. Actualmente, a la hora de realizar la distribución de la mercancía desde Yumbo y Bogotá, se cruzan los patrones de demanda para manejar prioridades y evitar faltantes al comienzo y final del mes.

5. Conclusiones

En el presente trabajo se expone la propuesta e implementación de un algoritmo para el reconocimiento de patrones de demanda en una empresa distribuidora de equipos de tecnología en

Colombia con el fin de mejorar la planeación a corto plazo de la logística de distribución. Este algoritmo hace el uso de la técnica de clustering, del aprendizaje automático para agrupar comportamiento de demanda similares y así inferir los patrones de demanda.

Luego de implementar en algoritmo en la empresa ACME, se reconocieron patrones de demanda anormales en el 34% de los productos que generan el 95% de los ingresos de la organización y con los cuales se puede mejorar considerablemente la disponibilidad de inventario y, por ende, aumentar las ventas.

Con esto, se concluye que el algoritmo es de gran utilidad para mejorar las operaciones logísticas de la empresa ACME y que se puede implementar en cualquier otra empresa que trabaje de manera similar. Por otra parte, como futuros proyectos se puede estudiar la integración del algoritmo con los sistemas ERP y de administración de inventarios de una empresa para recibir alertas en tiempo real y así disminuir los errores en la distribución.

6. Referencias

Cerreto, F. et al. (2018) "Application of Data Clustering to Railway Delay Pattern Reconition" Journal of Advanced Transportation. Volume 2018, Article ID 6164534, 18 pages

Chen, L. et al. (2016) "Dynamic Cluster-Based Over-Demand Prediction in Bike Sharing Systems" Ubicomp 2016, Septiembre 12 -16. Heilderberg, Alemania. DOI:

<http://dx.doi.org/10.1145/2971648.2971652>

Provost, F y Fawcett, T (2013) "Data science for Business" Sebastopol, Rusia. O'Reilly Media, Inc. pg 21.

Rousseeuw, P. (1987). "Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis". Computational and Applied Mathematics 20: 53-65.

Vidal, J. (2005) "Fundamentos de gestión de inventarios" Santiago de Cali, Colombia. Universidad del Valle. pg 44-46.