Aplicación de Cadenas de Markov como método de pronósticos de ventas para una empresa multinacional

P. Higareda-Ruiz*, R. Torres-Escobar,**
H.A. Pérez-Vicente, D.S. Reyna-Hernández*,**

**Facultad de Ingeniería, Universidad Anáhuac México Norte*
*** Centro de Alta Dirección en Ingeniería y Tecnología CADIT, Universidad Anáhuac México Norte*
Estado de México, MEX 52765
phigaredaruiz@gmail.com,danielsrh13@gmail.com, rafaeltorres.e@anahuac.mx, hugo.perez@anahuac.mx

Resumen: En este trabajo se presenta la aplicación de Cadenas de Markov para pronósticos de ventas en una empresa multinacional. El método usado es el propuesto por Liu, 2010, el cual se usa para analizar y predecir el precio y volumen de carne de vaca como una secuencia de datos. La información estudiada corresponde a las ventas mensuales de tres productos a partir del 2007 al 2015. Los datos presentan ventas nulas que deben representarse en los estados de la matriz de transición. Los resultados experimentales muestran la factibilidad del método en comparación con métodos clásicos de pronósticos siendo una opción viable y de fácil aplicación para la empresa bajo estudio.

Keywords: Cadenas de Markov, Pronósticos, Matriz de transición, Optimización, Modelación.

1. INTRODUCCIÓN

El mundo actual demanda una mayor precisión de los procesos que se llevan a cabo asegurando un menor riesgo, para lo cual todo tipo de empresas requiere tener conocimiento y hacer uso de modelos matemáticos para realizar pronósticos.

En la mayoría de las empresas, la predicción de las ventas es un problema que enfrentan con cierta frecuencia. Esta actividad casi siempre se reduce en la experiencia del negocio considerando las ventas de periodos pasados y estos pronósticos se modifican en cada periodo según el comportamiento del cliente para acercarse más a la realidad. Este método aun tradicional resulta fácil de aplicar pero puede no dar la certeza en muchos de los casos. En el mejor de los casos se usan métodos fiables en sus predicciones y en el que se pueden evaluar su desempeño. Estos métodos se pueden resumir en dos principales categorías: métodos cualitativos y métodos cuantitativos (Ghiani et al., 2004). En el primer caso están: la evaluación de la fuerza de ventas, investigación de mercados y el método Delphi. En el segundo caso se agrupan en: causales y series de tiempo, como regresión y promedio móvil, respectivamente (Ghiani et al., 2004). Como se menciona en (Chopra y Meindl, 2013) el objetivo de cualquier método de pronóstico es predecir el comportamiento sistemático de la demanda y estimar el componente aleatorio. No obstante, estos métodos si son usados en la práctica todavía continúan siendo considerados tradicionales desde nuestro punto de vista.

Las Cadenas de Markov se pueden ver como una serie de tiempo capaces de usar métodos potenciales para implementar la exactitud de pronóstico de ventas gracias al cálculo promedio de probabilidades de transición además de que su actualización no es brusca ni repentina. Lo impredecible de los análisis futuros puede hacer que los análisis de las Cadenas de Markov sean más útiles ya que pueden realizar un análisis sensible para la estimación a largo plazo el cual puede ser actualizado simplemente con el cambio de los datos de las probabilidades de transición.

2. ANTECEDENTES

Los pronósticos de ventas permiten que cierta empresa planee sus operaciones que favorezca atender todos los eslabones en su cadena de suministro de forma anticipada. Las series de tiempo han sido la base de cualquier estudio del comportamiento de un proceso o métrica observado en un periodo de tiempo. Estos han sido usados para diversos fines como predicción de ventas, del clima, estudios de inventarios, etc. (Desikan y Srivastava, 2014). En particular, en la predicción de ventas, los métodos encontrados en la literatura apuntan al uso de métodos sofisticados como redes neuronales artificiales (Omar et al., 2016) o algoritmos evolutivos (Arunraj y Ahrens, 2015) o su combinación (Chen y Kuo, 2015) y que, aunque podrían ser adecuados, son difíciles de implementar en la práctica sin un conocimiento vasto del área de conocimiento en particular.

En el estudio de Liu (2010), se muestra una alternativa para analizar y predecir series de tiempo con la aplicación de las cadenas de Markov. La idea surge debido a que algunas series de tiempo se pueden expresar como una cadena de Markov en tiempo discreto de primer orden y otras de orden superior mostrando efectividad y buen desempeño para los casos estudiados (Liu, 2010). En dicho estudio se explica detalladamente el método propuesto, sin embargo, en el presente trabajo se resume a continuación:

1) Proponer los estados de interés,

2) Construir la matriz de transición a partir de las frecuencias de cada estado normalizada,

3) Resolver el modelo matemático para predecir la serie de tiempo usando optimización lineal.

El presente trabajo consiste en la aplicación de cadenas de Markov de tiempo discreto de primer y segundo orden a un problema real para predecir las ventas de tres productos de una empresa trasnacional. Adicionalmente, se centra en dar un enfoque práctico dando énfasis a la efectividad del método incluso en comparación con algunos métodos cuantitativos tradicionales.

3. DESCRIPCIÓN DEL MÉTODO

En el trabajo se aplica la propuesta de Liu (2010), como un modelo de Cadenas de Markov para el análisis de series de tiempo.

Modelo de cadena de Markov de primer orden

Si se considera la modelación de una serie de tiempo x_t como una cadena de Markov de primer orden teniendo k estados $E=\{1,2,...,k\}$. Una cadena de Markov de primer orden en tiempo discreto con k estados que satisface la siguiente relación:

$$P(x_{t+1} = i_{t+1} | x_0 = i_0, x_1 = i_1, ..., x_t = i_t) = P(x_{t+1} = i_{t+1} | x_t = i_t),$$

donde x_t es el estado de una serie de tiempo en el tiempo t y i_j pertenece a E. Las probabilidades condicionales

$$P(x_{t+1} = i_{t+1} | x_t = i_t)$$

son llamadas probabilidades de transición de un paso de la Cadena de Markov. Estas probabilidades pueden ser escritas como $P(x_{t+1} = i | x_t = i_t) \forall i, j \in E$. La matriz $P = (p_{ij})_{kxk}$ es llamada matriz de probabilidad de transición de un paso. Se observa que los elementos de la matriz P satisfacen las siguientes dos propiedades:

$0 \leq p_{ij} \leq 1 \forall i, j \in E$ y $\sum_{i=1}^{k} p_{ij} = 1, \forall j \in E$. Luego, se construye un modelo de Cadena de Markov de primer orden para una serie de tiempo observada como $x_{t+1} = Px_t$. Dicho modelo está dado como sigue $x_{m+1} = \sum_{i=1}^{n} \lambda_i P_i x_{n-i+1}, i = m-1, m, \cdots$ con valores iniciales $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{n-1}$. Los pesos dados por λ_i son números reales no negativos tal que $\sum_{i=1}^{n} \lambda_i = 1$. Dado la serie de tiempo se cuenta la frecuencia de transición de los estados en la secuencia en el tiempo $t=m-i+1$ para los estados en la j-ésima secuencia en el tiempo $t=m+1$ para $1 \leq i \leq n$. Después de una normalización los estimados de las matrices de probabilidad de transición $\hat{P}_i$ son obtenidos. Se observa que el vector estacionario del vector x puede ser estimado de la proporción de ocurrencia de cada estado en cada secuencia (Ching et al., 2007). La estimación de los parámetros λ_i se resuelve planteando un problema de programación lineal que se detalla en (Liu, 2010). La evaluación del desempeño y la efectividad del modelo de cadena de Markov de orden superior se miden como resultado efectividad de la predicción definida por s que se muestra a continuación:

$$s = \frac{1}{N-n} \times \sum_{t=n+1}^{N} a_t \times 100\,\%$$

Donde N es la longitud de la secuencia de datos y

$$a_t = \begin{cases} 1, \hat{x}_t = x_t \\ 0. \end{cases}$$

Análisis de los datos

La información proviene de las ventas mensuales en libras de tres productos del departamento de una empresa trasnacional con sede en México que corresponde del 2007 al 2015, esto significa que se disponen de 108 datos por cada producto.

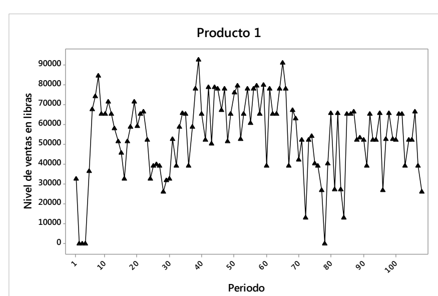

Fig. 1. Serie de tiempo para el Producto 1.

La serie de tiempo para el Producto 1 se puede observar en la Figura 1. Para esta serie se tuvo un valor mínimo de 0 y máximo de 92,300 con un coeficiente de variación del 37%. Para el Producto 2, en la Fig. 2 se observa la serie de tiempo correspondiente. Se obtuvo un valor entre 0 y 18,480, con coeficiente de variación del 145%. Es apreciable la frecuencia de ventas nula en varios periodos a diferencia de pocos casos para el producto anterior.

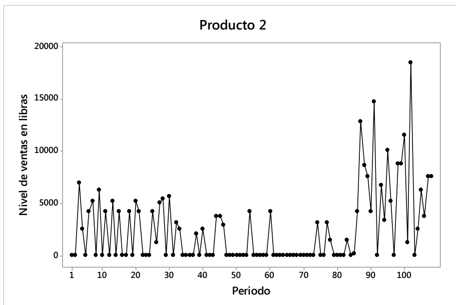

Fig. 2. Serie de tiempo para el Producto 2.

Finalmente, para el Producto 3 se tiene un valor mínimo de 0 y valor máximo de 218 con un coeficiente de variación del 76% (Fig. 3). Los tres productos muestran sus particularidades que pondrían a prueba el método aplicado.

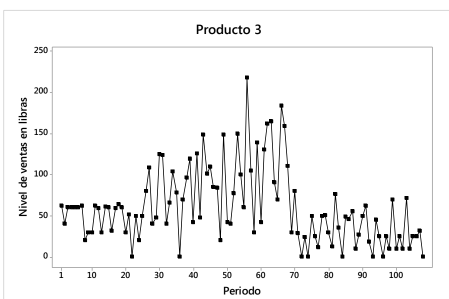

Fig. 3. Serie de tiempo para el Producto 3.

Definición de los estados de interés

Por conocimiento del negocio y por practicidad se propusieron dos casos: 1) Ventas bajas (Estado 1), Ventas medias (Estado 2) y Ventas altas (Estado 3) para los productos 1 y 3, 2) Ventas nulas (Estado 0), Ventas bajas (Estado 1), Ventas medias (Estado 2) y Ventas altas (Estado 3) para el Producto 2.

Una vez definidas los estados se capturan en una tabla de frecuencias los estados propuestos. Se establecieron distancias equidistantes para la acotación de estados de los productos. A excepción del Producto 2 que se fijó como valor mínimo 210 que es el valor más bajo de la demanda debido a que se tenían 58 casos con demanda nula. Los resultados se presentan en la Tabla 1.

Tabla 1. Tabla de frecuencias para los Estados de cada producto.

	Estado	Intervalo	Media del rango	Frecuencia
Producto 1	1	(0,30767]	15383	12
	2	(30767,61533]	46150	48
	3	(61533,92300]	76917	48
Producto 2	0	(0,210]	105	59
	1	(210,6300]	3255	36
	2	(6300,12390]	9345	10
	3	(12390,18480]	15435	3
Producto 3	1	(0,73]	36	78
	2	(73,145]	109	22
	3	(145,218]	182	8

Construcción de cadenas de Markov de primer y segundo orden

Con los datos obtenidos en la Tabla 2 se calcularon las probabilidades de transición entre cada estado para generar las cadenas de Markov de cada uno de los productos. Se registró el número de ocurrencias de un evento dado condicionado al evento anterior y posteriormente sus probabilidades de transición por cada producto (matriz de primer orden). Por ejemplo, para el Producto 1 se muestra el número de ocurrencias de cada estado en la Tabla 2 y su matriz de transición en la Tabla 3.

Tabla 2. Número de ocurrencias de cada estado para el Producto 1

Estado	**1**	**2**	**3**	Suma
1	4	5	2	11
2	5	25	18	48
3	3	17	28	48

Tabla 3. Matriz de transición para el Producto 1

Estado	**1**	**2**	**3**
1	0.3636	0.4545	0.1818
2	0.1042	0.5208	0.375
3	0.0625	0.3542	0.5833

Con esta misma lógica se construyeron las matrices de transición de primer orden y de segundo orden. Para este último, con la diferencia que ahora se tienen más estados debido a que se debe observar el número de ocurrencias de un evento dado que el evento anterior estaba precedido por otro evento.

Clasificación de los estados

La clasificación de los estados obtenidos nos da una mayor información de cómo se están comportando los datos y de los resultados que puede arrojar la matriz de transición al ser manipulada. Para esto se realizaron los diagramas de transición de cada matriz obtenida para observarlo de manera gráfica y así facilitar dicha clasificación. Para el Producto 1, se encontró que los tres estados son recurrentes y por tanto, se le considera una matriz ergódica, es decir, irreducible. El diagrama de transición de la cadena de primer orden se muestra en la Fig. 4.

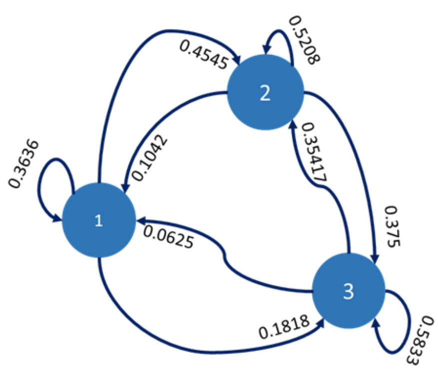

Fig. 4. Diagrama de transición de la cadena de primer orden para el Producto 1.

Para el Producto 2, la cadena de primer orden resultó ser una matriz irreducible y en la de segundo orden se encontró un estado transitorio. En el Producto 3 se tuvieron resultados iguales al Producto 2.

4. OBTENCIÓN DE RESULTADOS

4.1 Condición inicial

En primer lugar, se debe definir la condición inicial que servirá de referencia para determinar la situación actual, en qué punto se encuentra y a partir de ella obtener los resultados. Se tienen dos puntos de partida: el primero de ellos, tomará como condición inicial el promedio del último año. El segundo punto será más específico donde el último dato de la serie será la condición inicial.

En la Tabla 4 se muestra un ejemplo:

Tabla 4. Datos del último año para el Producto 1

Periodo	Valor	Escenario 1	Escenario 2
97	52650	2	2
98	65650	3	2
99	52650	2	2
100	52000	2	2
101	65000	3	2
102	65000	3	2
103	39000	2	1
104	52000	2	2
105	52000	2	2
106	66300	3	3
107	39000	2	1
108	26000	1	1
Promedio	**52270.83**	**2**	**2**

El Escenario 1 representa el análisis realizado con únicamente 3 estados correspondientes a la venta baja, media y alta. En el Escenario 2 se incluyó el estado 0 correspondiente a la venta nula. El promedio de datos del último año para el producto uno representa un valor que se encuentra dentro del estado 2 según los rangos obtenidos anteriormente. De aquí podemos decir que la condición inicial será partir de una venta media para ambos escenarios y para el caso del último dato sería partir de una venta baja como condición inicial ya que el valor del dato número 108 recae en el rango correspondiente al estado 1. Finalmente, este procedimiento se aplicó para la definición de las condiciones iniciales para cada uno de los productos.

4.2 Manipulación de la Cadena de Markov

En el trabajo se está interesado en encontrar los pronósticos de 5 a 10 años. Estos cálculos se obtienen con la siguiente fórmula: $a(n) = a(0) * p^n$ donde n es el año a pronosticar, $a(0)$ es la condición inicial y p es la matriz de transición. Por ejemplo, para el Producto 1 se obtuvo el siguiente resultado dado las dos condiciones iniciales propuestas:

Tabla 5. Pronóstico para el año 2 para el Producto 1 para las dos condiciones iniciales

Cond. Inicial	1	2	3
(0,1,0)	0.1156	0.4514	0.4330
(1,0,0)	0.1909	0.4664	0.3426

En la Tabla 5 se observan las dos condiciones iniciales propuestas. Los datos arrojados representan la probabilidad de tener una venta baja, media o alta respectivamente para los estados 1, 2 y 3 en este caso, partiendo de una venta media (condición inicial 1) o una venta baja (condición inicial 2). La probabilidad de obtener una venta media partiendo de esta misma es de 45.14% y de 43.30% de obtener una venta alta. Por otro lado, se obtiene una probabilidad de 46% de obtener una venta media partiendo de una venta baja para el segundo año y un 34% para una venta alta partiendo del mismo punto. En ambos casos la probabilidad de obtener una venta baja no es significativa. Se repitió el mismo esquema para las matrices de primer y segundo orden para cada escenario y para cada producto incluyendo el estado 0 para el Escenario 2, ver Tabla 6.

Tabla 6. Esquema de Pronóstico para el año 2 para el Producto 1 para las dos condiciones iniciales

	Matriz de transición de primer orden	
Matriz de transición de segundo orden	*Escenario 1*	*Escenario 2*
	Condición inicial 1: promedio del último año	
	Condición inicial 2: último dato de la serie	

4.3 Probabilidades de estado estable

Las probabilidades de transición son probabilidades límite de que el sistema se encuentre en estado j después de un número grande de transiciones, y que esta probabilidad sea independiente del estado inicial. Matemáticamente se puede expresar como $\pi_j = \pi P$ dónde $\pi = (\pi_0, \pi_0, ..., \pi_M)$. Las π_j se denominan probabilidades de estado estable de la Cadena de Markov (Liu, 2010). Una vez resuelto el sistema de ecuaciones se obtuvieron los valores de π para la matriz de primer y segundo orden por cada producto y cada escenario. En la Tabla 7 se observa como ejemplo los valores correspondientes a la matriz de primer orden considerando como condición inicial el promedio del último año para cada uno de los productos:

Tabla 7. Valores de π para matriz de transición de primer orden considerando la Condición inicial 1

		Producto 1	Producto 2	Producto 3
Escenario 1	π_1	**0.1156**	-	1.3974
	π_2	**0.4389**	-	4.7391
	π_3	**0.4455**	-	13.6250
Escenario	π_0	0.0374	0.5158	0.0777
	π_1	0.2150	0.3392	0.7090
	π_2	0.5047	0.1167	0.1307
	π_3	0.2430	0.0283	0.0826

La Tabla 7 indica que la probabilidad a largo plazo de tener una venta baja resulta de 11.56%, 43.89% para una venta media y para una venta alta es de 44.55%, considerando el Producto 1 del primer escenario.

5. DISCUSIÓN Y CONCLUSIONES

En la siguiente tabla se pueden observar los resultados que se obtuvieron para el porcentaje de precisión del modelo propuesto para la obtención de pronósticos de venta para los tres productos en los dos escenarios definidos, el primero excluyendo el estado 0 y el segundo incluyéndolo.

Primer producto	Primer orden	Segundo orden
3 estados	**54.21%**	**54.72%**
4 estados	**51.40%**	**51.89%**
5 estados	**33.64%**	**33.96%**
6 estados	**19.63%**	**19.81%**

Primer producto	Primer orden	Segundo orden
4 estados	**9.35%**	**22.64%**
5 estados	**3.74%**	**13.21%**
6 estados	**3.74%**	**9.43%**
7 estados	**3.74%**	**5.66%**

Fig. 6. Resultados de la evaluación de la cadena de Markov para el Producto 1.

Se puede distinguir la reducción en la precisión a medida que se agregan estados de transición a la matriz (rangos) debido a que al ir disminuyendo el rango de estos estados el resultado se vuelve más preciso. Para el Producto 3 se observó un comportamiento contrario para el segundo escenario con el aumento de precisión al ir aumentado la cantidad de estados dentro de la matriz de transición.

Existe un cambio importante en el porcentaje de precisión del modelo de escenario a escenario. Si se observa el primer escenario para el primer producto y se comparan los números con el segundo escenario, la diferencia entre resultados va de 44 a 15% al ir aumentando la cantidad de rangos. Para efectos prácticos se decide trabajar con el primer escenario para el primer producto al presentar mejores resultados en el nivel de acierto, esto se debe a la cantidad de venta nula que existe en la serie de tiempo la cual es mínima y por tal motivo no conviene trabajar los datos incluyendo un escenario únicamente para esa venta nula que se presenta en extrañas ocasiones. El resultado fue distinto con el Producto 2 que a pesar de que ambos resultados en ambos escenarios son muy buenos por la naturaleza de los datos se sabe que es necesario incluir un estado donde se concentre la venta nula para que el pronóstico no sea sesgado. El Producto 3 mostró un comportamiento similar al del primer producto. Por otro lado, se llevó a cabo el uso de modelos de pronóstico tradicionales como fueron promedio móvil con tres periodos, promedio móvil ponderado y nivelación exponencial. Se estimó el porcentaje de error medio absoluto (MAPE, por sus siglas en inglés) que variaron del 35% al 90%. Sin embargo, se usó lo pronosticado para el mes de enero y se comparó con el obtenido con la Cadena de Markov, los resultados se observan en la Tabla 8:

Tabla 8. Comparativo de resultados entre pronósticos tradicionales y usando Cadenas de Markov

Escenario 1	Métodos trad.	Cadena de Markov	Rangos	
Producto 1	54,088.43	2	30,766.00	615,533.33
Producto 2	2,471.39	-	6,160.00	123,200.00
Producto 3	59.89	1	-	72.67
Escenario 2				
Producto 1	54,088.43	2	30,766.00	615,533.33
Producto 2	2,471.39	2	6,160.00	123,200.00
Producto 3	59.89	1	10.00	79.33

Podemos concluir de acuerdo a la Tabla 8 que el uso de Cadenas de Markov acertó en 2 de 3 de los pronósticos obtenidos al compararlos con los resultados de modelos tradicionales contando con buenos porcentajes de precisión. No se puede comparar el resultado que se obtuvo para los siguientes periodos ya que los modelos tradicionales que se emplearon para el análisis solo entregan valores para el periodo inmediato siguiente desventaja que se observa al demostrar que cadenas de Markov nos arroja resultados para el periodo deseado e incluso resultados a largo plazo. Esto es una ventaja que se considera en el uso del modelo usado debido a que se reduce el error en el resultado.

5. TRABAJO FUTURO

Puede resultar interesante para futuras investigaciones, diseñar un experimento que permita ver el impacto de la discretización del espacio de estados, de esta manera se puede comprobar si la calidad de los pronósticos está vinculada con el uso de más estados de transición. Además, sería importante realizar un análisis con Cadenas de Markov Multivariantes como se propone en Ching et al. (2007).

REFERENCIAS

Arunjaj, N.S. and Ahrens, D. (2015). A hybrid seasonal autoregressive integrated moving average and quantile regression for daily food sales forecasting. *oInternation Journal of Productions Economics*, volume (170), Pp. 321-335.

Chen, Z.Y. and Kuo, R.J. (2015). Evolutionary Algorithm-Based Radial Basis Function Neural Network Training for Industrial Personal Computer Sales Forecasting. *Computational Intelligence*, DOI: 10.1111/coin.12073, Pp. 1467-8640.

Ching, W., Li, L.M., Li, T. and Zhang, S.Q. (2007). A new multivariate Markov Chain Model with applications to sales demand forecasting. *Proceedings of the International Conference on Industrial Engineering and Systems Management,Beijing may-30 –june 2.*

Chopra, S. and Meindl, P. (2013). *IAdministración de la cadena de suministro,* Pp. 183-184. Pearson Educación, México.

Ghiani, G., Laporte, G. and Musmanno, R. (2004). *Introduction to logistic systems planning and control.* Pp. 28-30. John Wiley and Sons, USA.

Liu, T. (2010). Application of Markov Chains to Analyze and Predict the Time Series. *Modern Applied Science*, volume (4), Pp. 162-166.

Omar, H., Hoang, V.H. and Liu, D.R. (2016). A Hybrid Neural Network Model for Sales Forecasting Based on ARIMA and Search Popularity of Article Titles. *Computational Intelligence and Neuroscience*, volume (2016), Pp. 1-9.