

Pronóstico de Ventas y Demanda de Componentes para Planta de Motocicletas MotoTec

Juan Andrés Muñoz, Paul Moreno, Enmanuel Velasquez Romero
Tecnología en Desarrollo de Software
Universidad Tecnológica de Pereira
Pereira, Risaralda, Colombia

Asignatura: Álgebra Lineal
Profesor: Daniel Alberto López
Fecha de entrega: 5 de junio de 2025

Abstract—Este informe presenta un prototipo analítico de pronóstico de ventas y demanda de componentes de motocicletas para la empresa MotoTec. Basado en datos históricos de ventas (2013–2022) y la metodología de ciencia de datos de John Rollins [4], se implementa un modelo de regresión lineal por mínimos cuadrados y su variante regularizada (Ridge). Se generan predicciones de ventas para 2024–2025 y se calcula la demanda de componentes necesarios para la fabricación de 4 tipos de motos.

I. INTRODUCCIÓN

MotoTec es un fabricante global con sede en Latinoamérica que ensambla cuatro familias de motocicletas—urbanas, turismo, off-road y eléctricas. La cadena de suministros se ha visto afectada en los últimos años por diversos factores (pandemia, logística marítima, escasez de chips), donde el 70 % de los componentes provienen de Asia. MotoTec quiere cambiar su modelo ERP para realizar sus compras, consolidar inventario y almacenar datos de producción, por un ecosistema digital más óptimo que anticipe cuellos de botella y optimice el flujo de piezas en tiempo real. Actualmente, la Gerencia de la compañía enfrenta el desafío de reducir un 25% el capital inmovilizado en inventarios sin afectar la disponibilidad de los productos en la línea de ensamblaje. Para reducir costos en la producción de las motocicletas, se implementará un modelo pronóstico para venta de motocicletas y compra de componentes para los años 2024–2025, basados en: capacidad instalada (horas-máquina), tiempo estándar de ensamblaje por modelo, datos históricos de ventas (2013–2022), matriz de componentes de los cuatro modelos de motocicletas, metodología de ciencia de datos de John Rollins-CRISP-DM [4], [7], modelo de regresión lineal por mínimos cuadrados y su variante regularizada (Ridge).

II. MARCO TEÓRICO DE ÁLGEBRA LINEAL

Para la realización de este modelo predictivo de venta de motocicletas y compra de componentes para los años 2024–2025, se describen conceptos básicos para la elaboración de este informe técnico.

A. Espacios Vectoriales

Es una estructura algebraica compuesta por un conjunto de vectores junto con dos operaciones: suma vectorial y multiplicación escalar, que cumplen ciertas propiedades (conmutatividad, asociatividad, existencia de neutro, inverso, etc.). Dado un conjunto V sobre un cuerpo F (usualmente \mathbb{R} o \mathbb{C}), decimos que V es un espacio vectorial si para todos $u, v \in V$ y $\alpha \in F$, se cumple: $u + v \in V$ y $\alpha v \in V$. Un concepto fundamental asociado es la base, que es un conjunto de vectores linealmente independientes que generan todo el espacio. El número de elementos en una base es la dimensión del espacio.

B. Transformaciones Lineales

Una transformación lineal $T : V \rightarrow W$ entre dos espacios vectoriales es una función que preserva la estructura lineal, es decir: $T(\alpha u + \beta v) = \alpha T(u) + \beta T(v)$. Toda transformación lineal puede representarse mediante una matriz A , tal que $T(x) = Ax$. Propiedades importantes incluyen el núcleo (o null space) y la imagen.

C. Descomposición en Valores Singulares (SVD)

Es una factorización de una matriz $A \in \mathbb{R}^{m \times n}$, como el producto de tres matrices: $A = U\Sigma V^T$, donde $U \in \mathbb{R}^{m \times m}$ es una matriz ortogonal cuyas columnas son los vectores propios de AA^T , $\Sigma \in \mathbb{R}^{m \times n}$ es una matriz diagonal con los valores singulares $\sigma_1 \geq \sigma_2 \geq \dots \geq 0$, y $V \in \mathbb{R}^{n \times n}$ es una matriz ortogonal cuyas columnas son los vectores propios de $A^T A$ [3]. La SVD es esencial para reducción de dimensionalidad, resolución de sistemas sobredeterminados y compresión de datos.

D. Pseudoinversa de Moore–Penrose

Se denota A^+ , es una generalización de la inversa de una matriz que existe incluso para matrices no cuadradas o no invertibles. Si $A \in \mathbb{R}^{m \times n}$, su pseudoinversa se define como $A^+ = V\Sigma^+U^T$, donde Σ^+ es la pseudoinversa de Σ . La pseudoinversa satisface las cuatro condiciones de Moore–Penrose. Su principal aplicación es encontrar soluciones de mínimos cuadrados al sistema $Ax = b$, mediante $x = A^+b$.

E. Método de Mínimos Cuadrados (Enfocado al Álgebra Lineal)

En muchos problemas prácticos, es común encontrarse con sistemas de ecuaciones lineales que no tienen solución exacta. El método de mínimos cuadrados permite encontrar una solución aproximada que minimiza el error. Sea $A \in \mathbb{R}^{m \times n}$ una matriz con $m \geq n$, y $b \in \mathbb{R}^m$ un vector columna. Queremos resolver el sistema $Ax = b$. Cuando no tiene solución exacta, se busca un vector $x \in \mathbb{R}^n$ tal que el residuo $r = b - Ax$ tenga norma mínima, es decir: $\min_{x \in \mathbb{R}^n} \|Ax - b\|^2$.

F. Regularización Ridge

Cuando los datos presentan colinealidad, alta dimensionalidad o ruido, el modelo de regresión lineal puede sobreajustarse. La regularización Ridge (o regresión de cresta o Tikhonov regularization [1]) introduce un término de penalización para reducir la varianza del modelo y evitar el sobreajuste. Dado un conjunto de datos $\{(x_i, y_i)\}_{i=1}^n$ con $x_i \in \mathbb{R}^p$, la regresión lineal clásica busca $\beta \in \mathbb{R}^p$ tal que $\hat{y}_i = x_i^T \beta$, minimizando el error cuadrático total $\min_{\beta} \|y - X\beta\|^2$. La regresión Ridge modifica la función de costo:

$$\min_{\beta} \{\|y - X\beta\|^2 + \lambda \|\beta\|^2\}$$

donde $\lambda \geq 0$ es el parámetro de regularización. Cuando $\lambda = 0$, se recupera la regresión lineal ordinaria. La solución analítica para β en Ridge es:

$$\beta_{\text{ridge}} = (X^T X + \lambda I)^{-1} X^T y$$

donde I es la matriz identidad. Geométricamente, Ridge impone una restricción esférica sobre los coeficientes β .

III. METODOLOGÍA

Se utiliza un enfoque de ciencia de datos siguiendo la metodología de John Rollins-CRISP-DM [4], [7], que comprende fases de definición, adquisición, preparación, modelado, evaluación y despliegue. Los datos consisten en un vector de ventas históricas por año y tipo de moto, y una matriz de componentes por modelo. La metodología CRISP-DM, según el documento, incluye 10 fases: Comprensión del negocio, Enfoque analítico, Requisitos de los datos, Recolección de datos, Comprensión de los datos, Preparación de los datos, Modelado, Evaluación, Despliegue y Retroalimentación.

A. Fase 1: Cargar datos y comprensión del problema

Esta fase prepara los datos e interpreta el problema industrial para su modelado matemático. **a. Carga de Datos:** Historial de Ventas históricas por tipo de moto (2013–2022), que proporciona la variable dependiente b . Matriz de Componentes por cada tipo de moto (10x4). **b. Comprensión del Problema:** La formulación matemática es $A \cdot x \approx b$. A : matriz de diseño, x : vector de coeficientes, b : vector de observaciones. Este planteamiento es ideal para regresión lineal por mínimos cuadrados.

	Tipo 1	Tipo 2	Tipo 3	Tipo 4
2013	172	89	18	28
2014	185	116	49	33
2015	202	155	98	49
2016	225	188	96	44
2017	252	200	148	59
2018	286	199	173	72
2019	316	240	204	70
2020	342	245	235	96
2021	371	280	266	140
2022	402	302	297	250

Fig. 1. Tabla de ventas históricas por tipo de moto (2013–2022).

B. Fase 2: Construcción de la matriz de diseño A y vector b

Esta fase es el puente entre el mundo real y el modelo matemático. Se construye la matriz de diseño A (variables independientes) y el vector de observaciones b (ventas históricas). Prepara los datos para aplicar mínimos cuadrados.

C. Fase 3A: Cálculo de la solución por mínimos cuadrados

Aquí se aplica el método de mínimos cuadrados clásicos para ajustar una recta de regresión lineal a los datos históricos de ventas de cada tipo de motocicleta. Permite proyectar ventas para 2024 y 2025. La solución de mínimos cuadrados se obtiene resolviendo $\beta = (A^T A)^{-1} A^T y$, y para Ridge $\beta = (A^T A + \lambda I)^{-1} A^T y$. El modelo entrega coeficientes a y b que definen la tendencia. Según este modelo y como se detalla en la Figura 2, las proyecciones indican que las motos más vendidas son la Tipo 1 (2024: 447.81, 2025: 474.35) y la menos vendida la Tipo 4 (2024: 205.16, 2025: 223.78).

	Tipo 1	Tipo 2	Tipo 3	Tipo 4
2024	447.806061	345.975758	357.890909	205.157576
2025	474.345455	368.218182	388.581818	223.781818

Fig. 2. Predicciones de ventas (unidades) para 2024 y 2025 por tipo de moto, modelo lineal individual.

D. Fase 4A: Proyección de ventas para 2024 y 2025 (y gráficas)

Componente visual y predictivo que aprovecha el modelo de regresión lineal para predecir ventas y generar gráficos comparativos. Esta fase traduce las matemáticas en decisiones visuales y estratégicas. Según las gráficas (Figuras 3–6), hubo mejor concordancia entre los datos históricos y las proyecciones para las motocicletas tipo 1 a 3. Sin embargo, el

comportamiento de la Tipo 4 experimentó una demanda o un éxito mucho mayor de lo previsto entre 2020-2022, lo que podría atribuirse a factores como un crecimiento inesperado del mercado, estrategias de marketing exitosas o factores externos como cambios en el comportamiento del consumidor.



Fig. 3. Ventas históricas y proyectadas - Tipo 1.

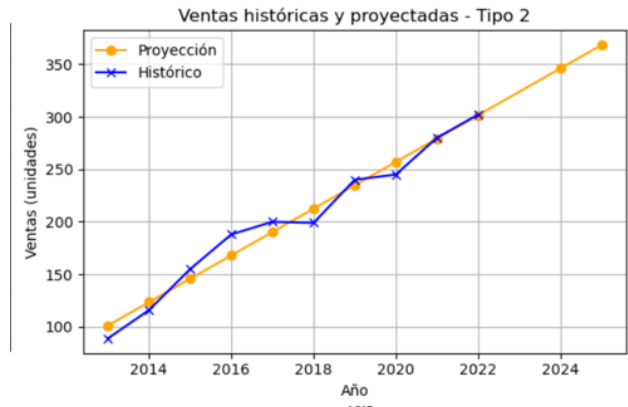


Fig. 4. Ventas históricas y proyectadas - Tipo 2.

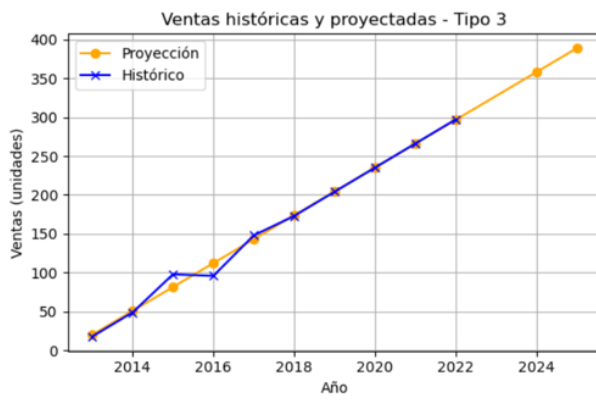


Fig. 5. Ventas históricas y proyectadas - Tipo 3.



Fig. 6. Ventas históricas y proyectadas - Tipo 4.

E. Fase 5A: Cálculo De Demanda de Componentes

Convierte las proyecciones de ventas en necesidad concreta de componentes. Aplica álgebra lineal al entorno industrial. Se traducen ventas proyectadas para 2024 y 2025 en cantidades de los 10 componentes para los 4 tipos de motocicletas. Cálculo: $d = C \cdot v$, donde C es la matriz de componentes y v el vector de ventas proyectadas. Esto permite planificar pedidos de forma optimizada.

	Componente	Demanda 2024	Demanda 2025
0	Comp 1	1151.672727	1231.145455
1	Comp 2	1458.660606	1561.054545
2	Comp 3	205.157576	223.781818
3	Comp 4	205.157576	223.781818
4	Comp 5	357.890909	388.581818
5	Comp 6	2035.369697	2159.472727
6	Comp 7	1831.709091	1947.218182
7	Comp 8	3136.139394	3331.945455
8	Comp 9	1509.563636	1619.727273
9	Comp 10	793.781818	842.563636

Fig. 7. Demanda proyectada de componentes por año (modelo lineal individual).

F. Fase 6A: Visualización De Resultados (Inventario)

Muestra de forma clara la demanda de componentes proyectada para 2024 y 2025. Transforma resultados numéricos

en gráficos comprensibles. Permite identificar componentes de mayor demanda y priorizar compras. El componente 8 presentará mayor demanda y los componentes 3 y 4 una tendencia menor.

G. Fase 7A: Análisis De Errores Y Documentación

Crítica para validar la calidad del modelo matemático usando el Error Cuadrático Medio (ECM). El análisis de errores valida cuantitativamente el modelo. Para la moto tipo 4 se presenta un ECM de 1174.7, que significa mal ajuste del modelo; se deben probar otros modelos (e.g. polinomio). Para los tipos 1, 2, 3, los ECM (59.4, 114.2, 57.0) no fueron uniformes pero se adaptan mejor a un modelo lineal.

ECM	
Tipo 1	59.409697
Tipo 2	114.155152
Tipo 3	56.901818
Tipo 4	1174.675152

Fig. 8. ECM de cada tipo de motocicleta para 2024-2025.

H. Solución Para El Margen De Error: Modelo Multivariable Global

Se propone un cambio a un modelo multivariable global que permite comprender mejor el comportamiento agregado. En vez de 4 regresiones independientes, se construye una sola regresión multivariable. Fundamento matemático: $A \cdot x \approx b$, donde A es una matriz $n \times 4$ (filas=años, columnas=ventas por tipo). Ejemplo de A (años 2013 a 2022):

TABLE I
EJEMPLO DE MATRIZ A PARA MODELO MULTIVARIABLE (ADAPTADO).

Año	Tipo 1	Tipo 2	Tipo 3	Tipo 4
2013	120	80	60	40
2014	130	85	63	42
...
2022	250	210	150	250

b es un vector $n \times 1$ (e.g., total de ventas por año). x es un vector de pesos 4×1 . Este modelo captura interdependencias, simplifica el análisis y es más robusto. Se implementa construyendo A y b y aplicando $x = (A^T A)^{-1} A^T b$. El modelo encuentra el hiperplano que mejor se ajusta. Cada componente de x es el peso de cada tipo de motocicleta sobre la variable total. Útil en estrategias globales.

I. Nueva Fase 3 (Modelo Multivariado): Enfoque alternativo – Ajuste global por año

Transformación a una perspectiva global y multidimensional. Aporta una visión más integral.

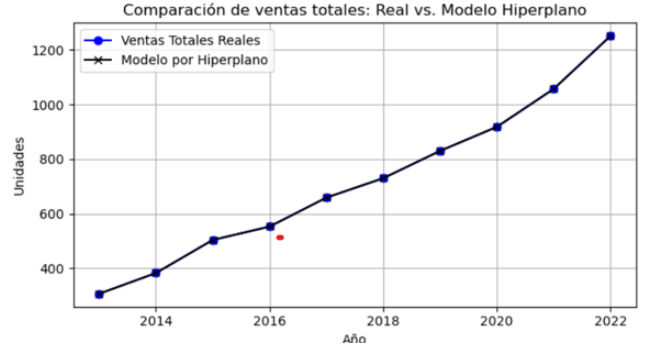


Fig. 9. Comparación de ventas totales: Real vs. Modelo Hiperplano.

J. Fase 3B (Modelo Multivariado): Proyección futura de ventas totales

Integra el enfoque individual y el global. Busca predecir ventas totales para 2024 y 2025 de forma más robusta. El vector x_{hiper} actúa como ponderación. Este método representa la sinergia entre la micro y macro perspectiva. Proyecciones:

$$v_{2024} = \begin{bmatrix} v_1^{2024} \\ v_2^{2024} \\ v_3^{2024} \\ v_4^{2024} \end{bmatrix}, \quad v_{2025} = \begin{bmatrix} v_1^{2025} \\ v_2^{2025} \\ v_3^{2025} \\ v_4^{2025} \end{bmatrix}, \quad x_{hiper} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix}.$$

$$Ventas_{totales\ 2024} = x_{hiper}^T \cdot v_{2024} \quad Ventas_{totales\ 2025} = x_{hiper}^T \cdot v_{2025}.$$

Mediante este enfoque, las ventas totales proyectadas para MotoTec son de 1,356 unidades en 2024 y 1,455 unidades en 2025.

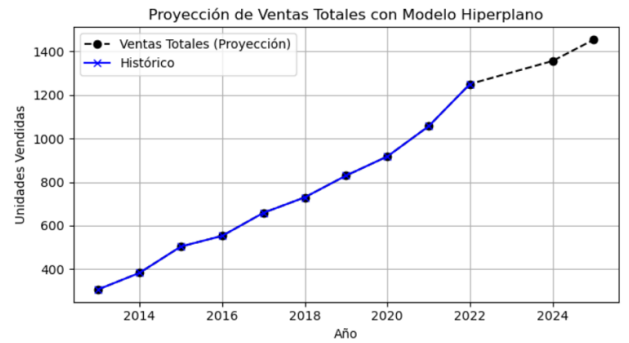


Fig. 10. Proyección de Ventas Totales con Modelo Hiperplano.

K. Fase 4B (Modelo Multivariado): Análisis de incertidumbre en las proyecciones

Clave para la validación estadística y toma de decisiones bajo riesgo. Se cuantifica la incertidumbre mediante simulación de Monte Carlo. Este es un método numérico que usa números aleatorios y repeticiones. Se parte de predicciones

puntuales, se añade error aleatorio simulado y se repite el proceso para generar un histograma de posibles resultados.

L. Fase 5B (Modelo Multivariado): Cálculo de demanda de componentes

Se observa la demanda de los 10 componentes.

	Componente	Demanda 2024	Demanda 2025
0	C1	1152	1231
1	C2	1459	1561
2	C3	205	224
3	C4	205	224
4	C5	358	389
5	C6	2035	2159
6	C7	1832	1947
7	C8	3136	3332
8	C9	1510	1620
9	C10	794	843

Fig. 11. Demanda de componentes 2024-2025 (Modelo Hiperplano).

La demanda de componentes es crucial en la transición a la aplicación industrial práctica. Convierte resultados del modelo en decisiones de planificación de inventario. Determinar con precisión cuántas unidades de cada componente se necesitarán. Este paso garantiza que la planta ordene cantidades adecuadas. El cálculo matricial es eficiente y replicable. Total proyectado de componentes (2025): 13530. Total proyectado de motos (2025): 1454.93. Promedio de componentes por moto (2025): 9.30. La validación asegura que los números tienen sentido técnico.

IV. CONCLUSIONES Y RECOMENDACIONES DE OPTIMIZACIÓN DE INVENTARIOS

Tras la implementación del modelo predictivo basado en regresión por hiperplano y su integración con la matriz técnica oficial de componentes, se logró construir un sistema robusto y coherente para estimar tanto la demanda total de motocicletas como los insumos requeridos para su producción. Las proyecciones para los años 2024 y 2025 indican un volumen estimado de ventas de 1,356 y 1,455 unidades respectivamente. Con base en estas proyecciones, se calculó la cantidad exacta de cada uno de los 10 componentes críticos, utilizando multiplicación matricial directa. La operación traduce efectivamente las estimaciones del modelo en un insumo logístico aplicable. En promedio, se requieren 9.35 componentes por motocicleta en 2024 y 9.30 en 2025. La ligera disminución del promedio para 2025 no representa un error, sino un ajuste natural en la mezcla de productos: el modelo proyecta una mayor proporción de ventas en motocicletas de menor complejidad técnica (como las de Tipo 4), que requieren menos insumos por unidad. Este cambio afecta positivamente la eficiencia logística. El proyecto avanza por las fases de evaluación,

despliegue y retroalimentación, permitiendo un sistema de pronóstico recurrente y adaptable.

****Recomendaciones de optimización de inventarios:****

- Utilizar las proyecciones de demanda de componentes para ajustar las políticas de compra y reducir el capital inmovilizado, con el objetivo de una reducción del 25%.
- Priorizar la gestión de componentes con mayor demanda (e.g., Componente 8).
- Considerar el cambio en la mezcla de productos (mayor proporción de Tipo 4) en la planificación logística y de almacenamiento.
- Para el Tipo 4, con alto ECM en el modelo lineal individual, explorar modelos alternativos (e.g., polinomiales).

V. ANEXO DE CÓDIGO (MATLAB Y PYTHON)

Ver repositorio de GitHub donde esta todo el proyecto:
<https://github.com/JuanTrader500/Proyecto-Algebra>

REFERENCES

- [1] IBM, "Ridge Regression", IBM Documentation.
- [2] Syntho, "Synthetic Data Generation", 2024.
- [3] Wikipedia, "Singular Value Decomposition".
- [4] J. Rollins, *Gestión de riesgos en TI: Cómo identificar, evaluar y gestionar riesgos en proyectos tecnológicos*, IT Governance Publishing, 2007.
- [5] Project Management Institute, *A Guide to the Project Management Body of Knowledge (PMBOK® Guide)*, 6th ed., Project Management Institute, 2017.
- [6] C. Chapman y S. Ward, *How to Manage Project Opportunity and Risk: Why Uncertainty Management can be a Much Better Approach than Risk Management*, 3rd ed., Wiley, 2011.
- [7] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer y R. Wirth, *CRISP-DM 1.0: Step-by-step data mining guide*, The CRISP-DM Consortium, 2000. [Online]. Available: <https://www.the-modeling-agency.com/crisp-dm.pdf>
- [8] D. C. Lay, S. R. Lay, y J. J. McDonald, *Álgebra lineal y sus aplicaciones*, 5.ª ed., México: Pearson, 2016.
- [9] T. Hastie, R. Tibshirani, y J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY, USA: Springer, 2009.
- [10] G. James, D. Witten, T. Hastie, y R. Tibshirani, *An Introduction to Statistical Learning*, 2nd ed. New York, NY, USA: Springer, 2021.
- [11] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*, Cambridge, MA: MIT Press, 2012.
- [12] J. J. Duderstadt and L. J. Hamilton, *Engineering Computation: An Introduction Using MATLAB and Excel*, Hoboken, NJ: Wiley, 2004.
- [13] C. M. Bishop, *Pattern Recognition and Machine Learning*, New York, NY: Springer, 2006.
- [14] M. S. Bazaraa, J. J. Jarvis, and H. D. Sherali, *Linear Programming and Network Flows*, 4th ed., Hoboken, NJ: Wiley, 2010.
- [15] NumPy Developers, "NumPy: Fundamental package for scientific computing with Python," [Online]. Available: <https://numpy.org/>
- [16] pandas Development Team, "pandas: Python Data Analysis Library," [Online]. Available: <https://pandas.pydata.org/>
- [17] Matplotlib Developers, "Matplotlib: Visualization with Python," [Online]. Available: <https://matplotlib.org/>