

 UNIVERSIDAD ECC I	SISTEMAS AVANZADOS DE PRODUCCIÓN Y PYTHON		Código: IN-IN-001 Versión:01
Taller 1	Fecha de emisión: 22-Nov-2009	Fecha de versión: 22-Nov-2009	

TALLER N° 1

**JESSICA ANDREA PARRA RIAÑO
JUAN PABLO VALERO BUITRAGO**

UNIVERSIDAD ECCI
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA INDUSTRIAL
BOGOTÁ, D.C.
AÑO 2026

 <p>UNIVERSIDAD ECCI</p>	SISTEMAS AVANZADOS DE PRODUCCIÓN Y PYTHON		Código: IN-IN-001 Versión:01
	Taller 1	Fecha de emisión: 22-Nov-2009	Fecha de versión: 22-Nov-2009

TALLER N° 1

**JESSICA ANDREA PARRA RIAÑO
JUAN PABLO VALERO BUITRAGO**

Taller N° 1

**FREDY ALEXANDER ORJUELA LOPEZ
Docente de Sistemas Avanzados de Producción y Python**

**UNIVERSIDAD ECCI
 FACULTAD DE INGENIERÍA
 PROGRAMA DE INGENIERÍA INDUSTRIAL
 BOGOTÁ, D.C.
 AÑO 2026**

 <p>UNIVERSIDAD ECC</p>	SISTEMAS AVANZADOS DE PRODUCCIÓN Y PYTHON		Código: IN-IN-001 Versión:01
	Taller 1	Fecha de emisión: 22-Nov-2009	Fecha de versión: 22-Nov-2009

DESCRIPCIÓN DEL PROBLEMA

Realizar un análisis de la relación entre la inversión en publicidad en diferentes medios y su influencia en el comportamiento de las ventas, esto mediante un proceso de modelamiento estadístico de regresión lineal simple, uno múltiple y por último con un modelo de árboles de decisión objetivo de comparar enfoques.

OBJETIVO DEL ANÁLISIS

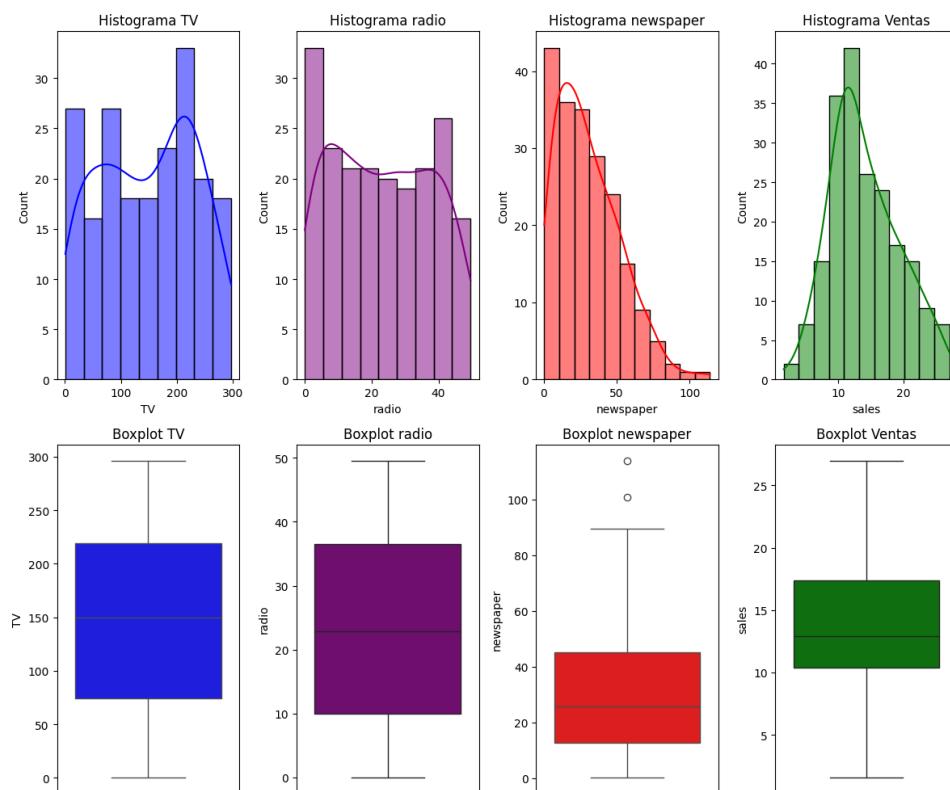
Analizar la relación entre la inversión publicitaria y el comportamiento de las ventas mediante el uso de herramientas estadísticas y modelos predictivos, con el fin de evaluar su capacidad explicativa y determinar el enfoque más adecuado para apoyar la toma de decisiones en un contexto empresarial.

Fase 1: Estadística descriptiva y análisis exploratorio

Variable	Mean	Median	Std	Min	Max	Skewness	Kurtosis
TV	147.0425	149.75	85.854236	0.7	296.4	-0.069853	-1.226495
Radio	23.2640	22.90	14.846809	0.0	49.6	0.094175	-1.260401
Newspaper	30.5540	25.75	21.778621	0.3	114.0	0.894720	0.649502
Sales	14.0225	12.90	5.217457	1.6	27.0	0.407571	-0.408869

Tabla (1): Medidas de tendencia central y dispersión de las variables independientes y la variable dependiente(SALES)

Análisis de distribución y atípicos



Figura(1): Las imágenes superiores de la gráfica representan los histogramas de las variables, y las imágenes inferiores corresponden a los diagramas de cajas y bigotes de cada variable.

 <p>UNIVERSIDAD ECCI</p>	SISTEMAS AVANZADOS DE PRODUCCIÓN Y PYTHON		Código: IN-IN-001 Versión:01
	Taller 1	Fecha de emisión: 22-Nov-2009	Fecha de versión: 22-Nov-2009

Determine la simetría de las distribuciones.

HISTOGRAMAS

TV

Observamos una distribución relativamente uniforme a lo largo del rango de inversión. No se aprecia una concentración marcada en valores extremos ni una cola pronunciada hacia alguno de los lados. La distribución es aproximadamente simétrica.

RADIO

La variable radio presenta un comportamiento bastante homogéneo en su rango de valores. No se identifican colas largas ni acumulaciones extremas. La curva KDE mantiene una forma relativamente balanceada. La distribución es aproximadamente simétrica.

NEWSPAPER

En este caso se observa una mayor concentración de datos en valores bajos y una disminución progresiva desde los valores altos. La curva KDE muestra una cola extendida hacia la derecha. Tiene una asimetría positiva (sesgo a la derecha).

VENTAS

La variable ventas presenta una forma más cercana a una distribución normal, aunque con ligera asimetría positiva. La mayor concentración de datos se ubica en valores medios, con una extensión moderada hacia valores altos. Distribución aproximadamente normal con leve sesgo positivo.

BOXPLOTS (DIAGRAMAS DE CAJAS Y BIGOTES)

TV

- **Mediana:** La mediana se encuentra prácticamente en el centro de la caja, lo que sugiere que no existe un sesgo marcado hacia valores altos o bajos.
- **Rango Intercuartílico:** El 50% central de los datos está distribuido de manera equilibrada dentro de la caja, mostrando una dispersión amplia pero uniforme en los niveles de inversión.
- **Ausencia de Outliers:** No se observan puntos aislados fuera de los bigotes. Esto es favorable para un modelo de regresión, ya que indica que no hay valores extremos que puedan distorsionar la pendiente de la recta estimada.

RADIO

- **Mediana:** La mediana está ubicada aproximadamente en el centro de la caja, lo que indica ausencia de asimetría significativa.
- **Rango Intercuartílico:** El 50% central de los datos se concentra dentro de un rango consistente, mostrando estabilidad en la variabilidad de la inversión en radio.
- **Ausencia de Outliers:** No se identifican valores atípicos. Esto contribuye a que cualquier modelo lineal estimado sea más estable y representativo del comportamiento general.

 <p>UNIVERSIDAD ECCI</p>	SISTEMAS AVANZADOS DE PRODUCCIÓN Y PYTHON		Código: IN-IN-001 Versión:01
	Taller 1	Fecha de emisión: 22-Nov-2009	Fecha de versión: 22-Nov-2009

NEWSPAPER

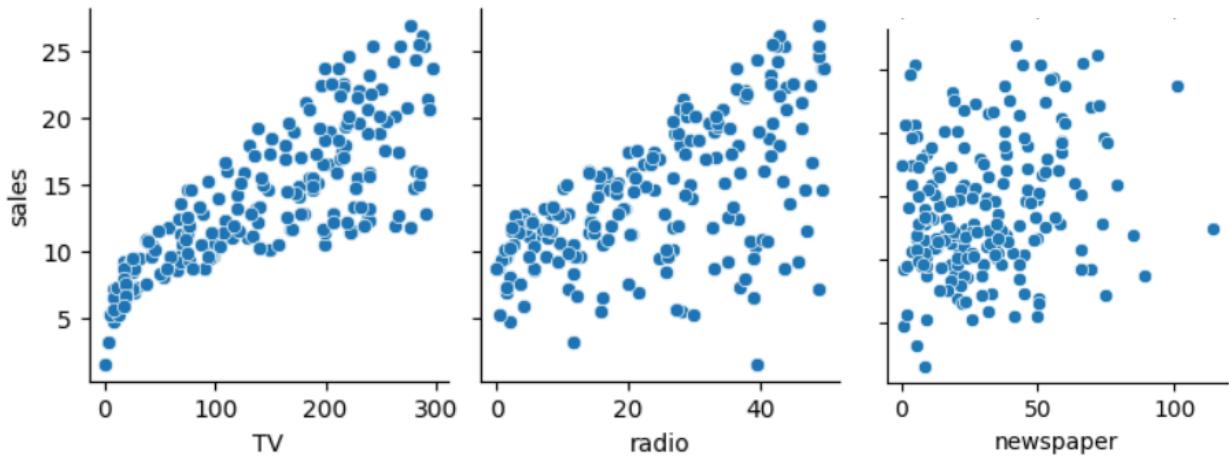
- **Mediana:** La mediana se encuentra desplazada hacia la parte inferior de la caja, lo que sugiere un sesgo hacia valores bajos.
- **Rango Intercuartílico:** El 50% central de los datos está concentrado en valores relativamente bajos, mientras que la dispersión hacia la parte superior es mayor.
- **Ausencia de Outliers:** Se observan puntos aislados por encima del bigote superior. Estos valores atípicos pueden afectar un modelo de regresión lineal, ya que podrían “jalar” la línea de ajuste hacia arriba, incrementando la pendiente o aumentando el error estándar de los coeficientes.

VENTAS

- **Mediana:** La mediana está ligeramente por debajo del centro de la caja, lo que indica una leve asimetría positiva, aunque no extrema.
- **Rango Intercuartílico:** El 50% central de los datos se encuentra relativamente concentrado, mostrando una dispersión moderada y estable.
- **Ausencia de Outliers:** No se observan valores atípicos significativos. Esto es positivo para el análisis de regresión, ya que garantiza que la estimación de la línea de ajuste no estará influenciada por observaciones extremas.
- Responda: ¿Se identifican datos atípicos o observaciones que se alejan significativamente de la masa de los datos? Explique cómo podrían influir estas observaciones en la trayectoria de una línea de regresión.

Se observan datos atípicos en la relación a la variable Newspaper esta podría generar una afectación de la pendiente del modelo de regresión el error estándar de los coeficientes, sin embargo las demás variables al tener distribuciones simétricas fortalecen y estabilizan el modelo de regresión.

Exploración de la Nube de Puntos



Figura(2): Exploración de la Nube de Puntos entre las variables TV, Radio, Newspaper contra la variable ventas.

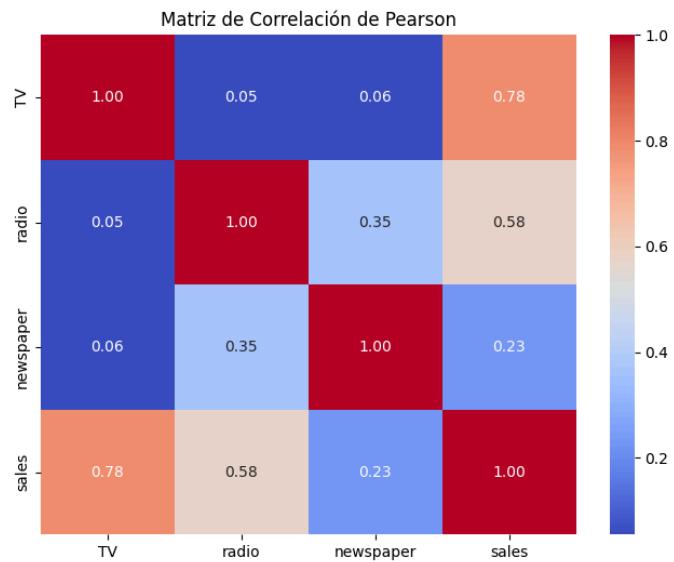
La nube de puntos que representa la relación entre las variables TV y ventas muestran un patrón ascendente positivo aproximadamente lineal donde a medida que aumenta la inversión en publicidad televisiva, las ventas tienden a incrementarse.

En el caso de la variable radio y ventas el patrón es ascendente y positivo pero la dispersión es mayor pese a que no se evidencia algún indicio de comportamiento parabólico esta condición hace que el modelo de regresión lineal se pueda ver afectado por dichos datos.

Para la relación entre Newspaper y ventas no se observa una relación lineal clara. Los puntos están bastante dispersos y no siguen una tendencia definida. No se aprecia una forma parabólica ni segmentos diferenciados, pero sí mucha variabilidad.

 UNIVERSIDAD ECC C I ECC	SISTEMAS AVANZADOS DE PRODUCCIÓN Y PYTHON	Código: IN-IN-001 Versión:01
Taller 1	Fecha de emisión: 22-Nov-2009	Fecha de versión: 22-Nov-2009

Matrices de correlación



Figura(3): Matriz de correlación de pearson de las variables

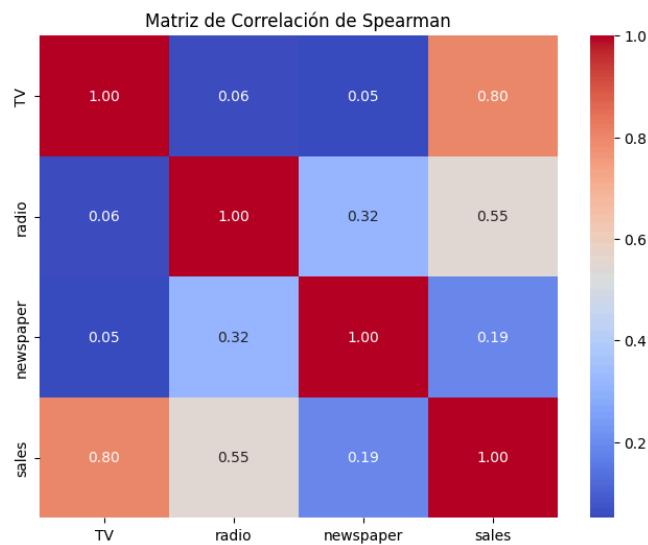


Figura (4) : Matriz de correlación de Spearman de las variables

 <p>UNIVERSIDAD ECCI</p>	SISTEMAS AVANZADOS DE PRODUCCIÓN Y PYTHON		Código: IN-IN-001 Versión:01
	Taller 1	Fecha de emisión: 22-Nov-2009	Fecha de versión: 22-Nov-2009

- **¿Qué significa un coeficiente cercano a 1 o -1 en el contexto de la inversión publicitaria y el retorno en ventas?**

Para el coeficiente cercano a 1 indica que a mayor inversión publicitaria tendrá como consecuencia un aumento de las ventas y existe una mayor relación entre las variables implicadas.

Mientras que para un coeficiente cercano a -1 indica que a mayor inversión publicitaria tendría una disminución en las ventas y una correlación negativa de las variables.

- **¿Cómo se interpreta un valor de correlación cercano a 0?**

Se interpreta como una relación de independencia entre las variables que no tendrá ninguna afectación de las ventas el invertir o no en publicidad de cualquier tipo.

- **Compare ambos coeficientes (Pearson y Spearman). ¿Sugieren estos resultados que las relaciones son estrictamente lineales? Justifique su respuesta basándose en la forma de nube de puntos observada.**

Para este caso sugiere que son mayormente lineales ya que la diferencia entre ambos coeficientes son similares para TV ($\approx 0.78-0.80$), radio ($\approx 0.55-0.58$) y newspaper ($\approx 0.19-0.23$) pero no se puede afirmar que la relación sea estrictamente lineal porque, aunque los coeficientes de Pearson y Spearman son altos y muy similares, no son iguales a 1 ni coinciden exactamente entre sí.

Fase 2: Regresión Lineal y Diagnóstico Modelamiento Múltiple

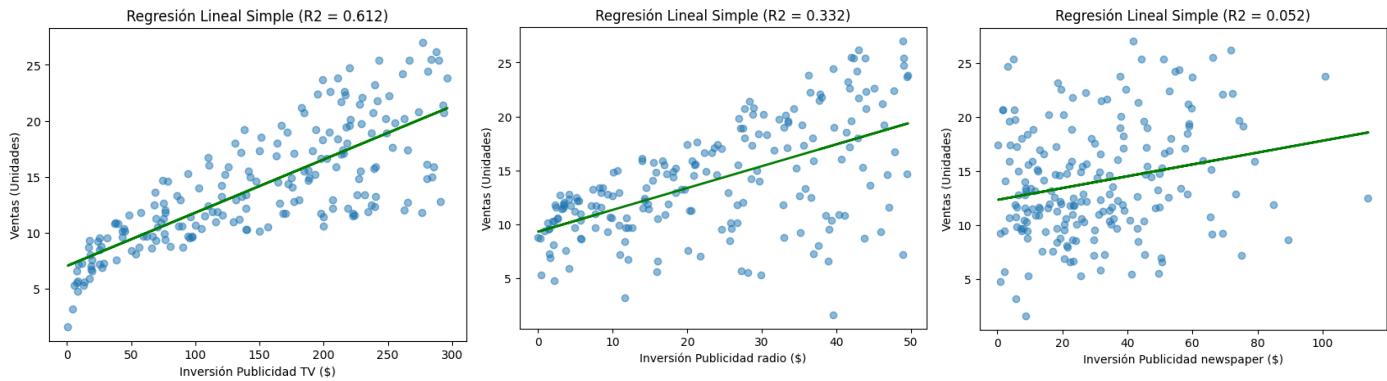


Figura (5) : Gráficos de regresión lineal de cada variable

Figura: Plano de Regresión Múltiple

● Observaciones

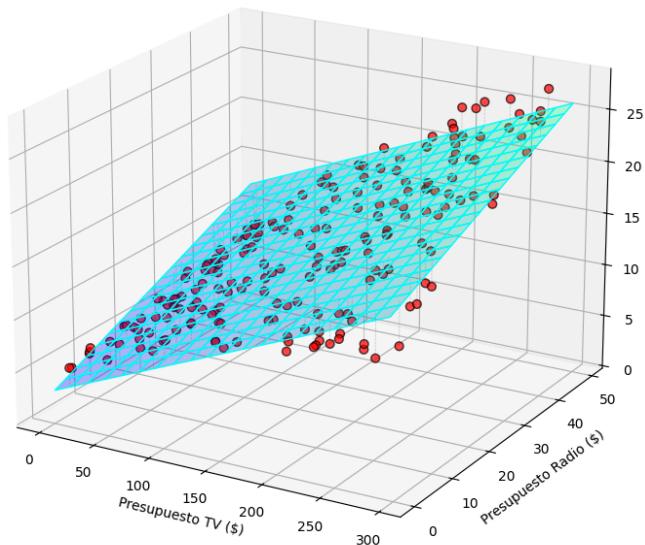


Figura (6) : Gráfico de regresión múltiple

 <p>UNIVERSIDAD ECC</p>	SISTEMAS AVANZADOS DE PRODUCCIÓN Y PYTHON		Código: IN-IN-001 Versión:01
	Taller 1	Fecha de emisión: 22-Nov-2009	Fecha de versión: 22-Nov-2009

Parámetro	Símbolo	Valor
Intercepto	β_0	2.9389
Coeficiente TV	β_1	0.0458
Coeficiente Radio	β_2	0.1885
Coeficiente Newspaper	β_3	-0.0010
Coeficiente de Determinación	R^2	0.8972

Tabla (2): Reporte los valores numéricos del intercepto y los coeficientes de pendiente.

Interpretación de Parámetros

- Interprete el intercepto β_0 indicando si tiene sentido conceptual que sea diferente de cero

Tiene sentido ya que las ventas no dependen exclusivamente de la inversión en publicidad en cualquier medio también existen factores como la demanda natural del producto, el posicionamiento de la marca y otras variables que no se tiene en cuenta en este modelo y este valor es necesario para ajustar correctamente la recta de regresión.

- Interprete los coeficientes β_1 , β_2 y β_3 bajo el concepto de variación de la media de Y cuando hay un incremento unitario de la covariable, manteniendo la otra constante (*ceteris paribus*).

Para el β_1 si la inversión en TV aumenta en una unidad, y la inversión en Radio y Newspaper se mantiene constante, las ventas aumentan en promedio 0.0458 unidades. Esto significa que la publicidad en televisión tiene un efecto positivo sobre las ventas, aunque su impacto por unidad invertida es moderado.

Para el β_2 si la inversión en Radio aumenta en una unidad, manteniendo constantes TV y Newspaper, las ventas aumentan en promedio 0.1885 unidades. Este es el mayor efecto entre las tres variables, lo que indica que la radio genera el mayor incremento promedio en ventas por unidad adicional invertida.

Para el β_3 si la inversión en Newspaper aumenta en una unidad, manteniendo constantes TV y Radio, las ventas disminuyen en promedio 0.001 unidades. Este efecto es prácticamente cero, lo que indica que la inversión en periódico no tiene un impacto relevante sobre las ventas dentro del modelo.

Estimación Matricial:

$$X = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{pmatrix} \quad Y = \begin{pmatrix} 2 \\ 4 \\ 6 \\ 7 \\ 9 \end{pmatrix}$$

$$X^T = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix}$$

$$X^T \cdot X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{pmatrix} = \begin{pmatrix} 5 & 15 \\ 15 & 55 \end{pmatrix}$$

$$(X^T \cdot X)^{-1} = \det(X^T \cdot X) = (5 \cdot 55) - (15 \cdot 15) = 50$$

$$(X^T \cdot X)^{-1} = \frac{1}{50} \begin{pmatrix} 55 & -15 \\ -15 & 5 \end{pmatrix}$$

$$(X^T \cdot X)^{-1} = \begin{pmatrix} \frac{55}{50} & -\frac{15}{50} \\ -\frac{15}{50} & \frac{5}{50} \end{pmatrix}$$

$$(X^T \cdot X)^{-1} = \begin{pmatrix} \frac{11}{10} & -\frac{3}{10} \\ -\frac{3}{10} & \frac{1}{10} \end{pmatrix}$$

$$X^T \cdot Y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix} \cdot \begin{pmatrix} 2 \\ 4 \\ 6 \\ 7 \\ 9 \end{pmatrix} = \begin{pmatrix} 28 \\ 101 \end{pmatrix}$$

Fase 3: Árboles de Decisión y Comparación de Modelos

Entrenamiento del Modelo:

Estructura del Árbol de Regresión: Ventas vs (TV y Newspaper)

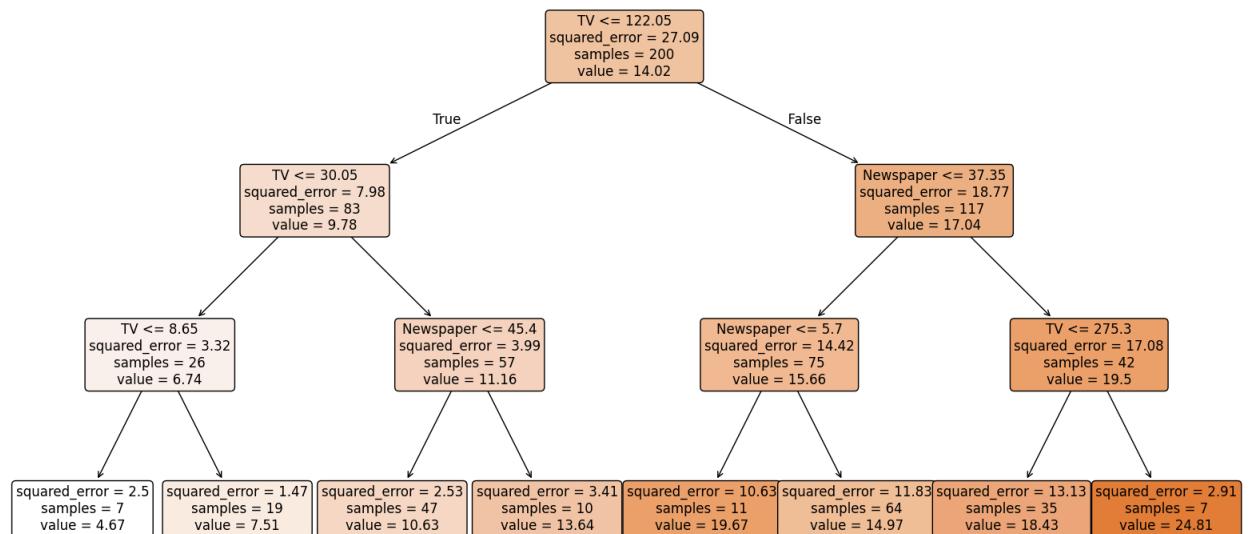


Figura (7) : árbol de regresión para predecir las ventas

 <p>UNIVERSIDAD ECC</p>	SISTEMAS AVANZADOS DE PRODUCCIÓN Y PYTHON		Código: IN-IN-001 Versión:01
	Taller 1	Fecha de emisión: 22-Nov-2009	Fecha de versión: 22-Nov-2009

Cálculo Manual de Incertidumbre (Clasificación):

Mercado	Inversión TV	Ventas (Clase)
1	Alta	Altas
2	Alta	Altas
3	Baja	Bajas
4	Baja	Altas

Tabla (3): Valores de referencia.

- **Calcule la Entropía del Nodo Padre:**

$$H(S)=-(0.75\log_2(0.75)+0.25\log_2(0.25))=0.8113$$

- **Calcule el Índice de Gini para el nodo padre:**

$$Gini(S)=1-(0.752+0.252)=0.375$$

Importancia de Predictores:

Variable	Importancia en el Árbol	Coeficiente de Regresión
TV	0.8241	0.469
Newspaper	0.1759	0.0442

Diagnóstico Comparativo:

Si se busca optimizar recursos y medir el impacto de cada variable es más conveniente usar la regresión lineal ya que el modelo nos ofrece datos del impacto de las ventas de cada variable lo que permite determinar una planificación y un control adecuado para los procesos administrativos, comerciales y gerenciales.

Para un caso donde se busque segmentar el mercado o los clientes el uso del árbol de decisión es más acertado ya que permite diseñar mejores estrategias, designar presupuestos en los que se prioricen los segmentos más rentables para la empresa.

 UNIVERSIDAD ECC I	SISTEMAS AVANZADOS DE PRODUCCIÓN Y PYTHON		Código: IN-IN-001 Versión:01
Taller 1	Fecha de emisión: 22-Nov-2009	Fecha de versión: 22-Nov-2009	

CONCLUSIÓN

Mediante el análisis de los resultados se identifica que para este caso particular es mucho más robusto el modelo de regresión lineal múltiple ya que este muestra un valor a cada variable el cual permite medir su efecto en las ventas lo cual ayuda a la toma de decisiones para asignar recursos económicos en las opciones que sean más rentables para una compañía.