

## Resumen:

La minería de datos y el análisis de regresión lineal ofrecen una oportunidad valiosa para descubrir patrones significativos en la delincuencia, incluso cuando parecen erráticos. Este estudio emplea un modelo de regresión lineal para prever tendencias delictivas a largo plazo en Bangladesh. Se utilizaron datos reales obtenidos del portal oficial de la policía de Bangladesh, seleccionados meticulosamente para entrenar y optimizar el modelo. Posteriormente, se realizó una evaluación proyectiva abarcando una amplia gama de infracciones y delitos. El propósito es proporcionar a las autoridades policiales una herramienta efectiva basada en regresión lineal para anticipar y prevenir futuros delitos. Al ofrecer pronósticos fundamentados y respaldados por datos concretos, se espera contribuir significativamente a la seguridad y protección en el país.

## Introducción:

La delincuencia es un problema social que perturba el bienestar público y el desarrollo socioeconómico. Por esta razón, en el siguiente estudio se aplicó la regresión lineal para predecir diversos tipos de delitos, utilizando conjuntos de datos recopilados del sitio web oficial de la policía de Bangladesh. Al examinar varios tipos de crímenes, como robo, asesinato y secuestro, la investigación tuvo como objetivo pronosticar las tendencias futuras de la delincuencia en las regiones metropolitanas y divisionales del país. Este enfoque permite identificar patrones y variaciones en la ocurrencia de delitos, lo que es crucial para la planificación y ejecución de estrategias de prevención. Además, el estudio busca favorecer la reducción y resolución eficiente de transgresiones, contribuyendo a una mejora significativa en la seguridad pública y, en última instancia, al bienestar general de la sociedad.

## UTILIZACIÓN DE LA REGRESIÓN LINEAL PARA PREDECIR LAS TENDENCIAS FUTURAS DE LA DELINCUENCIA EN BANGLADESH

Equipo 6:	Diana Valeria Ovalle Valdés	Emmanuel Sánchez Aranda
Leonardo Salvador Diego Solis	Emmanuel Alejandro Chavarría Buendía	Jorge Alberto Morales Reyes
	Mauricio De La Fuente Herrera	José Francisco Jasso González

### Metodología:

La minería de datos es algo indispensable en la disciplina de las bases de datos, lo cual consiste en recoger datos en bruto para luego extraer información que sirva para hacer predicciones. En esta investigación, la minería de datos fue usada para pronosticar la cantidad de crímenes en Bangladesh. Esta metodología se divide en dos partes: descripción del conjunto de datos y el modelo de selección.

**A. Descripción de conjunto de datos**

Para esta investigación se recopilaron datos reales sobre los delitos ocurridos en años anteriores en Bangladesh, esto del sitio web oficial de la policía del país ya mencionado. Los datos se dividen en 2 grupos según la región de Bangladesh: de la región metropolitana y de la región divisional. El conjunto de datos consta de 840 casos en todo el país, cada dato con 3 características predictivas (región, mes y año) y una característica objetivo (el valor de los diferentes tipos de delito).

**B. Modelo de selección**

Existen diferentes técnicas de minería de datos para predecir cuantitativamente la situación de la delincuencia, sin embargo, se seleccionó la regresión lineal para este artículo. Esta técnica consiste en predecir una variable objetivo Y como una función lineal de otra variable, que es de entrada, llamada X.

La hipótesis de la regresión lineal se puede expresar de la siguiente forma:

$$h_{\theta}(x) = \theta_0 + \theta_1x_1 + \theta_2x_2 + \dots + \theta_nx_n$$
$$h_{\theta} = \theta^Tx$$

Donde  $\theta_0, \theta_1, \theta_2, \dots, \theta_n$  son parámetros de regresión. Para encontrar dichos parámetros, se utiliza la siguiente función de coste:

$$J(\theta_0, \theta_1, \theta_2, \dots, \theta_n) = \frac{1}{2m} \sum_{i=1}^m \left( h_{\theta} \left( x^{(i)} \right) - y^{(i)} \right)^2$$

El objetivo de la regresión lineal es minimizar el costo de la función  $J(\theta_0, \theta_1, \theta_2, \dots, \theta_n)$  encontrando así  $\theta_0, \theta_1, \theta_2, \dots, \theta_n$  de modo que  $h_{\theta}(x)$  esté cerca de Y para cada ejemplo de (x,y). Una forma de minimizar el costo de la función es utilizando el algoritmo de descenso de gradiente por lotes. El algoritmo es el siguiente:

$$\text{Repeat until convergence} \{ \theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^m \left( h_{\theta} \left( x^{(i)} \right) - y^{(i)} \right) x_j^{(i)} \text{ (for every } j \text{)} \}$$

Donde  $\sigma$  es la tasa de aprendizaje. Los parámetros  $\theta_0, \theta_1, \theta_2, \dots, \theta_n$  se acercan a los valores óptimos que alcanzarán en el menor coste  $J(\theta_0, \theta_1, \theta_2, \dots, \theta_n)$  conforme al descenso de gradiente. Cada característica se normaliza en el rango [-1,1] para acelerar el proceso de minimización, y para normalizar las características se usa la siguiente ecuación:

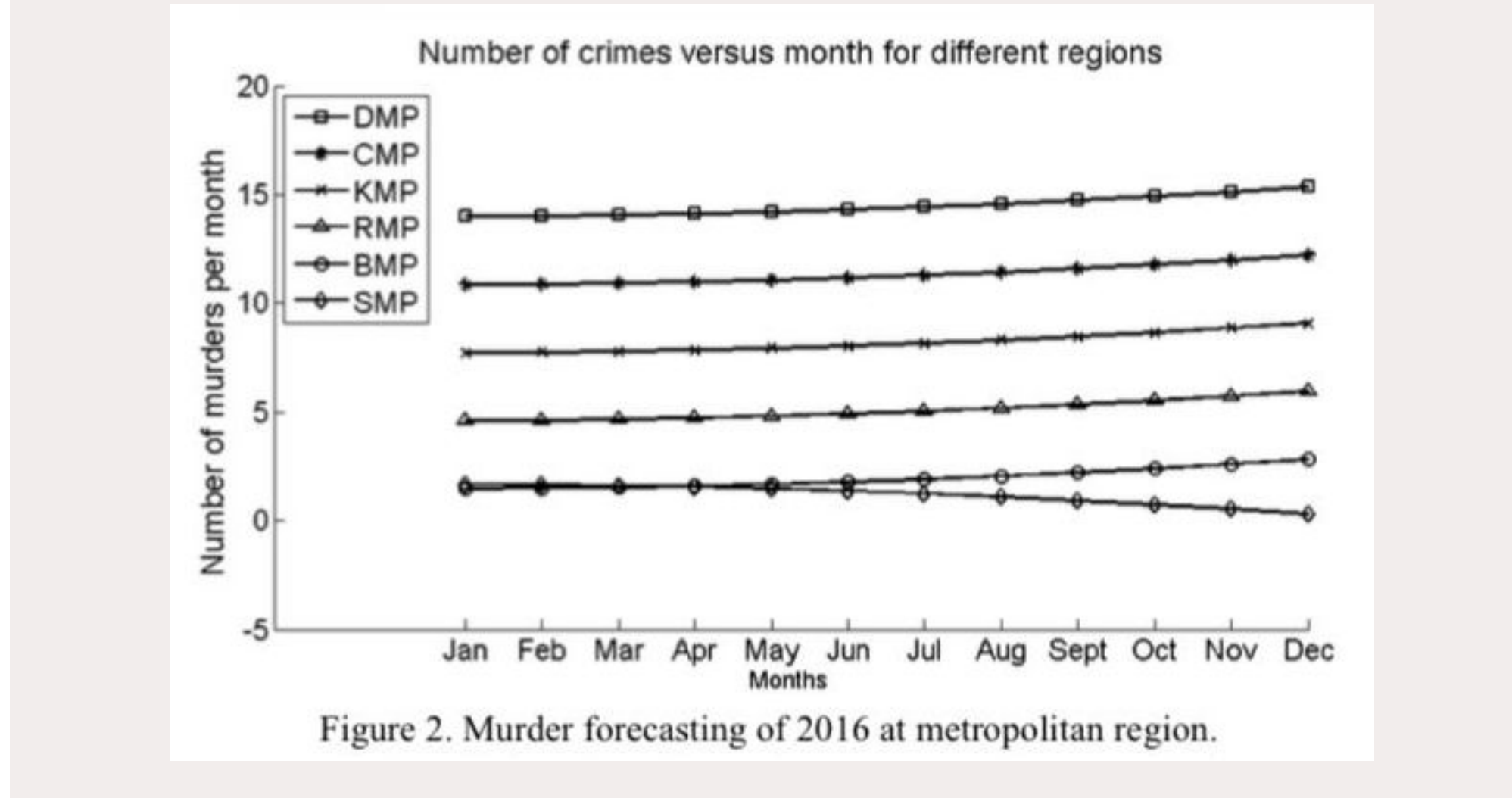
$$x' = \frac{x - \tilde{x}}{\sigma}$$

Donde  $\mathcal{X}$  es el vector original,  $\tilde{x}$  es la media de ese vector y  $\sigma$  la desviación estándar.

## Resultados:

El parámetro de aprendizaje  $\alpha$  se establece en 0.01 y el descenso del gradiente se itera 400 veces. La convergencia de la gradiente descendiente se muestra con una tasa de aprendizaje adecuada. Después del entrenamiento de regresión lineal, los parámetros de regresión para la región metropolitana son:  $\theta_0 = -212.69$ ,  $\theta_1 = -3.48$ ,  $\theta_2 = 0.24$  y  $\theta_3 = 0.11$ . Estos parámetros se usaron para pronosticar los homicidios en 2016 de la región donde se encuentra la policía metropolitana de Dhaka, definida como la región “10” en los datos, en el mes de Enero. Este procedimiento se realizó de la siguiente manera:

$$h(x) = \theta_0 + \theta_1x_1 + \theta_2x_2 + \theta_3x_3$$
$$h(x) = -212.69 + (-3.48) \cdot \text{region} + 0.24 \cdot \text{month} + 0.11 \cdot \text{year}$$
$$h(x) = -212.69 + (-3.48) \cdot 10 + 0.24 \cdot 1 + 0.11 \cdot 2016$$
$$h(x) = 15$$



## Conclusiones:

Los resultados de los gráficos estudiados respaldan la eficacia del uso de la regresión lineal para predecir futuros crímenes en Bangladesh. Destacan dos puntos importantes: primero, la región es el factor que más influye en la cantidad de homicidios, con un coeficiente de -3.48. Esto sugiere que a medida que nos acercamos a ciertas regiones designadas, la cantidad de homicidios tiende a aumentar. Segundo, se observa que la mayoría de estos crímenes tienden a incrementarse con el aumento de la población. Por lo tanto, un plan futuro implica prever los lugares específicos donde se producirán los delitos, para así tomar medidas preventivas que eviten su ocurrencia.

## Bibliografía:

Awal, M. A., Rabbi, J., Hossain, S. I., & Hashem, M. M. A. (2016, May). Using linear regression to forecast future trends in crime of Bangladesh. In 2016 5th international conference on informatics, electronics and vision (ICIEV) (pp. 333-338). IEEE.