26 de abril de 2022

ÁNGELES TERESA MENA DZUL

Facultad de ingeniería, uady

Análisis del Índice de Paz en México a través de algoritmos de Machine Learning para encontrar predicciones de interés.

Protocolo de investigación

**Análisis del Índice de Paz en México a través de algoritmos de Machine Learning para encontrar predicciones de interés.**

**Introducción**

El conocimiento que actualmente se tiene sobre el impacto económico de la violencia en el país es de carácter descriptivo de acuerdo con años previos al actual. Anualmente se recolecta información sobre este ámbito y se presenta en gráficos con la finalidad de analizar qué ocurrió en dicho año. Ante este ejercicio anual se ha logrado tener una noción de lo que ocurre alrededor de un tiempo establecido.

Estos análisis son anualmente elaborados por el Instituto para la Economía y la Paz (IEP), y su objetivo, es medir la ausencia de la violencia o temor a la violencia, lo cual indica qué tan pacífico es un país como nación y qué tan pacíficos son sus estados. Para el caso de México, tales análisis se presentan en el informe del Índice de Paz México (IPM). Para la edición de 2020 se tiene, a modo de resumen, que la Paz en México ha disminuido un 18.8% en los últimos seis años. El impacto económico de la violencia en México en 2020 fue de 4.71 billones de pesos mexicanos, lo cual es el equivalente al 22.5% del PIB nacional.

Aunque el análisis realizado por el IEP es relevante e informativo, no se ha intentado realizar un análisis de predicción respecto a lo que se espera en los próximos años. Se plantea entonces, la cuestión de hacia donde está dirigiéndose México en términos de violencia y cómo impactará esta en la economía en los próximos años. Tener las respuestas sobre las cuestiones anteriores, permitiría a las instituciones pertinentes tener un punto de partida para la formulación de estrategias para la contención de la violencia y así tratar de generar un menor impacto de esta en la economía y aprovechar tales recursos ahorrados para impulsar áreas como el desarrollo científico y tecnológico, salud y educación, por ejemplo.

El objetivo de esta investigación es realizar un análisis predictivo mediante algoritmos de Machine Learning, ya que, permite predecir qué es lo que pasará en el futuro, hacia dónde se está yendo de acuerdo con el camino que se está siguiendo en la actualidad y a datos del pasado. Otro tipo de análisis es el prescriptivo, este da una pauta para cambiar el futuro, en este caso, se usan algoritmos de inteligencia artificial, pues estos, permiten descubrir qué se tiene que hacer para obtener mejores resultados, si se llegaran a aplicar en esta investigación, se obtendría la pista o camino a seguir para disminuir los niveles de violencia y así el impacto económico de esta en el país, lo que permitirá destinar ese impacto a otros gastos gubernamentales; sin embargo, en este trabajo solamente se abordarán los análisis descriptivos.

**Objetivos**

Desarrollar algoritmos de Machine Learning que permitan un análisis predictivo del impacto económico de la violencia en México.

Replicar los gráficos del impacto económico de la violencia en México, presentados en el informe del IPM para los futuros 4 años, tomando como base de datos el portal del INEGI.

**Marco teórico**

En el trabajo de título “Entendiendo la Paz a través de las noticias del mundo”, se realizan modelos sacando los datos de GDELT, la cual es una importante fuente de datos de noticias que describe la situación socioeconómica y política mundial a través de los ojos de los medios de comunicación con técnicas de Machine Learning para estimar un índice de Paz Global, que abarca conflictos nacionales e internacionales, seguridad, fenómenos migratorios, etc.

Por otro lado, Jim Bang, Tinni Sen, John David, et al, en su reciente trabajo: “Predicción de conflictos civiles: Lo que el aprendizaje automático puede decirnos”, exploraron cómo el aprendizaje automático puede ampliar el conocimiento de los conflictos civiles y ofrecer respuestas políticas más precisas. Su punto principal en dicho trabajo fue demostrar que los algoritmos de aprendizaje automático pueden ser un sistema de alerta temprana que permita a la comunidad mundial actuar antes de que se inicie el conflicto.

Investigadores de la Universidad de Texas en Dallas han utilizado algoritmos de Machine Learning para predecir la violencia estatal en África, y esperan que la tecnología tenga aplicaciones predictivas aún más amplias. En este trabajo se utilizó una variante de algoritmo de Machine Learning, el autoMachine Learning (autoML). En el Machine Learning estándar un algoritmo analiza los datos para aprender patrones y, en última instancia, tomar decisiones sobre los nuevos datos con una mínima aportación humana, pero sigue requiriendo que un científico de datos seleccione qué algoritmo utilizar, entre otras decisiones. En el autoML, se resuelve el problema utilizando diferentes algoritmos de ML y se automatizan también otras decisiones. Esto reduce la necesidad de la participación humana durante el proceso de selección del modelo. Se concluye el trabajo con que, a efectos de elaboración de políticas, la tecnología podría utilizarse en marcos de acción anticipatoria para predecir el número de refugiados tras catástrofes naturales o de origen humano, o para prever la escasez de alimentos.

Este trabajo, al igual que los anteriores, pretende hacer uso de algoritmos de Machine Learning para predecir el impacto económico de la violencia en México, tomando como fuente de datos los mismos datos presentados en el IPM y también con datos directamente tomados de la INEGI.

**Metodología**

Los cinco indicadores del índice de paz en México son los siguientes: homicidio, delitos con violencia, delitos cometidos con armas de fuego, crímenes de la delincuencia organizada y cárcel sin sentencia. Se miden a nivel nacional y estatal, ponderando una calificación entre 1 y 5, donde 1 indica mayor grado de paz y 5 el menor grado de paz o mayor grado de violencia. Los costos que considera el impacto económico de la violencia son los siguientes: costos directos de la violencia, que son los gastos realizados por la víctima, el perpetrador y el gobierno; los costos indirectos, que son aquellos gastos que se acumulan después del acto violento e incluyen los costos a largo plazo que surgen de incidentes delictivos, como la pérdida de ingresos futuros y traumas físicos y psicológicos; por último, el efecto multiplicador, que representa los beneficios económicos que se hubieran generado si los gastos en el acto violento se hubieran dirigido a alternativas más productivas.

Se tienen datos de estos indicadores y costos de interés desde 2007, por parte del IPM, por lo que se considera que se tiene información necesaria para llevar a cabo la implementación de algoritmos de predicción. Adicionalmente se plantea el uso de datos de la INEGI para complementar y/o constatar los análisis realizados con los datos del IPM.

Como primer paso se plantea analizar los datos del índice de paz 2021 por estado y el impacto económico en un mismo dataframe mediante librerías de python, lenguaje elegido por su versatilidad para la implementación en machine learning. Por lo tanto, los pasos a seguir, de manera breve, son los siguientes:

1.- Recolección de los datos de los informes oficiales del Índice de Paz en México.

2.- Importación de los datos en python para crear un dataframe en el que se consideren todos los datos.

3.- Limpieza de los datos.

4.- Resumen estadístico.

5.- Normalización de los datos.

6.- Análisis de correlación.

7.- Implementación del algoritmo K-means

**Referencias**

1. Voukelatou, V., Miliou, I., Giannotti, F., & Pappalardo, L. (2022). Understanding peace through the world news. *EPJ Data Science*, *11*(1). doi: 10.1140/epjds/s13688-022-00315-z
2. Can we use artificial intelligence to predict civil conflict?. (2022). Recuperado 20 de Abril de 2022, de <https://www.visionofhumanity.org/predicting-civil-conflict-can-machine-learning-tell-us/>
3. Link al repositorio del proyecto: https://github.com/watashiwaangie/MexicoPeaceIndex\_Analysis/blob/master/IPM\_2020.ipynb