

Entrega del Proyecto Final

Entrenamiento y optimización de modelos de Machine Learning

Introducción:

En los últimos años, Brasil ha sido escenario de un marcado aumento en las manifestaciones y protestas que han sacudido las calles del país. Desde la gigantesca nación amazónica hasta las bulliciosas ciudades del sur, los brasileños han alzado sus voces en busca de cambios significativos en diversas áreas de la sociedad.

Estas protestas han reflejado la creciente insatisfacción de la población brasileña con una serie de problemas urgentes. La corrupción política, la crisis económica, los recortes en el gasto público, la violencia policial y las desigualdades sociales han sido algunos de los principales motivos que han llevado a las masas a movilizarse en todo el país.

Uno de los eventos más significativos en la historia reciente de Brasil fue la ola de manifestaciones que se desencadenó a partir del año 2018. Inicialmente impulsadas por el aumento de las tarifas del transporte público, estas protestas se convirtieron en una expresión generalizada de indignación contra la corrupción y la mala gestión gubernamental. Las manifestaciones congregaron a millones de personas en las calles de ciudades como Río de Janeiro, Sao Paulo y Brasilia, y tuvieron un impacto duradero en el país.

En años posteriores, las protestas se han diversificado en sus demandas y en las personas que participan en ellas. Movimientos sociales, estudiantes, trabajadores, grupos indígenas y defensores de los derechos humanos se han unido en un coro de voces exigiendo justicia y cambios en diversas áreas de la sociedad brasileña.

Estas protestas han llevado a avances significativos. Desde reformas políticas y cambios en la legislación hasta un mayor escrutinio de la corrupción y una mayor conciencia sobre los derechos humanos, las manifestaciones han demostrado su poder para generar impacto y catalizar transformaciones.

En este reporte , exploramos algunas de las protestas más destacadas ocurridas en Brasil en los últimos cinco años. Analizaremos las características de estas movilizaciones, así como también buscaremos indagar y mostrar al público de interés la situación del país. A través de estas narrativas y visualizaciones, podremos comprender mejor el espíritu de lucha y la resiliencia del pueblo brasileño en su búsqueda de un futuro más justo y equitativo.

Este reporte va dirigido a todo aquel que esté interesado por conocer la situación de la sociedad brasileña, es de carácter público por lo que puede utilizarse o tomar información relevante para otros análisis o investigaciones que enriquezcan la

información circundante. Nuestro objetivo base, es analizar la información y explotar nuestra fuente de datos sobre las movilizaciones de los últimos años en nuestro país y aportar conocimiento de causa con el fin de mejorar la situación política y social.

Preguntas Problemas:

A partir de lo comentado en los párrafos anteriores, pretendemos indagar sobre algunas preguntas problemas que nos han ido surgiendo en este último tiempo y nos movilizan a realizar dicha investigación, las cuales son las siguientes.

- 1) ¿Cuál es la distribución de los diferentes tipos de eventos? ¿Cuáles son los tipos de eventos más comunes y los menos comunes?
- 2) ¿Cuáles son los tipos de eventos que han resultado en el mayor número de fatalidades?
- 3) ¿Cómo varía la cantidad de eventos a lo largo del tiempo? ¿Hay patrones estacionales o tendencias a lo largo de los años?
- 4) ¿Cómo podemos clasificar la ocurrencia de los tipos de eventos en Brasil en función de la información con las que contamos? ¿Cuáles son las variables o factores más influyentes para lograr una clasificación precisa de los eventos?

Objetivos de Investigación:

Por otro lado nuestros objetivos a cumplir y destacar para dicha investigación se basan en lo comentado a continuación:

- 1) Analizar y describir la distribución de los diferentes tipos de eventos presentes en el conjunto de datos, identificando cuáles son los tipos de eventos más comunes y cuáles son los menos comunes.

2) Identificar y examinar en detalle los tipos de eventos que han resultado en el mayor número de fatalidades.

3) Evaluar la variación de la cantidad de eventos a lo largo del tiempo y determinar si existen patrones estacionales o tendencias a lo largo de los años.

4) Desarrollar modelos de Inteligencia Artificial para clasificar la ocurrencia de diferentes tipos de eventos en Brasil, utilizando la información disponible en el conjunto de datos

Fuente de Datos:

Se utilizó como fuente de datos un Dataset proveniente del Ministerio De Seguridad de la República Del Brasil. El Dataset se denomina “Violencia y Protestas en Brasil”.

El Dataset cuenta con las siguientes columnas:

EVENT_DATE : Fecha del Evento

EVENT_TYPE : Tipo de Evento

SUB_EVENT_TYPE : Subtipo de Evento

ACTOR1 : Actor Involucrado

ACTOR2 : Actor Involucrado

COUNTRY : Pais

LOCATION : Locacion

LATITUDE : Latitud

LONGITUDE : Longitud

SOURCE_SCALE : Escala

NOTES : Notas

FATALITIES : Fatalidades

Unnamed: 12 : Columna Sin Datos

Unnamed: 13 : Columna Sin Datos

Unnamed: 14 : Columna Sin Datos

Unnamed: 15 : Columna Sin Datos

Unnamed: 16 : Columna Sin Datos

Transformaciones Realizadas / Limpieza del Dataset:

- Eliminación de Columnas sin Valor (Unnamed: 12 ,

Unnamed: 13 ,
Unnamed: 14 ,
Unnamed: 15 ,
Unnamed: 16)

- Conversión de Valores Null a 'Sin Dato' en columna Actor 2.
- Cambio del tipo de dato de la columna Event_Date de Object a DateTime.
- Cambio del tipo de dato de la columna Fatalities de Float a Int.

Creación de Variables Sintéticas

- Se creó la Variable Fatalities_Category, dependiente de la columna Fatalities , siendo esta Categórica y tomando como referencia la cantidad de fatalidades y agrupandolas en diferentes categorías.

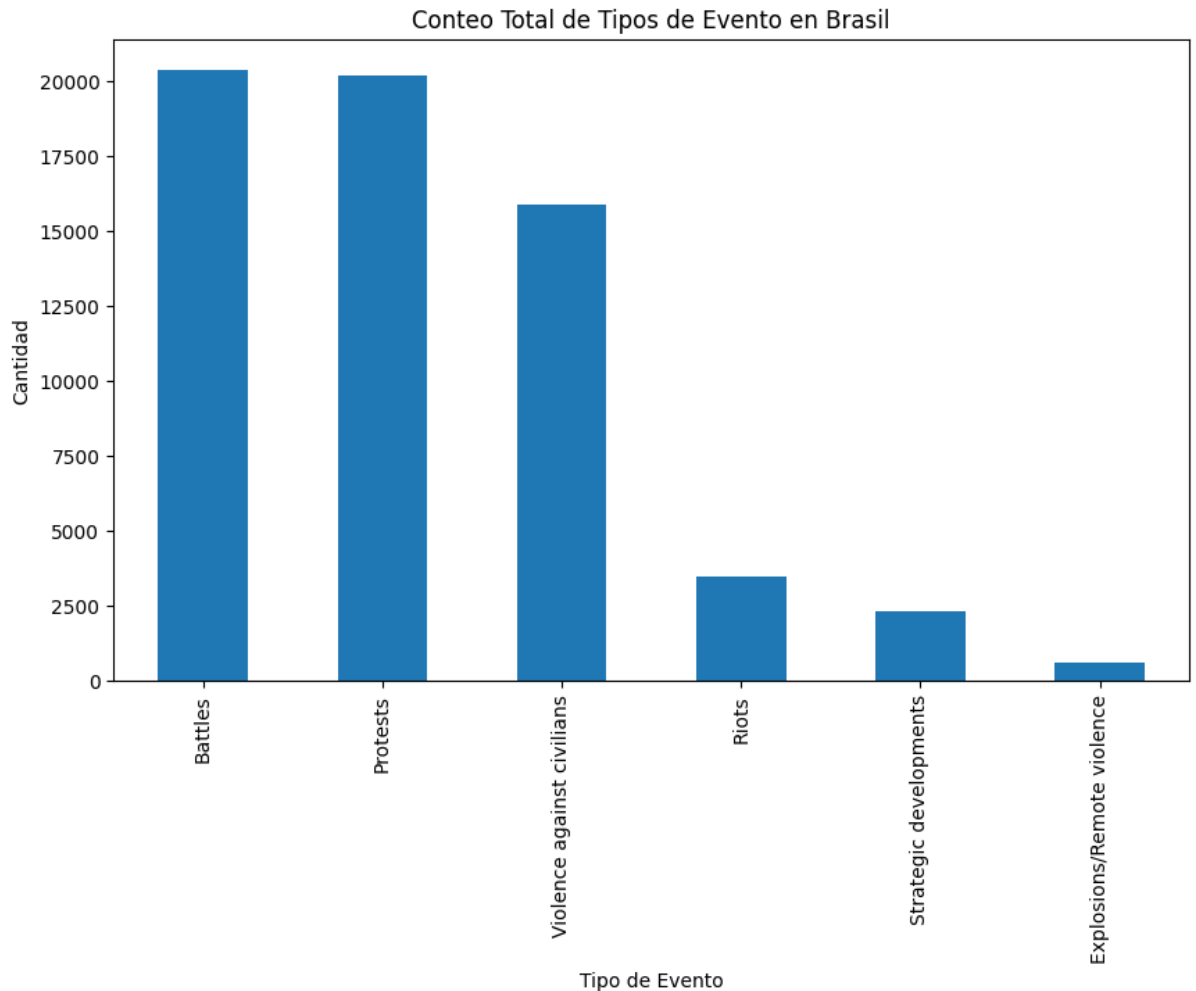
Análisis exploratorio

- A partir del exhaustivo estudio sobre diferentes tipos de eventos de protestas y violencia que ocurren en el país han sido categorizados seis tipos distintos, y para comprenderlos mejor se han utilizado diversas técnicas e instrumentos de recolección de información.

Un análisis detallado revela que el tipo de evento más común es conocido como "Batallas". Estas manifestaciones se caracterizan por ser protestas no pacíficas, cuyo único propósito es generar violencia física hacia ciudadanos o fuerzas gubernamentales.

Al observar el gráfico, podemos apreciar la distribución de cada uno de los tipos de eventos. Aunque las "Batallas" predominan en términos de cantidad,

también es importante destacar la presencia de otros tipos de protestas y violencia, cada uno con sus particularidades y motivaciones específicas.



- **Además de los tipos de eventos de protestas en Brasil hemos desarrollado una lista de Subtipos, que son acciones que se desenlazan dentro de los eventos.**

La más común son los enfrentamientos armados, Brasil enfrenta desafíos significativos en lo que respecta a la presencia y el acceso a armas de fuego. El país tiene una alta incidencia de armas de fuego en circulación, lo que ha contribuido a un aumento de la violencia armada y los delitos relacionados. Las protestas pacíficas también por detrás de los enfrentamientos armados son comunes en dicho territorio, en las cuales los ciudadanos expresan sus

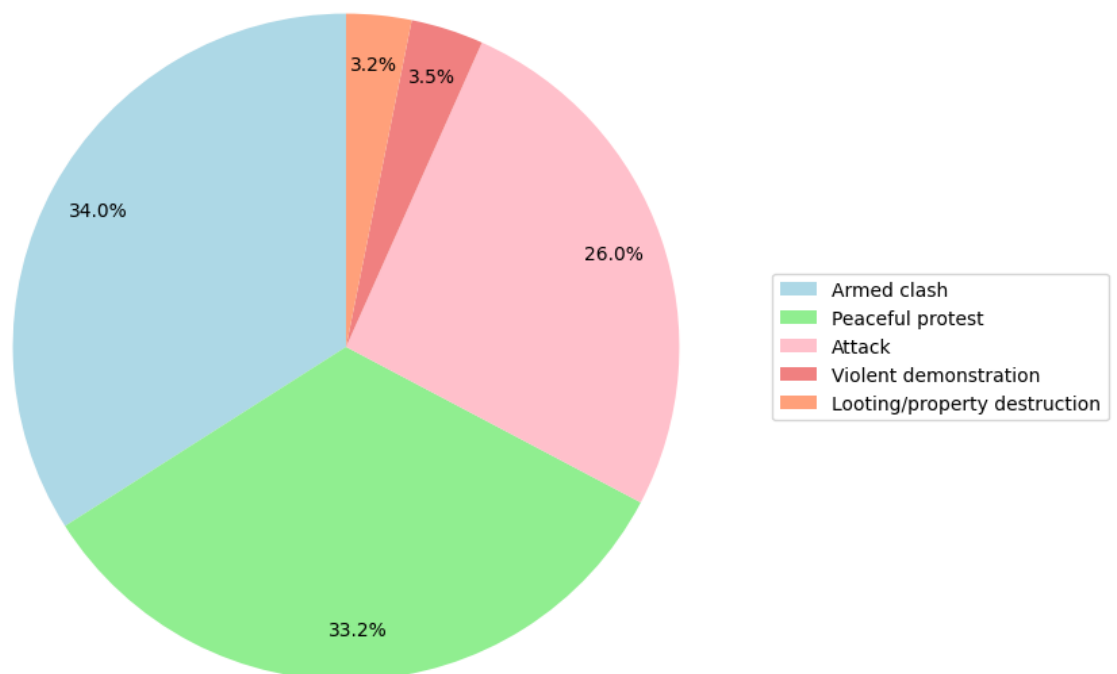
opiniones, demandas y preocupaciones de manera no violenta. Estas manifestaciones suelen ser parte del ejercicio del derecho a la libertad de expresión y reunión.

En muchas ocasiones, estas manifestaciones pacíficas se organizan mediante convocatorias en redes sociales, grupos de activistas y organizaciones de la sociedad civil. Los participantes suelen llevar pancartas, banderas, realizar marchas, concentraciones o actos simbólicos para visibilizar sus demandas y generar conciencia en la opinión pública y las autoridades.

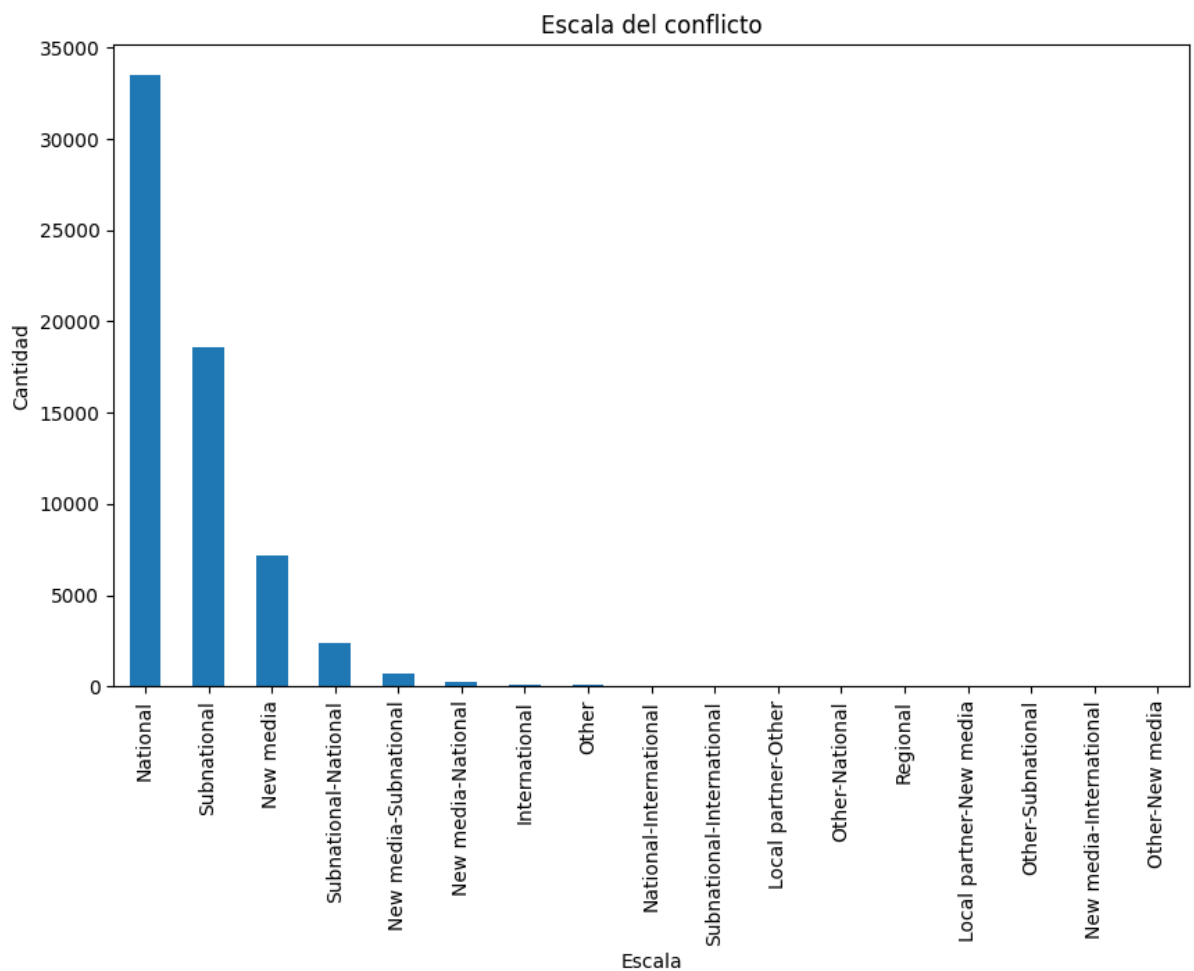
Es importante destacar que el derecho a la protesta pacífica está protegido por la Constitución brasileña y por instrumentos internacionales de derechos humanos. Sin embargo, es necesario que estas protestas se desarrollen dentro de los límites legales y respetando los derechos de los demás ciudadanos.

Las protestas pacíficas desempeñan un papel fundamental en la democracia, ya que permiten a la ciudadanía expresar sus inquietudes y contribuyen al debate público y a la búsqueda de soluciones a los problemas sociales.

Top 5 de Subtipos de Eventos más Frecuentes en Brasil



- La escala de conflicto a nivel territorial aquí utilizada , se refiere a una clasificación o categorización que se utiliza para medir y evaluar la intensidad o gravedad de un conflicto en un determinado territorio o área geográfica. Esta busca proporcionar una comprensión más clara y sistemática de la situación conflictiva y permite comparar diferentes conflictos entre sí. Como vemos en el gráfico siguiente, las escalas de conflicto nacional y subnacional son las más abarcativas. En forma explicativa un conflicto nacional, se refiere a un conflicto que abarca todo el territorio de un país, mientras un conflicto subnacional, se refiere a conflictos que ocurren en una región o área geográfica específica dentro de un país. Es importante señalar que los conflictos pueden evolucionar y cambiar de escala con el tiempo. Lo que puede comenzar como un conflicto subnacional en una región específica puede escalar y convertirse en un conflicto nacional si no se resuelven adecuadamente.



- **Brasil es el quinto país más grande del mundo en términos de área terrestre. Tiene una superficie total de aproximadamente 8.515.767 kilómetros cuadrados, lo que equivale a 3.287.956 millas cuadradas, En cuanto a la población, Brasil tiene una población estimada de alrededor de 213,99 millones de habitantes.**

Río de Janeiro, se estima en aproximadamente 6.75 millones de habitantes, es la segunda ciudad más poblada de Brasil, después de São Paulo.

Dada que en la gráfica realizada en Río de Janeiro la ciudad con mayor cantidad de protestas, es convincente poder agregar informacion al respecto para que nuestro público entienda el contexto relacionado a los datos:

Desigualdad social,Río de Janeiro es una ciudad con marcadas disparidades socioeconómicas.

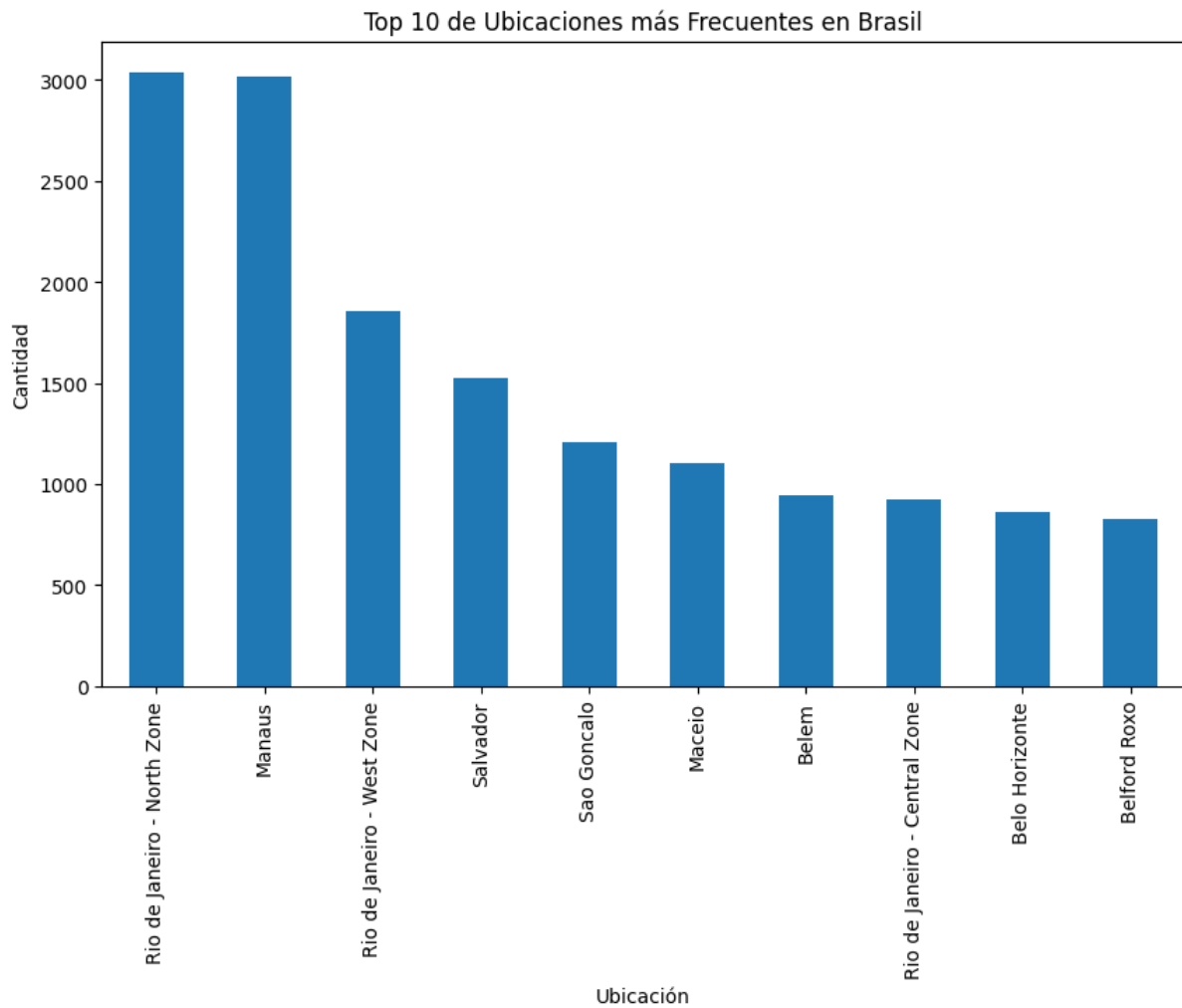
Problemas de seguridad,Río de Janeiro ha enfrentado desafíos en cuanto a la seguridad pública. El crimen organizado, la violencia urbana y la percepción de inseguridad pueden llevar a protestas para exigir medidas más efectivas por parte del gobierno en la lucha contra la delincuencia.

La corrupción es un problema que afecta a todo Brasil, y Río de Janeiro no es una excepción. Escándalos de corrupción en el gobierno local o en instituciones públicas pueden generar indignación y movilizar a la población para protestar contra la impunidad y exigir transparencia y rendición de cuentas.

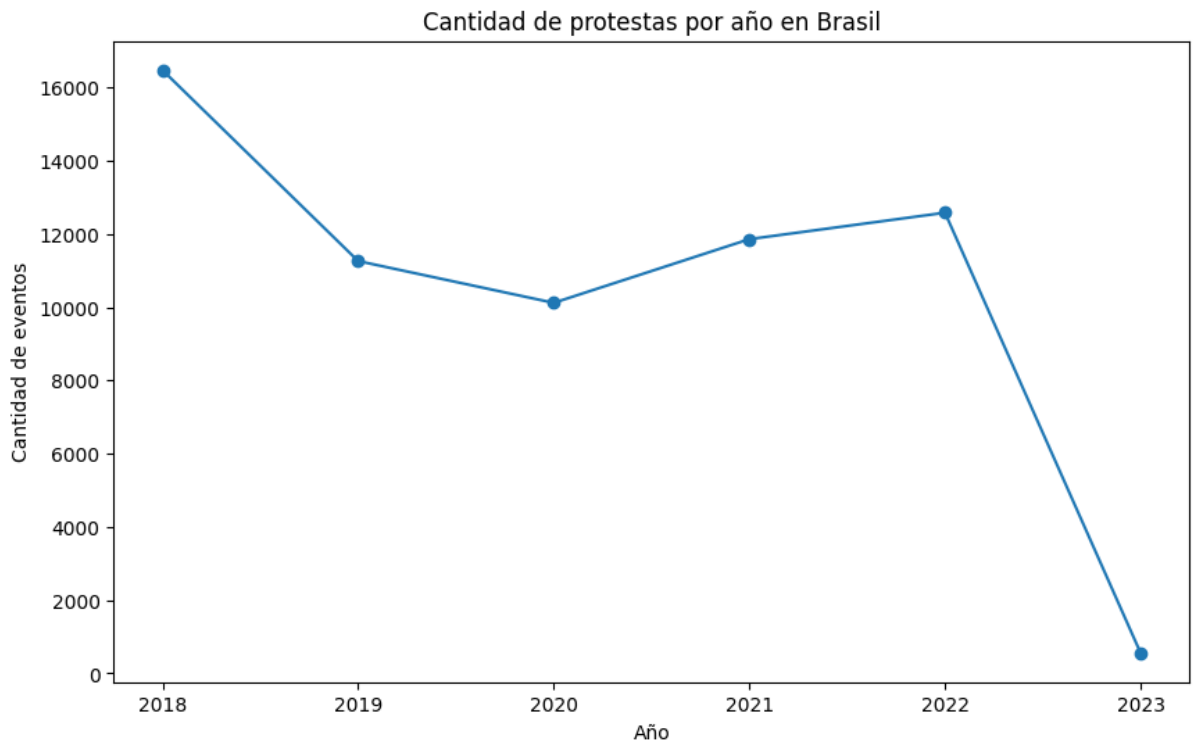
Problemas ambientales,Río de Janeiro se enfrenta a desafíos ambientales, como la contaminación del agua, la deforestación y la falta de acceso a servicios básicos en algunas áreas. Estos problemas pueden llevar a protestas por parte de grupos ambientalistas y de la sociedad civil que buscan proteger el medio ambiente y mejorar la calidad de vida de los residentes.

Descontento político,las protestas también son una expresión del descontento generalizado hacia el gobierno, ya sea a nivel local, estatal o federal. Las demandas pueden variar, desde críticas a políticas específicas hasta la exigencia de cambios en la estructura política y la participación ciudadana.

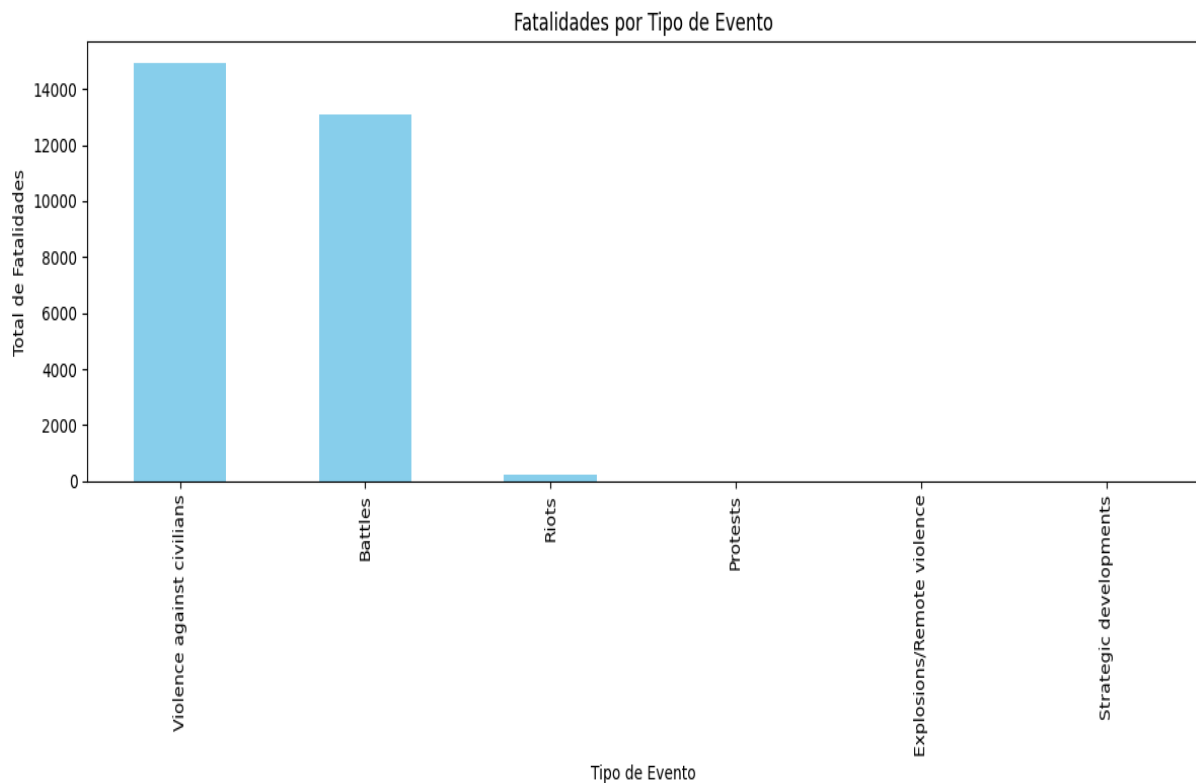
Las demás ciudades de la lista, aunque con sus propias características y demandas, en menor o mayor medida también comparten las problemáticas aquí descritas. una región específica puede escalar y convertirse en un conflicto nacional si no se resuelven adecuadamente.



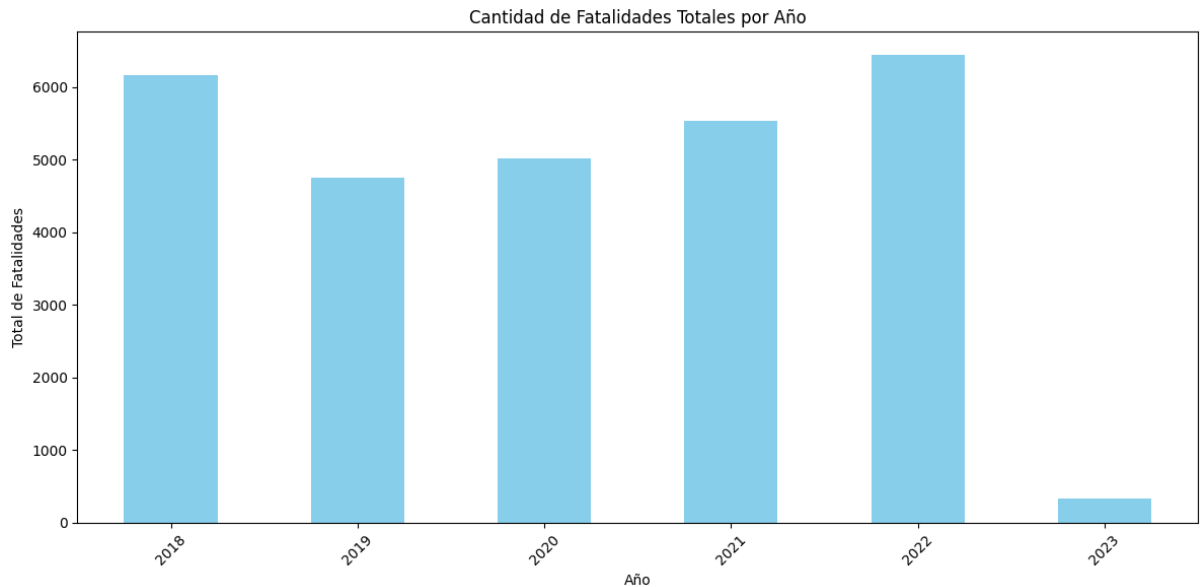
- En los últimos años la sumatoria de las protestas en Brasil nos arroja que se mantienen por encima de las diez mil anuales, siendo de diferentes índoles, propósitos y objetivos, pasando por diferentes gobiernos, políticas y leyes. La protesta es una forma de representación ciudadana ante situaciones que no les son beneficiosas en su totalidad. Como vemos el pico de los últimos cinco años fue en el 2018 superando las dieciséis mil protestas anuales. Actualmente la recolección de los datos solo llega hasta una parte del año en curso por razones obvias, pero siguiendo la tendencia esperamos que se mantenga en la media anual aquí demostrada.



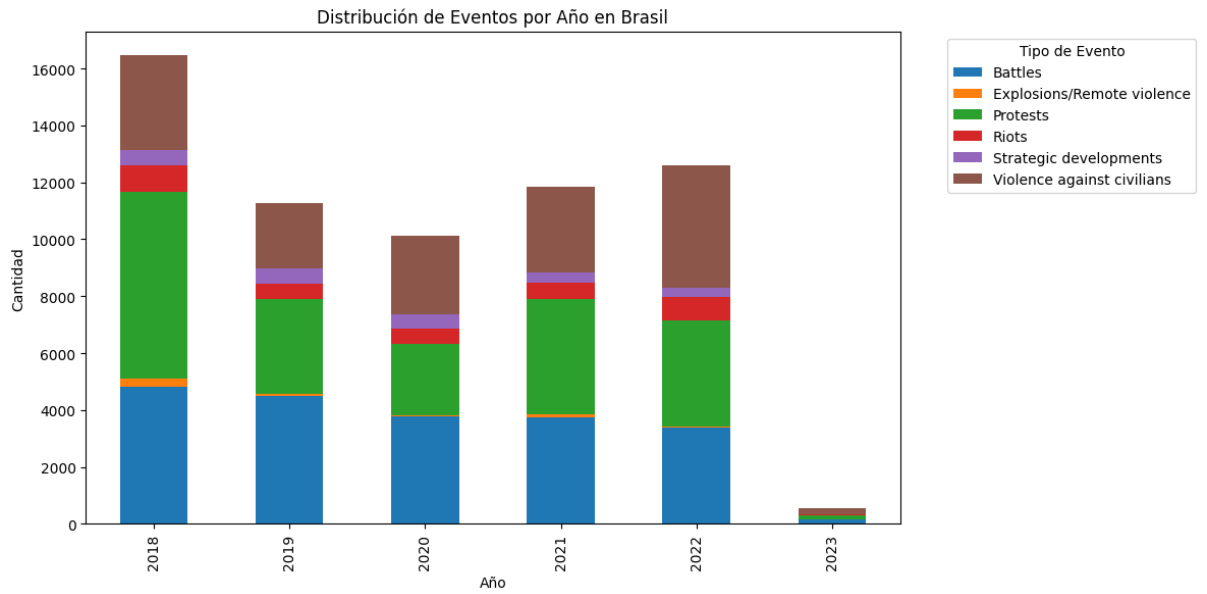
- **Aquí nos interesaba saber la cantidad de fatalidades que había en total por tipo de protestas en Brasil. Como bien puede observarse, hay una gran diferencia entre dos de los tipos de protestas con los demás. La 'Violencia con Civiles' y las 'Batallas' son quienes por mucho llevan la delantera en cuanto a decesos. Esta disparidad en el número de fatalidades sugiere que estos dos tipos de protestas pueden ser particularmente peligrosos tanto para los manifestantes como para las fuerzas de seguridad. Sería relevante a futuro investigar más a fondo las circunstancias y los factores que contribuyen a la alta mortalidad en estos tipos de eventos, con el objetivo de encontrar formas de reducir los riesgos y mejorar la seguridad durante las protestas en Brasil.**



- En este gráfico vemos que el año con mayor número de fatalidades es el 2022, con un marcado aumento en comparación con los años anteriores. Esto sugiere que el año 2022 fue particularmente peligroso en términos de eventos con fatalidades. Es importante destacar que este aumento significativo en las fatalidades coincide con un período específico y puede ser indicativo de desafíos de seguridad y situaciones críticas que se vivieron durante ese año en Brasil. Por otro lado, el año con la menor cantidad de fatalidades es el 2019, lo que indica un período relativamente más tranquilo en términos de eventos violentos con consecuencias fatales en el país. Sin embargo, es crucial recordar que detrás de cada cifra de fatalidad hay vidas y familias afectadas. Estos datos destacan la importancia de la seguridad y la prevención de conflictos en Brasil, así como la necesidad de un análisis detallado para comprender los factores que contribuyen a estas fluctuaciones en el número de fatalidades a lo largo de los años en el país.

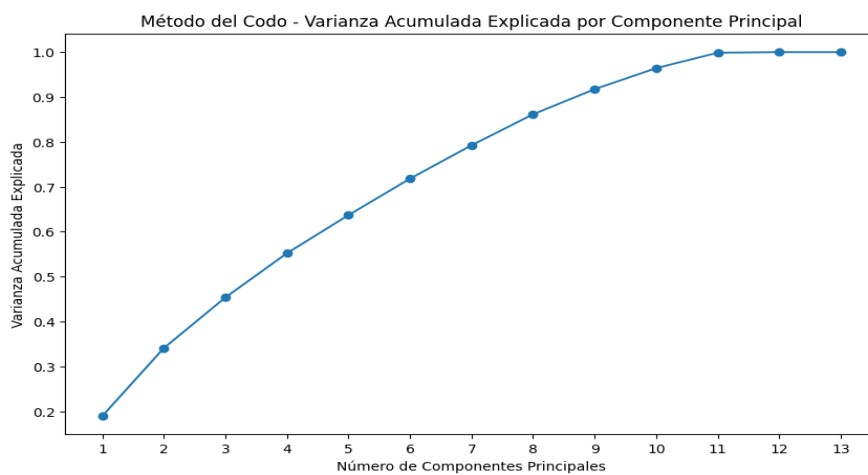


- En este gráfico vemos que el año con más eventos fue el 2018. Tal como lo analizamos anteriormente, los tres tipos de eventos más comunes son 'Batallas', 'Protestas' y 'Violencia con Civiles', con una gran ventaja en relación a los otros tipos de eventos. Esto sugiere que estos tres tipos de eventos son recurrentes y pueden estar relacionados con situaciones específicas en Brasil que generan movilizaciones y conflictos. Por otra parte, el año con menos eventos fue el 2020, lo que indica una disminución en la actividad en comparación con años anteriores. Sin embargo, es importante recordar que la cantidad de eventos no siempre refleja la intensidad o gravedad de las situaciones que subyacen a ellos.

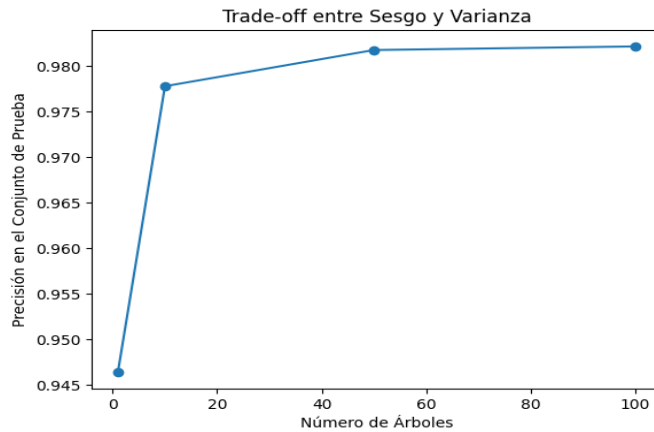


Modelo de Machine Learning

- Se aplicó Label Encoder a las Variables para convertirlas en numéricas.
- La Variable Objetivo a clasificar es Event_Type.
- Se estandarizaron las características mediante StandarScaler.
- Aplico PCA y veo el Número óptimo de Componentes Principales.



- Elijo en 11 mis componentes principales
- En primer lugar se entrenó un modelo de Random Forest Classifier con distinta cantidad de árboles, en este caso (1, 10, 50, 100)



- Se entrenó un Modelo de Random Forest Classifier con 100 árboles dando los siguientes resultados:

Precisión del modelo: 0.98

Precisión: 0.98

Recall: 0.93

F1-Score: 0.95

Matriz de Confusión:

```
[[4074  0  0  0  1 55]
 [  0 106  1 25  2  1]
 [  2  0 4008  8  3  0]
 [  7  0  21 675  1  4]
 [ 37  0  4 16 374  8]
 [ 19  0  0  6  3 3111]]
```

Importancia de Características:

EVENT_DATE: 0.1403

EVENT_TYPE: 0.1053

SUB_EVENT_TYPE: 0.1162

ACTOR1: 0.0542

ACTOR2: 0.0451

COUNTRY: 0.0097
LOCATION: 0.0497
SOURCE_SCALE: 0.0235
NOTES: 0.1035
FATALITIES_CATEGORY: 0.2029
FATALITIES: 0.1496

- Se Aplicó al modelo anterior Validación Cruzada mediante K-FOLD , dando el siguiente resultado:

Puntajes de Precisión en Validación Cruzada:

Pliegue 1: 0.98

Pliegue 2: 0.96

Pliegue 3: 0.98

Pliegue 4: 0.98

Pliegue 5: 0.75

Puntaje Promedio de Precisión en Validación Cruzada: 0.9293624873777713

- Se entrenó un modelo de SCV dando el siguiente resultado:

Precisión del modelo: 0.99

Precisión: 0.98

Recall: 0.98

F1-Score: 0.98

Matriz de Confusión:

```
[[4130  0  0  0  0  0]
 [  0 130  3  0  2  0]
 [  0  0 4018  1  2  0]
 [  0  0  0 708  0  0]
 [  5  1  6  3 418  6]
 [ 41 11  7  0  0 3080]]
```

Importancia de Características:

EVENT_DATE: 0.5534

EVENT_TYPE: 0.5229

SUB_EVENT_TYPE: 0.3503

ACTOR1: -0.4426

ACTOR2: -0.4277

COUNTRY: 0.1505

LOCATION: -0.5234
SOURCE_SCALE: -0.3751
NOTES: 0.9425
FATALITIES_CATEGORY: -0.4261
FATALITIES: -0.6748

- Se Aplicó al modelo anterior Validación Cruzada mediante K-FOLD , dando el siguiente resultado:

Puntajes de Precisión en Validación Cruzada:

Pliegue 1: 0.99

Pliegue 2: 0.99

Pliegue 3: 1.00

Pliegue 4: 0.99

Pliegue 5: 0.99

Puntaje Promedio de Precisión en Validación Cruzada: 0.9931273914911

Conclusiones:

En conclusión, las protestas en Brasil han sido, son, y serán una manifestación significativa de la participación ciudadana y la búsqueda de cambios en diversos aspectos de la sociedad. A lo largo de los años, Brasil ha sido testigo de una amplia gama de protestas que reflejan las preocupaciones y demandas de la población.

Estas manifestaciones han abordado temas políticos, sociales, económicos y medioambientales, entre otros. Desde huelgas de camioneros hasta movimientos estudiantiles y protestas contra la violencia policial, las personas han utilizado las protestas como una forma de expresar su descontento, exigir derechos, promover la justicia y luchar por un cambio positivo en el país.

Las protestas han demostrado la diversidad y la pluralidad de la sociedad brasileña, al tiempo que han resaltado problemas y desafíos específicos que enfrenta el país. Han generado debates públicos, han puesto temas importantes en la agenda política y han ejercido presión sobre las autoridades para abordar las demandas ciudadanas.

Sin embargo, también es importante destacar que las protestas en Brasil a menudo han estado acompañadas de desafíos y tensiones, como la polarización política, la represión y la violencia. Estos aspectos subrayan la complejidad del escenario social y político en el que se desarrollan las manifestaciones.

En última instancia, las protestas en Brasil reflejan la vitalidad de la democracia y la participación ciudadana en la búsqueda de un país más justo y equitativo. Si bien algunas demandas pueden resolverse con el tiempo, otras pueden requerir un diálogo continuo, la implementación de políticas efectivas y un compromiso sostenido por parte de los actores involucrados.

La comprensión y el respeto de los derechos de libre expresión y reunión son fundamentales para garantizar un entorno propicio para las protestas pacíficas y constructivas, donde las voces de los ciudadanos sean escuchadas y consideradas en la toma de decisiones.

Con respecto a los Modelos vemos que :

A) En el Ramdon Forest con diferentes cantidad de Árboles el ideal y que se Ajusta de la mejor manera para una clasificación en torno al tipo de evento es 100 Árboles.

B) En tanto el Ramdon Forest con 100 Árboles nos devuelve las siguientes métricas:

Precisión del modelo: El modelo generalmente acierta alrededor del 0.98% de las veces, lo que es bastante bueno.

Precisión por clase: En promedio, el modelo es correcto en un 0.98% al predecir cada tipo de evento.

Recall : El modelo identifica correctamente alrededor del 0.93% de los casos positivos en promedio.

F1 Score : En general, el modelo encuentra un equilibrio razonable entre la precisión y la sensibilidad, alrededor del 0.95%.

Confusión : se visualizan errores en el que el modelo confunde un tipo de evento con otro.

En cuanto a la importancia de cada característica que participa del modelo en resumen tenemos lo siguiente:

Fatalities_category: Tiene una importancia de 0.2029. Esto significa que en el contexto del modelo o análisis, la categoría de fatalidades es la característica más importante, con un peso significativamente alto.

Por otro lado, la característica con la importancia más baja en la lista proporcionada es **SOURCE_SCALE**

El resto de las demás características se distribuye entre ambas dos.

En cuanto a la validación cruzada del modelo, vemos que el puntaje promedio es aproximadamente 0.9293, lo que significa que, en promedio, el modelo logró una precisión del 92.93% en todos los pliegues.

C) Se entrenó un nuevo modelo, en este caso un SVC lo que nos arrojó los siguientes resultados:

Precisión del modelo: El modelo generalmente acierta alrededor del 0.99% de las veces, lo que es bastante bueno.

Precisión por clase: En promedio, el modelo es correcto en un 0.98% al predecir cada tipo de evento.

Recall : El modelo identifica correctamente alrededor del 0.98% de los casos positivos en promedio.

F1 Score : En general, el modelo encuentra un equilibrio razonable entre la precisión y la sensibilidad, alrededor del 0.98%.

Confusión : se visualizan errores en el que el modelo confunde un tipo de evento con otro.

En cuanto a la importancia de cada característica que participa del modelo en resumen tenemos lo siguiente:

NOTES: Tiene la importancia más alta en la lista, con un valor de 0.9425. Esto indica que, en este contexto, las notas son la característica más importante en términos de su capacidad para predecir o explicar resultados.

EVENT_DATE: Tiene una importancia de 0.5534. Aunque es una característica importante, su peso es menor en comparación con las notas, por lo que se encuentra en segundo lugar en importancia.

Por otro lado, las características con valores negativos pueden tener un impacto adverso en la capacidad del modelo para hacer predicciones precisas

En cuanto a la validación cruzada del modelo, vemos que el puntaje promedio es aproximadamente 0.9931, lo que significa que, en promedio, el modelo logró una precisión del 99.31% en todos los pliegues.

De esta manera, considero que ambos modelos cumplen las expectativas a la hora de clasificar la variable de tipo de evento.