



## Presentación y Descripción del Problema

Desde mediados de octubre de 2019, Chile se ha visto enfrentado a una seguidilla de movilizaciones y eventos sociales nunca antes vistos desde el retorno a la democracia. Lo que comenzó inicialmente con un llamado a la evasión del pago del metro por un alza en el valor de 30 pesos chilenos en el pasaje, ha escalado a hechos graves como, por ejemplo, la quema de más de 20 estaciones del Metro de Santiago. El denominado “estallido social” viene a sumarse a una serie de protestas a nivel global con características particulares, que las diferencian estructuralmente en la forma en que las comunidades habían manifestado tradicionalmente la solución a sus demandas. Juan Carlos Vargas, analista político declara: *“No había ocurrido en los últimos años una situación tan grave y de movilización. Este proceso es muy complejo, no hay líderes en este movimiento, es una especie de organización espontánea, de autoorganización por redes sociales que ha dificultado que la clase política pueda convertir esta demanda en soluciones y en políticas públicas”*<sup>1</sup>. Por otra parte, el escritor Moisés Naim menciona que *“estos procesos son nebulosos en su estructura -que tiene mucho de organización espontánea y autoorganización por redes sociales-, no tienen líderes claros, como en Hong Kong o Ecuador. Estas protestas no tienen liderazgos y dependen muchísimo de redes sociales. Lo hemos visto en Moscú, en Teherán y Bagdad. Lo otro llamativo es la velocidad con que ocurre todo. Y los gobiernos que quieren negociar no saben con quién hacerlo, porque no hay interlocutores.”*<sup>2</sup>

Según lo expuesto, podemos destacar la falta de interlocutores válidos en las negociaciones, y la autoorganización mediante redes sociales como puntos relevantes en nuestro análisis. Sin liderazgos claramente identificables y más bien distribuidos por todo lo extenso de la ciudadanía, las plataformas de redes sociales se vuelven una significativa fuente de información para poder explorar las demandas perseguidas, conocer los posibles lugares y horarios de las manifestaciones, y por lo tanto, inferir las consecuencias que estas puedan traer al desarrollo y estabilidad socio-económico del país. Esto es muy relevante tanto para el ámbito público (justicia social, orden público, seguridad, etc.) como para el sector privado (oportunidades de inversión, emplazamiento de negocios, impacto de indicadores económicos, etc.).

Investigaciones académicas ya han estudiado la utilidad del análisis de redes sociales orientado al ámbito económico<sup>3</sup>, en particular como una herramienta para apoyar las decisiones de inversión<sup>4</sup>. Si bien la literatura no indica un poder predictivo definitivo, sí se observa su rol como complementario a los análisis financieros. Conocer a más profundidad el rol de las redes sociales en las decisiones de inversión resulta ser de interés

---

<sup>1</sup> “No hay líderes en este movimiento, es una organización ....” 23 Oct. 2019, <http://www.radionacional.com.ar/no-hay-lideres-en-este-movimiento-es-una-organizacion-espontanea/>. Accessed 27 Nov. 2019.

<sup>2</sup> “Moisés Naim: “Este tipo de protestas son espontáneas y sin ....” 22 Oct. 2019, <https://www.latercera.com/mundo/noticia/mois-es-naim-este-tipo-protestas-espontaneas-sin-lideres/872088/>. Accessed 27 Nov. 2019.

<sup>3</sup> A. Groß-Klußmann, S. König & M. Ebner (2019). Buzzwords build momentum: Global financial Twitter sentiment and the aggregate stock market. *Expert Systems With Applications*, 136, 171–186.

<sup>4</sup> Ermanno Affuso, E., Lahtinen, K.D. (2018). Social media sentiment and market behavior.

para informar mejor a los que toman decisiones económicas en el país, en particular cuando se trata de decisiones en ventanas de tiempo comprendidas en semanas, como ocurre en el estallido social chileno.

Para este ejercicio ejemplificamos el análisis como si estuviese al servicio de una empresa ficticia que quiere desarrollar un emprendimiento de inversión en Chile, el cual consistirá principalmente en manejar fondos de inversiones de terceros, diversificando las colocaciones en varios negocios para minimizar los riesgos. Se contará con un capital inicial y una cartera de empresas donde invertir, las cuales tienen un retorno asegurado. El valor agregado de esta empresa será aplicar análisis inteligentes con herramientas de *machine learning* o similares sobre los distintos negocios, poniendo énfasis en el entorno en que estas se desenvuelven, considerando su riesgo y estabilidad económico-social.

Debido al “estallido social”, la empresa decide hacer honor a su metodología y desea implementar un modelo y responder a la pregunta: ¿Es conveniente arriesgarse con una empresa de inversión con la situación actual en Chile?

La empresa entonces define que es necesario realizar un análisis de sentimientos de la situación actual que se vive en Chile, recopilando la opiniones que se han expresado en la plataforma Twitter sobre los diferentes acontecimientos. Luego se analizará si los distintos acontecimiento han afectado a los principales indicadores económicos del país: Valor del Dólar, IPSA y Valor del Cobre. De esta forma contar con un análisis real sobre la situación de la economía actual y decidir el momento idóneo para comenzar con el emprendimiento.

## Objetivo General

Detectar la probabilidad de éxito de invertir, a corto plazo, con una empresa de inversiones en Chile, considerando la situación actual del país. Esto utilizando métodos estadísticos y de machine learning.

## Objetivos Específicos

- Extraer diversas opiniones con respecto a los distintos acontecimientos que han ocurrido en Chile desde el inicio de la explosión social y, desde dichas opiniones, establecer un perfil de sentimiento social.
- Recopilar y Observar el comportamiento económico del valor del dólar, el IPSA y el valor del cobre, antes y durante el estallido social en Chile.
- Establecer un modelo que permita evaluar la existencia y grado de impacto de los acontecimientos sociales sobre los indicadores económicos mencionados, en base a las opiniones declaradas en Twitter y el perfil de sentimiento detectado.

## Alcances

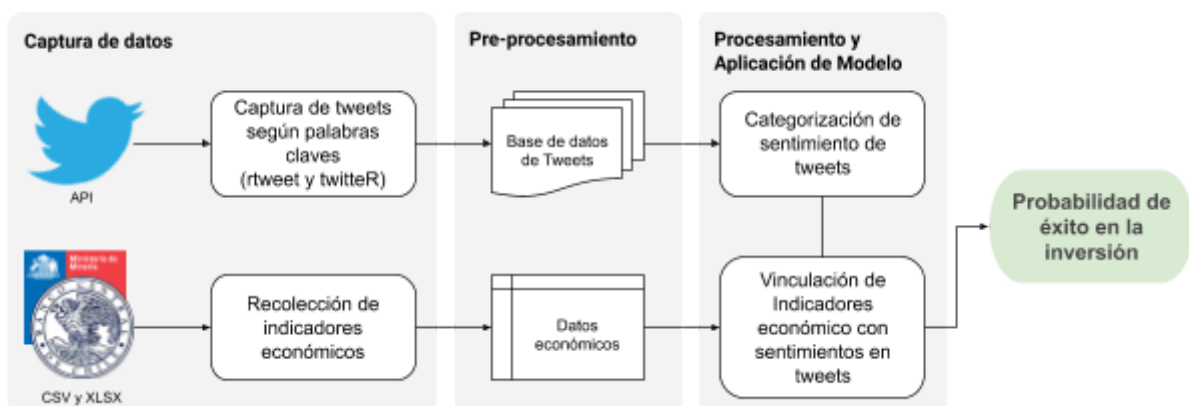
- Se descargara por medio de un Script en R y cuenta *Developer Standard* que ofrece Twitter, los distintos tweets que han emitido los usuarios desde los días previos al estallido social, tomando como referencia los primeros eventos de “evasión del pasaje del metro” (17-10-2019).
- Se recopila información histórica desde el 01-01-2016 hasta el 22-11-2019 de los indicadores: Valor del Dólar, IPSA y Valor del Cobre. Con el objetivo de estudiar los comportamientos previos a los conflictos de estos indicadores y, en caso de ser necesario, establecer comparativas entre mismos periodos de diferentes años.
- Debido a complejidad de procesamiento y de arquitectura de modelos, se asumirán ciertas hipótesis que deberán ser probadas o corregidas posteriormente. En particular no se abordará la detección de sarcasmo y no se profundizará en la precisión semántica o comportamental de la detección de sentimientos, asumiendo que datos que se comporten como falsos positivos no son significativos para el volumen de datos analizado. El volumen de datos también será restringido especialmente por capacidades del hardware disponible para este estudio y las limitaciones de la API de Twitter.

## Bosquejo esquemático:

Para poder cumplir los objetivos planteados se proponen las siguientes etapas:

1. Recopilación de datos
2. Procesamiento de datos
3. Exploración de datos
4. Modelado de datos
5. Presentación y automatización de datos.

El esquema simplificado sobre la captura de datos



## Recopilación de datos:

Una vez fijados los objetivos, y para poder concretar el análisis planteado, el paso siguiente será establecer la fuentes de los datos. Por las características del problema, se definirán fuentes externas y públicas de datos.

De las redes sociales actuales se elige Twitter debido a que sus datos son en su mayoría públicos, y determinan la tendencia o “sentimiento” de quienes los emiten.

Los datos a utilizar pertenecen principalmente a dos categorías: datos de texto y datos numéricos financieros. Debido a sus diferentes naturalezas, serán tratados con métodos diferentes.

Respecto a los datos de texto, se escogió como fuente la red social Twitter, de la cual se recopilaron datos sobre tweets emitidos entre el día 01 de octubre de 2019 y el 25 de noviembre de 2019. Como palabras claves se utilizaron las siguientes:

*"chile", "estallido", "social", "pacos", "constitucion", "DDHH", "dictadura",  
"pinochet", "represion", "derechos", "protesta", "marcha", "AFP", "salud",  
"educación", "ojos", "ocular", "disparar", "desperto", "justicia"*

Para recopilar los datos se usaron las funciones `search_tweets` de la librería `rtweet` y la función `searchTwitter` de la librería `twitterR`. Ambas librerías forman parte del ecosistema de R. El detalle de uso de estas funciones se puede encontrar en el script de R Markdown que acompaña a este informe, o accediendo a [este sitio de GitHub](#).

Para obtener los datos relativos al impacto económico, elegimos como indicadores al Valor del Dólar, al IPSA y al Valor del Cobre, de los cuales obtuvimos datos de frecuencia diaria para el periodo comprendido entre el 01-01-2016 hasta el día de hoy, para cada indicador. A continuación describimos las fuentes para cada conjunto de datos.

Dato	Fuente	Link
Valor del Dólar	Banco Central de Chile	<a href="https://si3.bcentral.cl/Bdemovil/BDE/Series/MOV_SC_TC1">https://si3.bcentral.cl/Bdemovil/BDE/Series/MOV_SC_TC1</a>
IPSA	Bolsa de Santiago	<a href="https://datosmacro.expansion.com/bolsa/chile">https://datosmacro.expansion.com/bolsa/chile</a> <a href="https://www.bolsadesantiago.com/detalle_indice/SP%20IPSA">https://www.bolsadesantiago.com/detalle_indice/SP%20IPSA</a>
Valor del Cobre	Ministerio de Minería, Comisión Chilena del Cobre	<a href="https://www.cochilco.cl/Paginas/Estadisticas/Bases%20de%20Datos/Precio-de-los-Metales.aspx">https://www.cochilco.cl/Paginas/Estadisticas/Bases%20de%20Datos/Precio-de-los-Metales.aspx</a>

## Procesamiento de datos:

Respecto a los datos de texto, utilizando las librerías de R **tidytext** (para manipulación de datos de tipo string), **dplyr** (para manipulación de tablas de datos) y **stopword** (para definición de palabras no elementales para análisis de sentimiento). Con estas librerías se realizó una limpieza de las palabras no elementales o adjuntivas, que no son útiles para el análisis de sentimiento y no tienen mayor valor semántico. La librería **stopword** otorga una interfaz con la API de Twitter. Es importante considerar algunas limitaciones. La API de Twitter gratuita tiene la limitación de entregar sólo los 9 días anteriores a la fecha de consulta, la idea propuesta era realizar un análisis de la primera semana del estallido, pero terminamos analizando desde el 20-11-2019 debido a esta limitación. De cualquier manera creemos que esta limitación no impide mostrar este desarrollo como prueba de concepto.

Dada la naturaleza string de la data, la transformación, ETL y agregación de datos numéricos se acota a la frecuencia de palabras clave usadas. Además de la limpieza de palabras no elementales, también se realizó eliminación de vínculos, emoticones, puntuaciones, símbolos especiales, entre otros elementos no literales. Vale notar que los emoticones se podrían considerar para la evaluación de sentimiento en un desarrollo futuro, pero no lo consideraremos en este estudio.

El procesamiento de los datos de los indicadores económicos no requiere de mucha limpieza o pre-procesamiento, ya que los datos se obtienen de bases estadísticas públicas y oficiales, por lo que ya se encuentran procesadas y limpias. En esencia, el tratamiento de estos datos se redujo a su compilación, confirmación de formato fecha y número decimal cuando correspondiese, y revisión de que no existiesen datos corruptos o de un tipo no correspondiente a la definida en su columna. Los datos mínimos a utilizar consisten en fechas y en datos numéricos.

Respecto al análisis de sentimientos, se utiliza la librería **sentimentr** de R, que clasifica en un puntaje si el sentimiento es positivo o negativo, mediante la función **sentiment\_by** aplicada a un vector de datos de texto. Esta asigna un puntaje promedio de sentimiento para cada tweet, el cual puede ser contrastado posteriormente.

## **Análisis descriptivo y exploratorio**

Para poder comprender de manera profunda la base de datos se explorará la información en búsqueda de patrones, correlaciones y desviaciones mediante estadística descriptiva y gráficos de apoyo.

### **Indicadores Económicos**

Se realizan diferentes gráficos con los tres indicadores económicos anteriormente mencionados: Valor de Dólar, Precio del Cobre y IPSA, este último se grafican dos vistas, en el primero se puede apreciar el comportamiento de los puntos de cierre, y en el segundo se pueden observar las variaciones que tiene este en el tiempo.

Para generar los gráficos se utiliza el programa de visualización Tableau. Se escoge el gráfico de línea para cada uno de los indicadores, de esta manera observar el comportamiento en el tiempo. Para destacar la conducta de cada indicador desde que se produce el “Estallido Social” (17-10-2019) se resalta la línea en rojo y además se agrega una línea de tendencia como referencia.



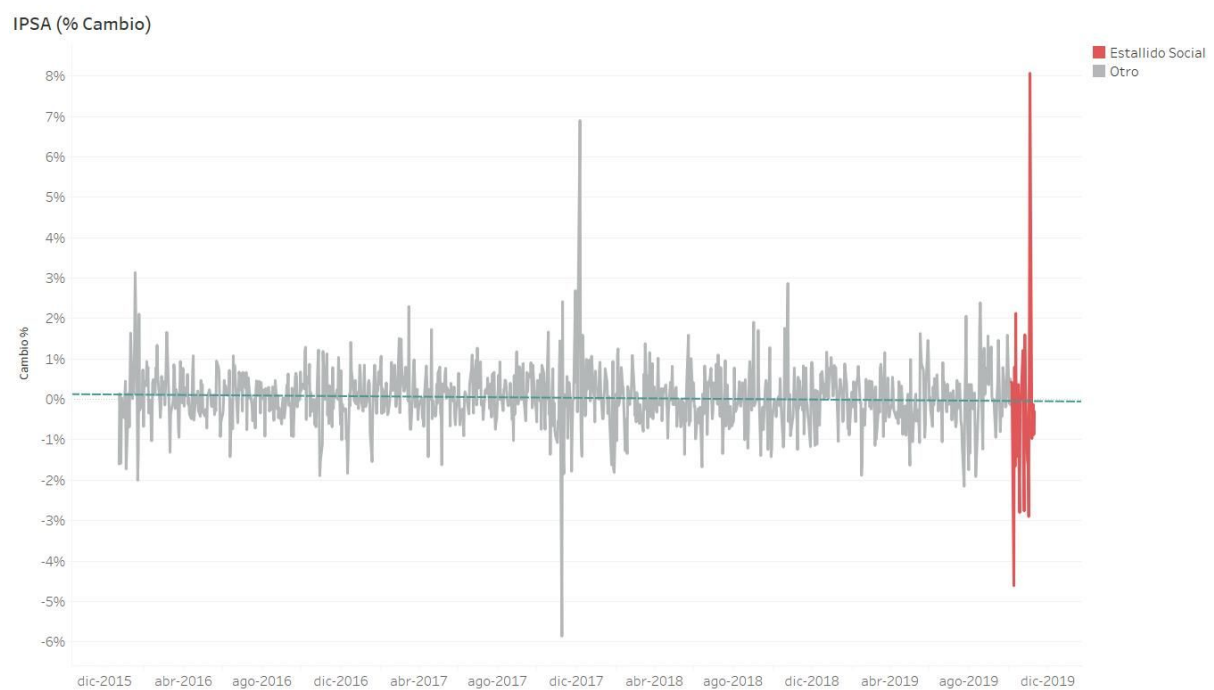
Se observa que el Valor del Dólar tuvo una fuerte baja en comienzo del año 2018, en donde ya ha mediados del 2018 comenzó a estabilizarse, y a principios del 2019 ya se observa una tendencia al alza. Sin embargo, se observa un peak en el “Estallido Social” que supera los 3 años anteriores.



El precio del cobre ha tenido bastantes variaciones a través de los últimos tres años, sin embargo se observa una baja sobre la tendencia a principios del 2018. Para el momento del estallido social no se aprecia mayor cambio o variación significativa.



Los puntos de cierre del IPSA se mantuvieron sobre la tendencia entre el 2017 y 2018, a comienzo del 2019 se observa que esto comienza a cambiar a la baja. Se puede apreciar que en el “Estallido Social” se obtienen los puntos de cierre más bajos del 2019.

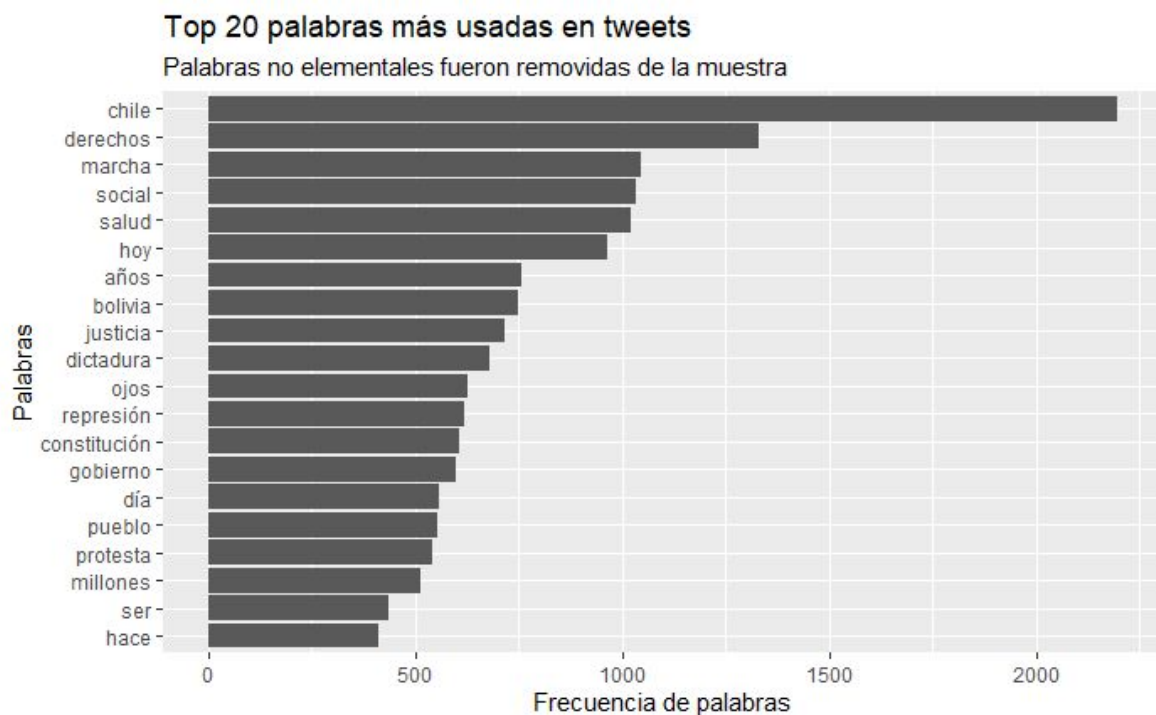




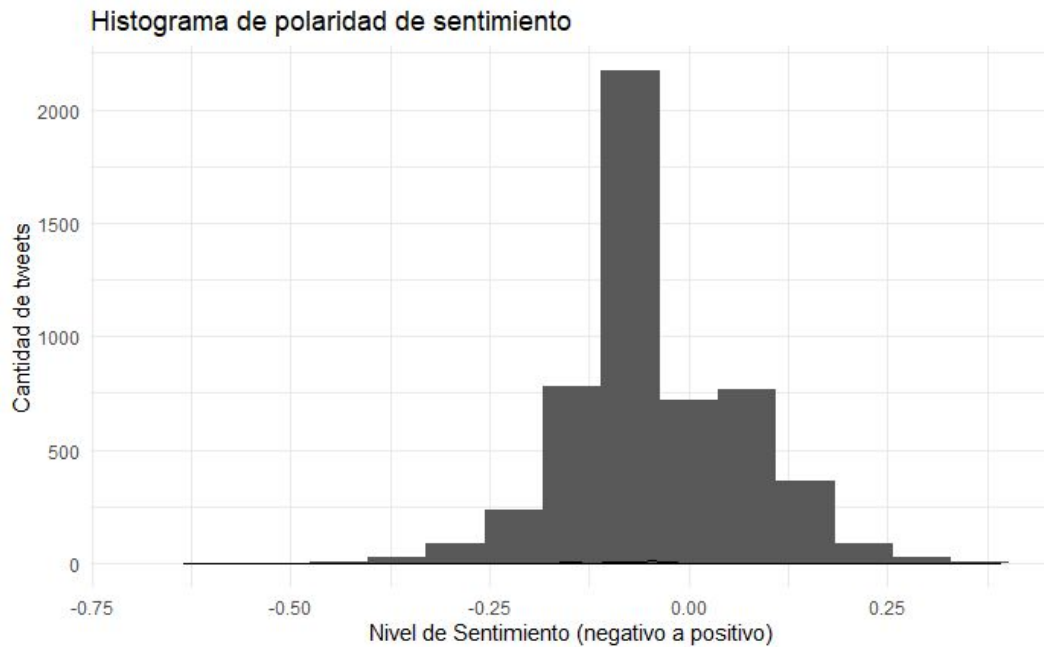
Los cambios del IPSA se mantienen relativamente estables, salvo en dos fechas que se observan, finales del 2017 y en el “Estallido Social”. La vista de este indicador, podría tener un segundo análisis para comparar los sucesos que ocurrieron en las fechas que se observan cambios bruscos.

### **Tweets y Análisis de Sentimiento**

En primer lugar, podemos observar las 20 palabras más usadas en los tweets recopilados. Esto nos muestra las temáticas más comunicadas en esa red social, como reflejo de lo que está ocurriendo en la sociedad chilena.



Respecto al análisis de sentimiento, un primer análisis muestra que los sentimientos, desde la perspectiva de polaridad positiva o negativa, no se alejan demasiado del puntaje neutro con un leve sesgo a sentimientos negativos, lo que se puede ver reflejado en el histograma:



Esto también puede ser indicador de que hace falta un mayor tratamiento de los datos de texto por tratarse de texto en lenguaje español, cuando comúnmente, las librerías dedicadas a análisis de sentimiento están construidas para procesar texto en español

## Propuesta para próximos pasos

En etapas futuras del proyecto se propone, mediante los resultados obtenidos en el análisis descriptivo y exploratorio, generar un modelo que permita vincular y detectar la relación que existe entre los datos de análisis de sentimientos de twitter y los indicadores económicos definidos en el alcance. Para ello, se utilizarán las variables determinantes y se confirmará que estas sean realmente influyentes mediante métodos tales como la regresión lineal u otros recomendados en la literatura como los Support Vector Machines (SVM). Una vez ejecutado el modelo, este permitiría categorizar algunos conceptos o sentimientos como mejores predictores o co-variables de los indicadores económicos de interés, y se avanzará en evaluar si la hipótesis inicial (que el sentimiento social informa a la probabilidad de éxito de inversión) es corroborada, para luego continuar con tareas de perfeccionamiento y aumento de precisión y efectividad del modelo.

Creemos que, en tiempos como este, contar con herramientas que permitan reducir la incertidumbre social y económica de las personas permiten contribuir a desarrollar y avanzar en los cambios sociales que el país necesita. Al mismo tiempo estamos convencidos que el valor multiplicativo que tiene la ciencia de datos es un elemento catalizador de cambio positivo en la sociedad, sobre todo en hitos tan significativos para la construcción de la futura historia de nuestro país.