

Banco de México
Documentos de Investigación

N° 2022-18

**Índices de Sentimiento e Indicadores de Actividad
Económica en México 2016-2021**

Leonardo Torre
Banco de México

Eva González
Banco de México

Ramón Casillas
Banco de México

Jorge Alvarado
Banco de México

Diciembre 2022

La serie de Documentos de Investigación del Banco de México divulga resultados preliminares de trabajos de investigación económica realizados en el Banco de México con la finalidad de propiciar el intercambio y debate de ideas. El contenido de los Documentos de Investigación, así como las conclusiones que de ellos se derivan, son responsabilidad exclusiva de los autores y no reflejan necesariamente las del Banco de México.

The Working Papers series of Banco de México disseminates preliminary results of economic research conducted at Banco de México in order to promote the exchange and debate of ideas. The views and conclusions presented in the Working Papers are exclusively the responsibility of the authors and do not necessarily reflect those of Banco de México.

Índices de Sentimiento e Indicadores de Actividad Económica en México 2016-2021*

Leonardo Torre[†]
Banco de México

Eva González[‡]
Banco de México

Ramón Casillas[§]
Banco de México

Jorge Alvarado^{**}
Banco de México

Resumen: Este trabajo utiliza, por vez primera, información en formato de texto proveniente de 9,802 entrevistas realizadas entre enero de 2016 y enero de 2021 del Programa Trimestral de Entrevistas a Directivos, empleado en la elaboración del Reporte sobre las Economías Regionales del Banco de México, para estimar índices de sentimiento regionales y nacionales. Estos índices son posteriormente asociados con diferentes indicadores, "suaves" y "duros", de actividad económica. Los resultados muestran correlaciones positivas y estadísticamente significativas entre ambos tipos de indicadores, principalmente en el nivel nacional, sugiriendo que la información en formato de texto obtenida en el Programa Trimestral de Entrevistas a Directivos puede ser útil para complementar la que proveen indicadores tradicionales de actividad económica.

Palabras Clave: Análisis de Sentimientos, Análisis Regional, México

JEL Classification: C45, R11, R15

Abstract: This paper uses, for the first time, information in text format from 9,802 interviews performed between January 2016 and January 2021 from the Programa Trimestral de Entrevistas a Directivos, employed to elaborate Banco de México's Regional Economic Report, to estimate regional and national sentiment indexes. These indexes are next associated with different "soft" and "hard" data of economic activity. The results show positive and statistically significant correlations between both types of indicators, mainly at the national level, suggesting that data in text format contained in the Programa Trimestral de Entrevistas a Directivos can be useful to complement the information provided by traditional indicators of economic activity.

Keywords: Sentiment Analysis, Regional Analysis, Mexico

*Agradecemos a Víctor Muñiz, Hairo Miranda y Tatiana Rueda, de CIMAT-Monterrey por sus consejos y asesoría durante la elaboración de este documento; así como a Alejandro Noriega, Diana Martínez, Ramsés Franco y Eduardo Sandoval por su apoyo a lo largo de este proyecto. Agradecemos también los comentarios de Caterina Rho, Raúl Fernández, Alejandrina Salcedo y dos dictaminadores anónimos. Todos los errores son responsabilidad de los autores.

[†] Dirección General de Investigación Económica. Correo: leonardo.torre@banxico.org.mx.

[‡] Dirección General de Investigación Económica. Correo: egonzalezg@banxico.org.mx.

[§] Dirección General de Investigación Económica. Correo: lecasillas@banxico.org.mx.

^{**} Dirección General de Investigación Económica. Correo: jorge.alvarado@banxico.org.mx.

1. Introducción

El Banco de México permanentemente realiza encuestas y entrevistas en el nivel nacional a diversos agentes económicos con el objetivo de obtener información oportuna sobre las condiciones del ambiente de negocios en los diferentes sectores productivos. Una de estas fuentes de información es la derivada del Programa Trimestral de Entrevistas a Directivos Empresariales (PED) que el Banco de México efectúa de manera telefónica o en persona, a directivos de empresas, así como a representantes de organismos empresariales (o *stakeholders*) del país para elaborar su publicación trimestral *Reporte sobre las Economías Regionales* (RER).¹ Las entrevistas del PED, implementadas a partir de 2016 en los meses de enero, abril, julio y octubre de cada año, capturan información cualitativa y en formato de texto sobre i) Actividad Económica, ii) Perspectivas y iii) Riesgos para la Actividad Económica de las regiones en las cuales el Banco de México divide al país (norte, centro norte, centro y sur).^{2,3} Actualmente, la información cualitativa capturada en el PED es empleada para generar índices de difusión y porcentajes de respuesta en el nivel regional que se presentan en el RER, en tanto que la información en formato de texto ha sido utilizada principalmente para dar contexto al comportamiento de diversos indicadores construidos con datos “duros” y datos “suaves” del RER.⁴

Si bien dar contexto al comportamiento de los indicadores cualitativos y cuantitativos representa ya una contribución de la información en formato de texto al RER, también existe abundante literatura, identificada como Análisis de Sentimientos, que muestra que la

¹ Conviene señalar que la información obtenida de las fuentes entrevistadas en los PED es tratada de manera estrictamente confidencial por el Banco de México.

² En realidad, los PED correspondientes al RER del primer trimestre de un año determinado se realizan generalmente en marzo-abril de ese mismo año; el correspondiente al RER del segundo trimestre entre los meses de junio y julio; el correspondiente al RER del tercer trimestre se lleva a cabo entre septiembre y octubre, en tanto que el del cuarto trimestre se realiza en el mes de enero siguiente al cierre del trimestre señalado. La mayor parte de las entrevistas, no obstante, se lleva a cabo en enero, abril, julio y octubre, por lo que en adelante haremos referencia a estos meses para indicar el periodo de realización del PED utilizado para elaborar su correspondiente RER.

³ La región norte incluye: Baja California, Chihuahua, Coahuila, Nuevo León, Sonora y Tamaulipas; el centro norte considera: Aguascalientes, Baja California Sur, Colima, Durango, Jalisco, Michoacán, Nayarit, San Luis Potosí, Sinaloa y Zacatecas; el centro lo integran: Ciudad de México, Estado de México, Guanajuato, Hidalgo, Morelos, Puebla, Querétaro y Tlaxcala, y el sud está compuesto por: Campeche, Chiapas, Guerrero, Oaxaca, Quintana Roo, Tabasco, Veracruz y Yucatán.

⁴ Los “datos duros” (*hard-data*) hacen referencia a información constituida por variables objetivas y directamente cuantificables, como producción, empleo, o precios. Por su parte, los “datos suaves” (*soft-data*) hacen referencia a datos cualitativos que son posteriormente cuantificados y se obtienen comúnmente mediante de encuestas o entrevistas para medir, por ejemplo, la satisfacción de consumidores en relación con un bien o servicio, o revisar su calidad. Vea Stsiopkina (2022).

información en formato de texto refleja sentimientos o actitudes de individuos, empresas, grupos especializados, etc., en torno a condiciones económicas presentes o futuras, por lo que un procesamiento adecuado de la misma puede ser útil para la toma de decisiones. De aquí que en la actualidad existan y continúen desarrollándose metodologías de análisis de sentimientos encaminadas a extraer de la información en formato de texto, como la que ofrece el PED, índices numéricos que puedan ser comparados con indicadores tradicionales de actividad económica (producción, empleo, inversión, etc.) a fin de determinar si entre ellos existe algún tipo de asociación. Más aún, dada la naturaleza de la información en formato de texto que pudiera utilizarse en su elaboración, esos índices numéricos, también llamados “índices de sentimiento”, pueden resultar más oportunos que los indicadores económicos tradicionales basados en datos duros o suaves.

Tomando en cuenta lo expuesto arriba, este documento ofrece, de acuerdo con nuestro conocimiento, las primeras estimaciones de índices de sentimiento a partir de la información en formato de texto contenida en el PED. En particular, se calcularon *índices de sentimiento ponderados* derivados a partir de las respuestas en formato de texto obtenidas de los directivos entrevistados en torno a la situación económica pasada, presente y futura, tanto de sus empresas como de las entidades federativas en las que estas se ubican.⁵ Para obtener estos índices, cada una de las respuestas en formato de texto se clasificó en una de tres categorías, dependiendo de si transmitían un sentimiento positivo, un sentimiento negativo o un sentimiento neutral (o no definido). Esta clasificación, o etiquetado, se realizó de manera manual, así como con el apoyo de tres algoritmos de Aprendizaje de Máquina (Máquina de Soporte Vectorial, Redes Neuronales Recurrentes y Representación de Codificador Bidireccional de Transformadores).

El trabajo utiliza la información en formato de texto de 9,802 entrevistas realizadas a directivos y representantes empresariales de las cuatro regiones del país en los PED de enero de 2016 a enero de 2021, de las que se desprendieron 76,895 documentos. Aquí, un “documento” se refiere a una respuesta en formato de texto que un directivo entrevistado

⁵ Los índices se ponderan con base en la participación del PIB de un sector en una región (o en el nivel nacional) determinada en el PIB total de la región (o del país). Esta ponderación busca ajustar el peso de las respuestas por la relevancia del sector al que pertenece el documento contabilizado en el índice respectivo. Esto se explica con mayor detalle en la sección 4.

ofrece a una pregunta de la entrevista, lo que implica que de una entrevista a un directivo pueden surgir varios documentos.

Adicionalmente, se investiga en qué medida estos índices de sentimiento ponderados se asocian con diferentes indicadores de actividad económica regional y nacional publicados por el INEGI. La relevancia de este ejercicio reside en que los índices de sentimiento se pueden generar tan pronto se captura la información de un PED, mientras que algunos indicadores del INEGI se publican normalmente con al menos un mes de rezago. Así, este ejercicio permite evaluar la capacidad de los primeros para complementar la información provista por los indicadores oportunos del INEGI que se publican con rezago, o bien, reforzar la señal que deriva de indicadores que se publican sin retraso.

La relación entre los índices de sentimiento ponderados por región y nacionales, con los indicadores del INEGI se capturó mediante de coeficientes de correlación de Pearson. Así, se obtuvieron correlaciones entre los índices de sentimiento ponderados con dos indicadores mensuales suaves de actividad económica nacional: el Indicador de Pedidos Manufactureros y el Indicador de Confianza Empresarial del Sector Manufacturero; cuatro indicadores duros nacionales: las tasas de crecimiento trimestral del PIB real, del Indicador de Actividad Industrial Total y del Indicador de Actividad Industrial Manufacturera, además del crecimiento mensual del Indicador Global de la Actividad Económica, y dos indicadores duros regionales: las tasas de crecimiento trimestral del Indicador de la Actividad Económica Regional y del Indicador Regional de la Actividad Manufacturera.

Entre los principales resultados del trabajo destacan (i) que los valores de los índices de sentimiento ponderados nacionales obtenidos con los diferentes etiquetados muestran patrones en el tiempo muy similares entre ellos, lo cual se refleja en coeficientes de correlación positivos y estadísticamente distintos de cero; (ii) los índices de sentimiento regionales muestran patrones en el tiempo similares a los nacionales, si bien las correlaciones de estos al interior de las regiones no son tan fuertes como las nacionales; (iii) los índices de sentimiento nacionales muestran correlaciones positivas y elevadas con los dos indicadores oportunos nacionales de actividad económica; (iv) los índices de sentimiento nacionales se correlacionan positivamente con los indicadores duros de actividad económica nacionales a

niveles de significancia del 15%, con excepción de las correlaciones con la tasa de crecimiento trimestral del IGAE, donde las correlaciones no alcanzan este último nivel de significancia, y (v) las correlaciones entre los índices de sentimiento regionales y los dos indicadores duros regionales de actividad son positivas y distintas de cero a niveles de significancia del 15% en casi todos los casos de la región norte; no así en el resto de las regiones. Los resultados previos sugieren que los índices de sentimiento estimados en este trabajo, y en particular los que utilizan la información nacional, pueden ser útiles para señalar, de una manera más oportuna, el comportamiento tanto de la actividad económica en general, como del manufacturero en particular.

El trabajo se organiza como sigue. La sección 2 presenta una breve reseña del enfoque de análisis de sentimientos y de la utilización de técnicas de aprendizaje de máquina para generar índices de sentimiento. La sección 3 revisa la metodología utilizada para estimar los índices de sentimiento que se presentan en este trabajo. La sección 4 presenta los índices de sentimiento estimados y sus respectivas correlaciones con los ocho indicadores de actividad económica regional y nacional. La sección 5 presenta comentarios finales.

2. Indicadores de Sentimientos Basados en Texto

La aparición y el desarrollo de las redes sociales, al facilitar la generación y comunicación de conocimiento y experiencias, el intercambio de opiniones, y la creación de métodos más eficientes para conservar y procesar esa información, generaron incentivos para que empresas, proveedores de bienes y servicios, políticos, investigadores, entre otros actores, intentaran aprovecharla para su toma de decisiones (D'Andrea et al. 2015). Un porcentaje significativo de esta información se caracterizó, no obstante, por ser información en formato de texto, poco sistematizada, y sobre la cual las metodologías econométricas tradicionales no podían aplicarse. De aquí que se haya impulsado la adopción de técnicas capaces de aprovechar esta información, siendo una de ellas, el análisis de sentimientos.⁶ El análisis de sentimientos se define como aquellos procesos o métodos que permiten detectar el

⁶ El análisis de sentimientos surge, básicamente, con Turney (2002) y Pang et al. (2002). En el primer caso, aplicado a reseñas de restaurantes, automóviles, bancos y destinos turísticos. En el segundo, en el análisis de revisiones de películas. Otros trabajos precursores son Nasukawa y Yi (2003) y Kim y Hovy (2004).

sentimiento positivo o negativo contenido en un texto, es decir, en una frase o una palabra. Para detectar dicho sentimiento, estos métodos se apoyan en la polaridad de una frase o una palabra (positiva, negativa o neutral), aunque también pueden capturar sentimientos y emociones (felicidad, enojo, tristeza, etc.), o incluso urgencia (urgente, no urgente).

En la medida que los métodos para el análisis de sentimientos se desarrollaron, sus campos de aplicación también se extendieron y hoy su uso se aprecia en ámbitos tan diversos como las ciencias computacionales, las ciencias sociales, las ciencias administrativas, los negocios, así como en bancos centrales. Su aplicación fundamental ha sido extraer información, a partir de expresiones subjetivas, del sentimiento de palabras, oraciones subjetivas y tópicos.⁷ El objetivo final de este campo de análisis es extraer, a partir de información en formato de texto, indicadores numéricos, o índices de sentimiento.⁸

Dos definiciones básicas de índices de sentimiento se presentan en (1) y (2):

$$(1) \quad \text{Índice de Sentimiento } A = \frac{\text{Positivos} - \text{Negativos}}{\text{Positivos} + \text{Negativos}}$$

$$(2) \quad \text{Índice de Sentimiento } B = \frac{\text{Positivos} - \text{Negativos}}{\text{Positivos} + \text{Negativos} + \text{Neutrales}}$$

Estas definiciones se basan en los conteos de los documentos obtenidos mediante de algoritmos, donde *Positivos* se refiere a la cuenta total de documentos que de acuerdo con los criterios de clasificación transmiten un sentimiento positivo; *Negativos* se refiere a la

⁷ D'Andrea et al. (2015), p.26.

⁸ Es importante distinguir entre un “índice de sentimiento” y un “indicador de confianza.” De acuerdo con Santero y Westernlund (1996), los “indicadores de confianza” se desprenden de encuestas sencillas y breves a agentes económicos (individuos, hogares, empresas), caracterizadas por un reducido número de preguntas relacionadas con el sentir de los agentes consultados en torno a tendencias recientes, la situación actual o sus expectativas para el comportamiento de corto, mediano y largo plazos de diferentes variables. Estas preguntas requieren, en general, respuestas cualitativas como “mejor”, “sin cambio” o “peor. Las respuestas a cada pregunta se asocian, a su vez, con un valor (por ejemplo, 1, 0 o -1, cuando se tienen solo tres opciones de respuesta) lo que permite obtener distribuciones de frecuencia y generar “índices”, siendo estos últimos, por lo general, balances simples que resultan de sustraer, por ejemplo, el número de respuestas “peor” al número de respuestas “mejor” en una pregunta. Esto permite presentar, mediante un solo valor, el resumen de las respuestas a cada pregunta, y obtener una representación en el tiempo de los cambios de esas respuestas. Las series de tiempo de los indicadores de confianza así obtenidas han sido tradicionalmente contrastadas con series de tiempo de indicadores duros de actividad económica, como PIB, inversión, consumo, empleo, etc. Para ejemplos de cálculos de “indicadores de confianza” y sus aplicaciones, ver Garret et al. (2004), De Bondt y Schiaffi (2015), Salhin et al. (2016) y Benhabib y Spiegel (2017). Para el caso de México, ver Díaz y Huerta (2020).

cuenta total de documentos clasificados con sentimiento negativo, y *Neutrales* al total de documentos clasificados con sentimiento neutral (o bien, que no pueden clasificarse).

Existen diferentes métodos para generar índices de sentimiento a partir de información en formato de texto, entre los que destacan: (i) el análisis de textos basado en lexicones,⁹ (ii) el aprendizaje de máquina,¹⁰ y (iii) los enfoques híbridos.¹¹ Enseguida se explica brevemente en qué consiste el segundo enfoque, ya que algunas técnicas pertenecientes al mismo serán utilizadas para clasificar los documentos.

Es conveniente mencionar aquí que una vez obtenidos estos índices, se les comparará con índices que se desprenden de conteos parciales o totales de los documentos clasificados de manera manual por los investigadores interesados, como referencia de qué tan aceptable es la clasificación obtenida de los algoritmos utilizados. El diseño de un buen algoritmo debería generar una correlación elevada entre los índices que de él se desprenden y los realizados de manera manual (también llamados “anotados”).

2.1 Algoritmos de Aprendizaje de Máquina

El aprendizaje de máquina hace referencia a algoritmos enfocados principalmente en predecir, clasificar, agrupar o generar clústeres de datos, siendo particularmente útiles cuando se tiene gran cantidad de información en formato de texto cuya clasificación manual resulta muy costosa.¹² En estos algoritmos el problema se aborda, en términos generales, como un problema de clasificación, en el que el *clasificador* es provisto de texto y regresa una categoría, por ejemplo, “positivo”, “negativo” o “neutral.”¹³

Los algoritmos se clasifican en cuatro tipos: *aprendizaje supervisado*, *aprendizaje no supervisado*, *semi-supervisados* y *por refuerzo*.¹⁴ En este trabajo, no obstante, solo

⁹ Para una revisión del enfoque basado en lexicones, ver Turney (2002) y Hatzivassiloglou y McKeown (2002).

¹⁰ Ver Medhat et al. (2014) para una exposición sobre aprendizaje de máquina.

¹¹ Ver Prabowo y Thelwall (2009) para una exposición sobre el enfoque híbrido.

¹² Es oportuno mencionar que el aprendizaje de máquina es, a su vez, solo una de una variedad de técnicas de la llamada Minería de Datos, siendo esta última la disciplina encargada de buscar estructuras desconocidas en un conjunto de datos. Ver Arora (2021) y Witten et al. (2017).

¹³ Un “clasificador” es el algoritmo que permite ordenar o catalogar de manera automática los datos en una o más clases.

¹⁴ Vea Alloghani et al. (2020) para la distinción entre algoritmos de aprendizaje supervisado y no supervisado; Weng (2021) presenta una descripción de aprendizaje semi-supervisado, en tanto que Sutton y Barto (2018) tratan con detalle el aprendizaje por refuerzo.

utilizaremos algoritmos pertenecientes a las primeras dos categorías. Los algoritmos de *aprendizaje supervisado* “utilizan un conjunto de características X para predecir un resultado Y . Aquí, el término *predicción* se refiere a que con un conjunto de datos observados y etiquetados, tanto de X como de Y a los que se les denomina datos de entrenamiento, se desea anticipar los valores de Y contenidos en un conjunto de datos de prueba independientes, con los valores observados de X en el conjunto de prueba.¹⁵ En otras palabras, el objetivo es construir una función $\hat{\mu}(x)$, que sea un estimador de $\mu(x) = E[Y|X = x]$ que haga un buen trabajo para predecir los verdaderos valores de Y en un conjunto independiente de datos.” [Athey 2018, p.4, traducción de los autores]

Por su parte, “los algoritmos de aprendizaje no supervisado se orientan a encontrar clústeres de observaciones que son similares en términos de sus características, lo que puede interpretarse como una reducción de dimensiones, y es principalmente utilizado para clasificar video, imágenes y tópicos en información en formato de texto... En general, el producto final de esta clase de modelos es una partición de un conjunto de observaciones, donde las observaciones dentro de cada partición son similares de acuerdo con alguna métrica, o con un vector de probabilidades o ponderaciones que describen una mezcla de tópicos o grupos a los que pudiera pertenecer una observación. Estos algoritmos reciben el calificativo de no supervisados ya que no requieren de una clasificación inicial de la información, sino que es el algoritmo el que encuentra las categorías con base en el reconocimiento de patrones o anomalías en los datos.” [Athey 2018, p.3, traducción de los autores].

2.2 Algunas Aplicaciones en Bancos Centrales

En cuanto al uso específico de estas herramientas en algunos de los principales bancos centrales del mundo puede mencionarse, por ejemplo, el trabajo de Suss y Trietel (2019) del Banco de Inglaterra. Estos autores utilizan un algoritmo de aprendizaje de máquina supervisado (árboles de decisión aleatorios) para desarrollar un sistema de alerta temprana

¹⁵ Las observaciones se suponen independientes, además de que se supone que la distribución conjunta (X, Y) en el conjunto de entrenamiento es la misma que en el conjunto de prueba. Estos dos supuestos son los que, en general, se requieren para que la mayoría de los métodos de aprendizaje de máquina funcionen (Athey, 2018).

en torno a vulnerabilidades en el sistema bancario de Inglaterra el cual, de acuerdo con su evaluación, genera mejores resultados que los obtenidos mediante métodos de regresión.

En el caso de Estados Unidos, Pinto (2019) se apoya en una versión modificada del algoritmo de bolsa de palabras para clasificar comentarios vertidos en las Encuestas Manufacturera y de Servicios del Quinto Distrito del FED de Richmond, y obtener índices de sentimiento que posteriormente correlaciona con un índice de difusión compuesto de actividad económica provisto en las encuestas. El autor reporta que, en general, los índices de sentimiento y el índice de difusión tienden a comportarse en la misma dirección.

Uno y Adachi (2019), del Banco Central de Japón, utilizan métodos de aprendizaje de máquina para identificar a directivos de empresas que afirman no tener una expectativa cuantitativa de inflación, siendo que existe la posibilidad de que sí la tengan. Una vez identificado a este grupo de directivos, utilizan “propensity score matching” para obtener una estimación contrafactual de esas expectativas de inflación, reportando que estas no son estadísticamente distintas de las que se obtienen de directivos encuestados que sí comparten una cifra de inflación esperada.¹⁶

Por su parte, Azqueta-Gavaldón et al. (2020), del Banco Central Europeo, utilizan técnicas de aprendizaje de máquina no supervisado para clasificar notas diarias de la prensa escrita de Alemania, Francia, Italia y España de enero de 2000 a mayo de 2019, lo que les sirve de base para construir diversos indicadores de incertidumbre. Encuentran que sus índices de incertidumbre capturan eventos de ese periodo como reformas laborales, ajustes fiscales, el sentido del voto del Brexit y tensiones geopolíticas con mayor anticipación que otros índices de incertidumbre o bien, que no fueron capturados por estos últimos.

En el caso de México, Rho, Fernández y Palma (2021) aplican técnicas de análisis de texto a mensajes de Twitter en español del periodo 2006-2019 para construir un índice de riesgo basado en el sentimiento para el sector financiero en México y lo comparan con distintos indicadores de estrés financiero, encontrando que su indicador de riesgo captura choques que no se reflejan en los índices existentes. También muestran que su índice se

¹⁶ Para una exposición de la técnica de “propensity score matching”, ver Rosenbaum y Rubin (1983).

correlaciona positivamente con medidas de riesgo financiero, volatilidad del mercado accionario, riesgo de default soberano y volatilidad cambiaria.¹⁷

En cuanto a la aplicación de estas técnicas en la disciplina económica en general, Korab (2021) muestra que el número de trabajos de investigación que usan indicadores apoyados en algoritmos de aprendizaje de máquina publicados en cuatro de las más influyentes revistas de Economía (*Quarterly Journal of Economics*, *American Economic Review*, *Econometrica* y *The Review of Economic Studies*) se elevó notoriamente a partir de 2010. Otras referencias que revisan aplicaciones de análisis de sentimientos utilizando técnicas de aprendizaje de máquina en Economía son Gentzkow et al. (2019) y Algaba et al. (2020).

3. Análisis de Sentimientos a partir del PED del Banco de México

La implementación de los algoritmos de aprendizaje de máquina involucra, en términos generales, los siguientes pasos: (1) identificar la base datos; (2) realizar un análisis exploratorio de la base de datos; (3) realizar el preprocesamiento de la información; (4) etiquetar información y elegir los modelos que se utilizarán para clasificar los documentos; (5) entrenar los modelos y evaluarlos, y (6) proceder con la predicción, es decir, clasificar con base en el modelo entrenado, los documentos que se encuentran fuera del conjunto de datos de prueba.¹⁸ Enseguida describimos, a la luz de estos pasos, el proceso para identificar los modelos de aprendizaje de máquina adoptados en este trabajo para generar el etiquetado necesario para calcular los índices de sentimiento.

¹⁷ Si bien la revisión de trabajos que aplican técnicas de aprendizaje de máquina al análisis en bancos centrales puede extenderse fácilmente, no debe dejar de mencionarse el trabajo de Doerr et al. (2021), quienes reportan los resultados de una encuesta implementada en 2020 por el Banco de Pagos Internacionales y que respondieron 52 representantes de bancos centrales de todas las regiones del mundo. La encuesta se enfocó en los retos y oportunidades que, en la práctica, enfrentan sus bancos centrales ante la disponibilidad de grandes bases de datos (*Big Data*) y la posibilidad de utilizar técnicas de aprendizaje de máquina para aprovecharlas. De acuerdo con estos autores, la mayoría de los bancos centrales ya discuten estos temas de manera formal a su interior, además de que ya aplican técnicas de aprendizaje de máquina en distintas áreas, como investigación, política monetaria y estabilidad financiera. También reportaron estar utilizando esa información para labores de supervisión y regulación. No obstante, destacaron como retos principales para sus bancos centrales la necesidad de mejorar la calidad de la información que capturan, la calidad en el muestreo y, por tanto, la representatividad de la información que generan con dicha información. De igual manera, los encargados de responder la encuesta mostraron preocupación en torno a temas de privacidad, protección y seguridad de la información que pudieran utilizar; en tanto que algunos entrevistados reportaron dificultades para establecer una infraestructura de Internet adecuada y de desarrollo de capital humano para darle servicio.

¹⁸ Ver Cuadro A1 del Apéndice para una descripción más detallada de estos pasos.

3.1 Base de Datos

El PED del Banco de México tuvo sus inicios en enero de 2011. Las entrevistas, como se mencionó previamente, se realizan con frecuencia trimestral en los meses de enero, abril, julio y octubre de cada año y constan, en general, de tres secciones: 1) Actividad Económica, 2) Perspectivas y 3) Riesgos.^{19,20} La primera recaba información sobre el desempeño de la empresa o sector económico durante el trimestre que termina; en la segunda se consulta a los directivos sobre sus perspectivas de producción a doce meses, y la tercera les consulta sobre los riesgos económicos con probabilidad de ocurrir en el corto plazo en la entidad federativa donde opera su empresa o el organismo que representan. En el periodo que cubre este trabajo, enero de 2011 a enero de 2021, se realizaron 19,364 entrevistas a directivos de empresas y a directivos de asociaciones empresariales (*stakeholders*) de las cuatro regiones en las que el Banco de México divide al país: norte, centro norte, centro y sur.²¹ Las entrevistas constan de preguntas cerradas, donde el entrevistado selecciona el comportamiento observado o esperado de variables como demanda, empleo, inversión y precios; así como de preguntas abiertas, donde se solicita al entrevistado respuestas en formato de texto relacionadas con factores que afectaron, afectan o que espera pudieran afectar, la actividad económica en su sector o entidad federativa, preguntas que, por cierto, han experimentado algunas modificaciones a lo largo del tiempo.²²

La información recabada en las entrevistas ha sido utilizada para (i) generar índices de difusión de actividad económica, y (ii) como complemento descriptivo en los análisis sectoriales y regionales publicados en el RER.²³ Las respuestas en formato de texto, o documentos, obtenidos a partir de las preguntas abiertas relacionadas con factores que

¹⁹ En el RER normalmente aparece una cuarta sección, a la que se le denomina Sección Especial, y en la que se solicita la opinión de los entrevistados sobre temas de coyuntura que pueden variar entre trimestres.

²⁰ Conviene tener presente que el RER del cuarto trimestre de un año determinado utiliza información del PED realizado en enero del año entrante; el RER del primer trimestre utiliza información del PED de abril, el RER del segundo trimestre se apoya en las entrevistas del PED de julio, y el RER correspondiente al tercer trimestre utiliza la información del PED realizado en octubre.

²¹ Ver nota al pie 2 para identificar las entidades federativas que integran cada una de las regiones y el Cuadro A2 del Apéndice para el conteo de entrevistas por región y tipo de entrevistado en el periodo señalado.

²² Ver Cuadro A3 del Apéndice para una descripción de las preguntas.

²³ En el RER, los sectores se clasifican en: 1) Agropecuario-Industria Alimentaria; 2) Comercio; 3) Minería, electricidad, agua y gas; 4) Construcción; 5) Manufacturas; 6) Transportes-Comunicaciones; 7) Otros Servicios y 8) Turismo. El Cuadro A4 del Apéndice presenta la relación de esta clasificación con el Sistema de Clasificación Industrial de América del Norte.

afectaron, afectan o afectarán, la actividad económica en su sector o entidad federativa han sido utilizados hasta la fecha como se indica en el punto (ii). La intención de este trabajo es utilizar estos documentos para construir índices de sentimiento.²⁴ Descartando la “no respuesta,” el universo de información disponible para la elaboración de este trabajo (enero de 2011 a enero de 2021) se integra de 116,197 documentos.

3.2. Análisis Exploratorio

En esta etapa se lleva a cabo el análisis exploratorio o de tópicos, que tiene como objetivo descubrir la estructura temática oculta en los documentos que se utilizarán en el análisis (Blei, 2011). Para este análisis se consideró la base de datos completa. Existen diversos algoritmos para identificar los tópicos subyacentes de los documentos, entre los que se encuentran el LDA (*Latent Dirichlet Allocation*) y el LSA (*Latent Semantic Analysis*). En este trabajo se probaron ambos modelos de detección de tópicos (LDA y LSA), obteniéndose una mejor detección con el segundo. Para determinar el algoritmo de identificación de tópicos a utilizar, se evaluó de forma subjetiva cuál de ellos generaba a partir del conjunto de datos, tópicos en los cuales pudiera percibirse una separación clara entre los temas que engloban el conjunto de palabras que compone cada uno de dichos tópicos. De igual forma, el número de tópicos se determinó probando con distintos valores hasta encontrar una separación aceptable. De esta evaluación de la base de datos, resultó una separación más clara de tópicos con LSA. Debe enfatizarse que el parámetro de la cantidad de tópicos lo determina el diseñador, lo que implica cierta subjetividad. Así, la decisión sobre cuántos tópicos calcular se realizó empíricamente, buscando identificarlos con base en palabras o frases utilizadas en Economía.²⁵ Así, como primer paso del análisis exploratorio se lematizaron las palabras (es decir, se agruparon de acuerdo con su significado) y se identificaron términos y secuencias de palabras que se repitieron con mayor frecuencia en el conjunto de documentos a utilizar. El Cuadro 1 muestra los 25 términos y frases más comunes, clasificados por unigramas,

²⁴ Cabe mencionar que en los casos de preguntas abiertas puede ocurrir que el entrevistado no ofrezca comentarios.

²⁵ El algoritmo LDA es un modelo estadístico que define una distribución probabilística sobre las variables observadas (palabras en los documentos) y las variables ocultas (la estructura temática o tópicos), y forma clústeres de documentos tomando en cuenta la presencia de cada palabra en cada uno de los documentos (ver Blei, 2011). Por su parte, el algoritmo LSA es un modelo matemático más intuitivo, capaz de descubrir un significado subyacente o latente en el texto utilizando una matriz término-documento y la reducción de dimensionalidad, obteniendo una representación más compacta que mantiene las propiedades semánticas, tanto de los documentos, como de los términos (ver Anandarajan, et al. 2019).

bigramas y trigramas (secuencias de una, dos y tres palabras), excluyendo previamente palabras funcionales presentes en los documentos. Enseguida se llevó a cabo el análisis de tópicos, que busca encontrar la mezcla de tópicos que conforma un documento. Este se realizó para todos los PED (enero de 2011 a enero de 2021) en cada una de las tres secciones de la entrevista: 1) Actividad Económica, 2) Perspectivas y 3) Riesgos. Los resultados por sección fueron similares, por lo que aquí presentamos los tópicos agrupando las tres secciones.

**Cuadro 1. Top 25 Palabras más Frecuentes en los PED
Enero 2011-Enero 2021**

a) Unigrama				b) Bigrama			
Palabra	Frecuencia	Palabra	Frecuencia	Palabra	Frecuencia	Palabra	Frecuencia
actividad	12,002	inversión	5,792	actividad económico	5,095	materia prima	728
año	9,563	mayor	5,709	cambio gobierno	4,485	mayor demanda	725
cambio	9,203	mercado	5,581	covid 19	1,563	nuevo gobierno	712
demanda	8,897	nuevo	5,182	demanda producto	1,446	nuevo proyecto	689
económico	8,727	poder	5,179	demanda servicio	1,412	obra público	642
empresa	8,343	precio	5,171	estados unidos	1,141	parte gobierno	632
esperar	7,950	público	5,159	gasto público	1,016	poder adquisitivo	631
estado	7,666	sector	5,142	gobierno estatal	1,005	reforma fiscal	621
estados	7,354	ser	5,124	gobierno federal	845	sector automotriz	617
gobierno	7,023	tener	5,101	incremento precio	829	tasa interés	602
incertidumbre	6,946	unidos	5,067	industria automotriz	814	tipo cambio	591
incremento	6,593	venta	5,043	inversión extranjero	778	volatilidad tipo	570
inseguridad	5,821			inversión privado	759		

c) Trigrama

Palabra	Frecuencia	Palabra	Frecuencia
alza tasa interés	563	inversión extranjero directo	145
comercial estados unidos	534	inversión público privado	142
demanda estados unidos	342	pandemia covid 19	138
depreciacion tipo cambio	333	parte estados unidos	133
depreciación tipo cambio	252	parte gobierno estatal	132
deterioro condición seguridad	229	parte gobierno federal	132
deterioro seguridad público	212	precio materia primo	123
economía estados unidos	179	tipo cambio afectar	121
económico estados unidos	173	tipo cambio alto	114
ejercicio gasto público	169	tipo cambio estable	113
estabilidad tipo cambio	153	vivienda interés social	113
hacia estados unidos	148	volatilidad tipo cambio	111
incremento tasa interés	145		

Nota: Como parte del preprocesamiento de la información se lematizaron las palabras. Por esa razón, aparecen términos como “actividad económico” en lugar de “actividad económica”.

Fuente: Elaboración propia con información del PED del Banco de México.

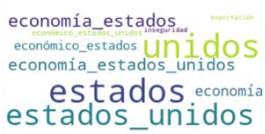
Los tópicos, a su vez, están formados por una distribución de palabras que aparecen en mayor o menor medida en cada uno de ellos. Como resultado, lo que se obtiene son n tópicos representados como una mezcla de palabras y con esto, los investigadores determinan cuáles son los tópicos más relevantes. El Cuadro 2 muestra los siete tópicos detectados, así como los términos más recurrentes en cada uno de los tópicos. Dichos tópicos se obtuvieron utilizando LSA, tomando en cuenta unigramas, bigramas y trigramas.

Adicionalmente, se realizó el análisis de tópicos de los comentarios por trimestre.²⁶ Este análisis permitió determinar tópicos que dominaron los comentarios en cada periodo, cómo variaron en el tiempo y cómo se relacionaron con la evolución de indicadores de actividad económica. La Gráfica 1 ilustra un ejemplo de cómo los tópicos pueden brindar contexto a un indicador, en este caso, al IGAE. La gráfica destaca, por ejemplo, como la relevancia de los tópicos cambia en el tiempo. Así, la reforma fiscal es el tópico más relevante en el 3T2013; un año más tarde dominan los relativos a la reforma energética y la economía de Estados Unidos, en tanto que la elección presidencial en Estados Unidos es el principal tópico en el 3T2016. En el 2T2018 el proceso electoral en México y la renegociación del TLCAN son los que resaltan; mientras que en el 1T2020 el tipo de cambio y el COVID-19 son los relevantes. Finalmente, en el 2T2020, cuando se contrae la actividad económica en México, el tópico COVID-19 es el que destaca; en tanto que la vacunación en México y la economía de Estados Unidos aparecen como los tópicos dominantes durante el 1T2021, acompañando así a la recuperación económica. Debe destacarse que al realizar el análisis de tópicos por región y trimestre, la distribución de términos no permitió identificar claramente tópicos relevantes. A partir del PED de enero de 2016, sin embargo, cambiaron las preguntas para

²⁶ La gráfica se construyó asociando el IGAE de un trimestre determinado, con los tópicos correspondientes al mes más cercano al cierre de dicho trimestre. Por ejemplo, al IGAE del cuarto trimestre de un año determinado se le asocian los tópicos identificados en las entrevistas del primer mes del año siguiente.

elaborar el RER del 4T2015, lo que elevó el número de respuestas por sección, permitiendo visualizar mejor tópicos regionales.^{27,28}

Cuadro 2. Tópicos Enero 2011 – Enero 2021

Tópico	Términos más recurrentes
1. Ninguno	
2. Seguridad Pública:	
3. Mercado Cambiario	
4. Obras Públicas/Inversión Pública	
5. Inversión	
6. Actividad Económica de Estados Unidos	
7. Incertidumbre	

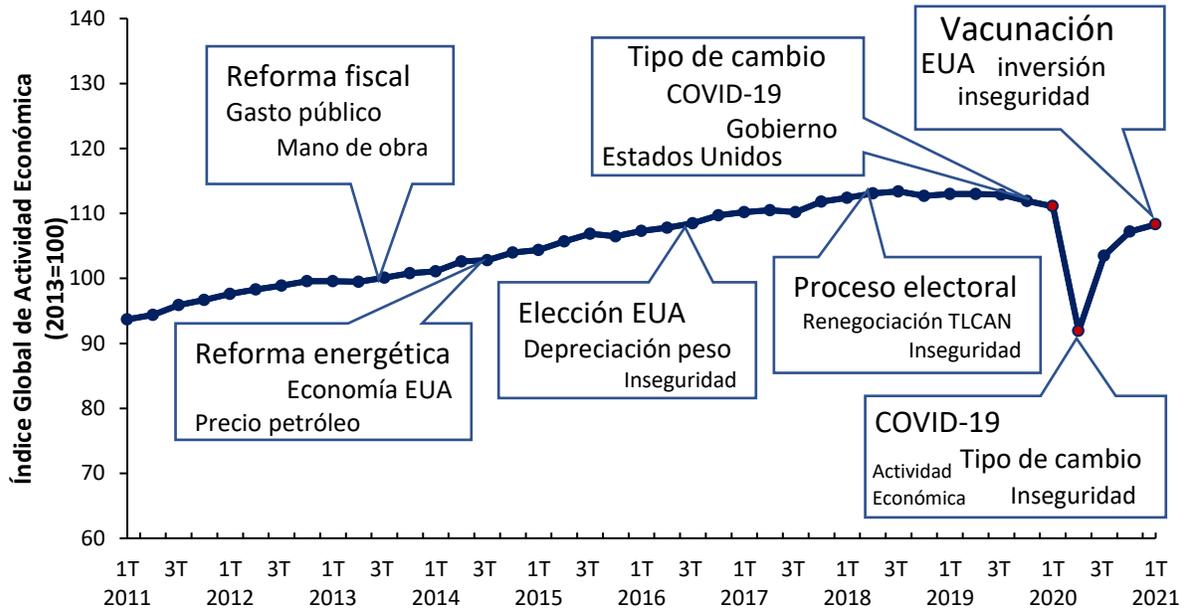
Nota: Se presentaron casos en los que los entrevistados no ofrecieron comentarios en las preguntas realizadas. Como se indica más adelante, estos casos recibieron una etiqueta “neutral” en la clasificación de la información. El tamaño de las palabras en las nubes de palabras corresponde a la frecuencia de la palabra en el tópico.

Fuente: Elaboración propia con información del PED del Banco de México.

²⁷ Si bien el número de entrevistas es similar en cada trimestre, el número de respuestas (o “documentos”) en el periodo 2016-2021 es aproximadamente el doble del periodo 2011-2015. Esto es así ya que la formulación de las preguntas se ha vuelto más detallada en años recientes. Por ejemplo, a partir del 4T2016 en la sección de actividad económica es posible identificar cuatro respuestas divididas en: i) factores externos de impulso, ii) factores internos de impulso, iii) factores externos limitantes y iv) factores internos limitantes. Ver Cuadro A3 del Apéndice.

²⁸ El Cuadro A5 del Apéndice presenta los tópicos relevantes por región en los PED de enero de 2016 a enero de 2021. Este no incluye los términos más recurrentes, pero están disponibles a solicitud del interesado.

Gráfica 1. IGAE y Tópicos Relevantes



Fuente: Elaboración propia con datos del PED de Banco de México e INEGI. El tamaño de las palabras en las nubes de palabras corresponde a la frecuencia de la palabra en el tópico.

3.3. Preprocesamiento de la Información

Antes de entrenar un modelo de aprendizaje de máquina se requiere realizar el preprocesamiento de la información. Este paso involucra realizar diversas modificaciones para eliminar el mayor ruido posible de las bases de información. Cabe mencionar, que los índices de sentimiento que se presentarán en la sección 4 utilizarán solo los documentos obtenidos de los PED enero de 2016 a enero de 2021. Para este trabajo se llevaron a cabo las siguientes acciones:

- Se realizó una corrección ortográfica de los textos.
- Se identificaron palabras en inglés que se tradujeron al idioma español.
- Se eliminaron palabras funcionales que, en sí mismas, carecen de información relevante para el estudio, como son artículos, pronombres y preposiciones (también conocidas en la literatura como *stop words*).
- Se eliminaron números y signos de puntuación.
- Se lematizaron las palabras, lo que significa agruparlas de acuerdo con su significado.
- Todas las letras mayúsculas se cambiaron a minúsculas.

- g) Se descartaron documentos que en sí mismos no contenían información relevante para el estudio; por ejemplo, documentos conformados por una sola letra o un punto. En este sentido, el total de los documentos de los PED de enero de 2016 a enero de 2021 se redujo de 77,256 a 76,895.

3.4. Etiquetado y Elección del Modelo para la Clasificación de los Documentos

Después de realizar el análisis exploratorio y preprocesar la información, procedimos a revisar diversas metodologías de aprendizaje de máquina para la predicción del sentimiento. Estos modelos requieren una base previamente clasificada, o de prueba, por lo que el primer paso en su implementación fue asignar etiquetas “Positivo”, “Negativo” o “Neutral” a cada uno de los documentos obtenidos en las entrevistas de los PED de enero de 2016 a enero de 2021.²⁹ Si bien los PED del RER iniciaron en el enero de 2011, optamos por etiquetar documentos de las entrevistas solo de los PED de enero de 2016 a enero de 2021 ya que en este periodo se elevó el número observaciones por región y las preguntas fueron más homogéneas en el tiempo. Esto permitió conformar una base de 76,895 documentos a partir de 9,802 entrevistas realizadas.

3.4.1 Asignación de Etiquetas

En la literatura no existe una metodología estándar para realizar el etiquetado manual de los documentos, requisito para elaborar tanto los índices anotados, como para utilizar los algoritmos de aprendizaje de máquina. En relación con esta asignación de etiquetas, algunos trabajos de investigación reportan haber etiquetado todos los documentos, o bien, una fracción de estos, mediante una clasificación manual con dos etiquetadores.³⁰ En ese trabajo, dos de los autores se encargaron de revisar y etiquetar, por separado, cada uno de los 76,895 documentos del periodo de enero 2016 a enero de 2021 en positivo, negativo y neutral, de

²⁹ Existen modelos más simples que entrenan lexicones para clasificar el sentimiento, como el utilizado por Pinto (2019). Al inicio de la investigación se hicieron algunas pruebas. No obstante, los modelos de clasificación basados en diccionarios son menos flexibles a la evolución del idioma y la estructura temática. En ese sentido, se optó por realizar la inversión de tiempo para generar nuestras propias etiquetas y entrenar modelos que pudieran utilizarse en futuros ejercicios trimestrales de entrevistas del Banco de México a bajo costo.

³⁰ Para una referencia en la que el etiquetado es realizado de manera independiente por dos investigadores, vea Aragón et al. (2019). Sobre el tema también consultamos, en enero de 2020, al Dr. Víctor Muñoz, catedrático del Centro de Investigación en Matemáticas A.C. en Monterrey (CIMAT Monterrey) y especialista en ciencia de datos.

acuerdo con criterios como los que se muestran en el Cuadro 3.³¹ Es importante insistir aquí que si bien se realizó un etiquetado manual de estos 76,895 documentos, dicho conjunto se dividió, como se explica más adelante, entre documentos para ajustar los modelos y los documentos reservados para la predicción.

Para determinar la etiqueta final de cada documento se tomó en consideración la clasificación realizada por ambos etiquetadores. En los casos en los que los etiquetadores discreparon en la asignación de una etiqueta, estas se sumaron a la categoría neutral. La distribución de las etiquetas fue la siguiente: 30,715 positivos; 31,156 negativos; y 15,024 neutrales.

Cuadro 3: Criterios para la Asignación de Etiquetas

Cualquier factor que sugiera crecimiento económico:	
Positivo:	Mayor inversión privada, mayor inversión extranjera directa, mayor inversión pública, mayor exportación, mayor demanda interna, más seguridad pública, mayor competencia económica, mayor estabilidad de precios, mayor estabilidad cambiaria, menores tasas de interés, entre otros.
Cuando el comentario no indique claramente positivo o negativo para el crecimiento de la economía regional:	
Neutral:	Alza o baja del tipo de cambio, estacionalidad en la producción, estrategias de ventas, alza o baja en precios de bienes. Asimismo, en esta categoría se incluyeron los comentarios de “ninguno”, “no comentó”, y otros similares, que denotaban una “no respuesta.”
Cualquier factor que apunte a una desaceleración, contracción o recesión económica:	
Negativo:	Menor inversión privada, falta de infraestructura, mayor incertidumbre cambiaria, mayor incertidumbre en política interna, mayor incertidumbre electoral, mayores costos de producción, repunte en inseguridad pública, desaceleración económica, menor disponibilidad de insumos, incremento en cierre de negocios, agravamiento de la pandemia, mayor volatilidad financiera, entre otros.

Fuente: Elaboración propia con información del PED del Banco de México.

³¹ Debemos reconocer que los etiquetados “manuales” son susceptibles a errores. Adicionalmente, emplear solo las clasificaciones de dos “etiquetadores” y asignar la categoría “neutral” a aquellos casos en los que hay discrepancia entre ambos investigadores puede dar lugar a una pérdida de información. De allí que pudiera resultar conveniente emplear un tercer etiquetador y tomar, por ejemplo, la mediana de las clasificaciones para asignar el sentimiento del conjunto de datos a etiquetar. En reconocimiento a estas debilidades, un tercer colaborador se encuentra trabajando en un etiquetado adicional, mismo que podría utilizarse en una futura versión de este trabajo.

3.4.2 Modelos de Predicción del Sentimiento

La elección de los modelos que pueden ser útiles en el proceso de generar índices de sentimiento se apoya en las características mismas de los documentos. Así, para la clasificación del sentimiento a partir de las características de la información en formato de texto de la que disponemos, se analizaron varios modelos de aprendizaje de máquina, entre los que se encuentran modelos clásicos como Máquina de Soporte Vectorial, Naive Bayes, Clasificador Ridge y Árbol de Decisión, además de modelos de Redes Neuronales Recurrentes (RNN) y modelos de Representación de Codificador Bidireccional de Transformadores (BERT).

3.5 Entrenamiento, Ejercicio de Predicción y Selección de Modelos

Siguiendo la literatura, se seleccionó aleatoriamente el 80% de la muestra de los documentos para *entrenar* los modelos, y el 20% restante se reservó como conjunto de prueba. Esto nos dejó con 61,516 documentos para entrenar los modelos y 15,379 para la evaluación del desempeño de los modelos.

Para encontrar el modelo “base”, es decir, el que servirá como punto de referencia para la comparación de los modelos de redes propuestos en este trabajo, se probaron inicialmente los métodos clásicos de aprendizaje automatizado SVM, Naive Bayes, Clasificador Ridge y Árbol de Decisión. Cada clasificador se probó con las siguientes representaciones del texto: Bolsa de Palabras (Bag of Words), TF-IDF (Frecuencia de Término–Frecuencia Inversa de Documento) y LSA-100 (Análisis Semántico Latente-100), LSA-300, LSA-500, LSA-1000.³² Como resultado de estas pruebas, se seleccionaron varios modelos, siendo el algoritmo SVM con la representación de texto TF-IDF (SVM+TF-IDF) el que mostró el mejor desempeño, indicado por el mayor valor del estadístico F1-Macro, por lo cual se utilizará como referencia para comparar otros modelos (ver Cuadro 4).³³ El F1-Macro,

³² En la literatura de aprendizaje de máquina, una *representación de texto* consiste en convertir palabras a representaciones numéricas para que los algoritmos comprendan y decodifiquen patrones dentro de un lenguaje. Para conocer cómo funcionan distintas representaciones de texto, ver Jurafsky y Martin (2020).

³³ El F1 es una medida diseñada para evaluar qué tan bien predice un modelo y se define como:

$$F1 = 2 * \left(\frac{\text{Precisión} * \text{Recuperación}}{\text{Precisión} + \text{Recuperación}} \right)$$

donde “Precisión” cuantifica la fracción de la clase correctamente clasificada en el total de los clasificados en dicha clase; en tanto que “Recuperación” mide la clasificación correcta de una clase en el total de los correctamente clasificados en

conviene señalar, se utiliza cuando se tienen datos desbalanceados, como ocurre en nuestra base de datos, ya que esta se integra con 30,715 documentos positivos, 31,156 negativos y 15,024 neutrales. El F1-Micro es otro indicador de ajuste; este, sin embargo, otorga el mismo peso a cada uno de los documentos clasificados, y que aquí se presenta solo para propósitos de comparación. Note, sin embargo, que ambos F1 arrojan el mismo resultado.

Cuadro 4. Evaluación de los Modelos Clásicos

Representación	Clasificador	F1-Macro (Test)	F1-Micro (Test)
TF-IDF	SVM	0.7643	0.7981
LSA-1000	SVM	0.7492	0.7791
BOW	SVM	0.7627	0.796
TF-IDF	Ridge	0.7375	0.7808

Fuente: Elaboración propia con información del PED del Banco de México.

Finalmente, se probaron los modelos de RNN y BERT.³⁴ El mejor modelo de RNN fue el que utilizó celdas GRU y vectores preentrenados FastText.³⁵ En el caso de los modelos basados en BERT, se evaluaron dos modelos preentrenados en español (beto) y otro ajustado a la tarea de análisis de sentimiento (beto-sentiment-analysis), siendo este último el que mejor predijo el sentimiento para el conjunto de datos de que se dispone (ver Cuadro 5).³⁶

Al final, los modelos con mejor ajuste empleando el criterio F1-Macro son: SVM TF+IDF (0.7643), RNN GRU 10L + FastText (0.7462) y BERT (beto-sentiment-analysis) (0.8397). En lo que resta del trabajo, nos referiremos a estos tres modelos simplemente como SVM, RNN y BERT, respectivamente.³⁷

dicha clase, así como aquellos que erróneamente se clasificaron en una categoría diferente. El estadístico fluctúa entre 0 y 1, y entre más cercano a 1, mejor el desempeño de un modelo. El Cuadro A6 del Apéndice presenta este y otros criterios de evaluación de modelos.

³⁴ Ver Cuadro A7 del Apéndice para una explicación en torno al entrenamiento de los modelos RNN y BERT.

³⁵ Ver Cho et al. (2014) para información sobre las celdas GRU. Ver Bojanowski et al. (2017) para una revisión de los vectores preentrenados FastText.

³⁶ Para consultar sobre modelos BERT, ver Devlin et al. (2018).

³⁷ Conviene señalar que el modelo SVM es un clasificador capaz de detectar los patrones en las representaciones vectoriales del texto utilizando una función de similitud que permite transformar el espacio original de los documentos a un espacio vectorial en el que es posible separar los documentos de acuerdo a sus características. Por su parte, los modelos de redes RNN y BERT son procesadores de secuencias que se adaptan muy bien al procesamiento de texto, toda vez que el texto en sí mismo es una secuencia de palabras. En este sentido, mientras que la representación de vectores que utiliza el modelo SVM no toma en cuenta el orden de las palabras, los modelos RNN y BERT sí lo considera. Otra diferencia entre el modelo de SVM y los modelos RNN y BERT es que el primero requiere relativamente menos datos que los modelos de redes. No obstante, con demasiados datos el modelo SVM se vuelve intratable por el tamaño de las matrices que necesita. Por su parte, los modelos basados en redes neuronales, si bien son también muy demandantes computacionalmente, tienen la ventaja de

Cuadro 5. Evaluación de los Modelos RNN y BERT

Conjunto de Entrenamiento		
Modelo	F1-Macro	F1-Micro
RNN GRU 10L + FastText	0.7462	0.7786
BERT (beto)	0.8254	0.8566
BERT (beto-sentiment-analysis)	0.8387	0.8651
Conjunto de Prueba		
Modelo	F1-Macro (Test)	F1-Micro (Test)
RNN GRU 10L + FastText	0.7662	0.7957
BERT (beto)	0.8289	0.8592
BERT (beto-sentiment-analysis)	0.8397	0.8664

Fuente: Elaboración propia con información del PED del Banco de México.

4. Índices de Sentimiento a Partir del RER

Este trabajo tiene como objetivos principales obtener índices de sentimiento utilizando la información en formato de texto contenida en los PED y determinar si estos se asocian con indicadores de suaves e indicadores duros de actividad económica.

Los índices de sentimiento en el nivel regional y nacional que aquí se presentan son los que se desprenden de los clasificadores SVM, RNN y BERT y se obtienen como sigue. Una vez que se tiene etiquetada la información, se toman como base los conteos de cada categoría en la que los documentos fueron catalogados por cada uno de los clasificadores utilizados (SVM, RNN y BERT) con valor 1 (para documentos clasificados con sentimiento positivo), 0 (documentos clasificados como neutrales) y -1 (documentos clasificados con sentimiento negativo).³⁸ Los índices de sentimiento regional (ISR) se construyen, a su vez, con las formulaciones que se definen en las expresiones 3 y 4:

ser altamente “paralelizables”, lo que facilita el manejo de los datos disponibles. Aquí, la “paralelización” se refiere a cómo se ejecuta el álgebra matricial, tanto en el entrenamiento como en la inferencia de redes neuronales.

³⁸ En este trabajo, la decisión de utilizar solo (-1, 0, 1) para identificar las etiquetas que utilizaron los algoritmos obedeció a que en las preguntas realizadas a los directivos consultados solo se les pide distinguir entre factores “negativos o positivos”. Es decir, no se les solicita que distingan entre factores “muy negativos, negativos, neutrales, positivos y muy positivos”, lo que daría margen a generar, por ejemplo, una clasificación más amplia, como (-2, -1, 0, 1, 2).

$$(3) \quad ISRPA_t^r = \sum_{i=1}^n \frac{Positivos_t^r - Negativos_t^r}{Positivos_t^r + Negativos_t^r} * \alpha_{i,j(t)}^r$$

$$(4) \quad ISRPB_t^r = \sum_{i=1}^n \frac{Positivos_t^r - Negativos_t^r}{Positivos_t^r + Negativos_t^r + Neutrales_t^r} * \alpha_{i,j(t)}^r$$

Al índice en (3) le llamamos “índices sin neutrales”, ya que deja fuera los documentos clasificados como neutrales; y al índice (4) le llamamos “índice con neutrales”, ya que cuenta a los documentos neutrales. En ambos índices, *Positivos* hace referencia a la cuenta de documentos clasificados como positivos, *Negativos* a la cuenta de negativos, y *Neutrales* a la cuenta de documentos neutrales, de la región *r* (= norte, centro norte, centro, sur), o del país. En estos índices *n* es el número de sectores, *i* hace referencia al sector productivo al que pertenece la empresa del directivo consultado de acuerdo con el PED, y *j(t)* es el trimestre más cercano al trimestre *t* en el que se tiene información del PIB regional. Por su parte, $\alpha_{i,j(t)}^r$ captura la participación del PIB del sector *i* en el trimestre más cercano a la realización de la entrevista, en el PIB total de la región *r*, o del país, en el trimestre más cercano a la realización de la entrevista *j(t)*.³⁹ Por ejemplo, si la entrevista se realizó en enero de 2017, la ponderación se realizaría utilizando la información del PIB del 4T-2016. Esta ponderación pretende ajustar el peso de las respuestas por la relevancia del sector al que pertenece el documento contabilizado en el índice respectivo. Esto es, la ponderación busca capturar de la mejor manera la estructura de la economía mexicana, dado que toma en cuenta el peso de los sectores productivos en un contexto regional y nacional.^{40,41} De acuerdo con estas

³⁹ Las estimaciones de los ponderadores sectoriales en nivel estatal toman como base el PIB estatal publicado por el INEGI. Dado que el PIB estatal publicado no tiene una periodicidad trimestral, nuestros ponderadores trimestrales de cada año se apoyan en los datos del año correspondiente. Otro punto a destacar es el ponderador usado para 2020. En particular, al momento de realizar este documento se contaba solo con el dato del PIB hasta 2019; de ahí que se haya adoptado el supuesto de que los ponderadores de 2020 son iguales a los de 2019, lo que supone que la estructura de la economía en 2020 es la misma a la del año previo. Lo anterior es discutible ya que el 2020 fue un año atípico por la pandemia del COVID-19.

⁴⁰ Para apreciar la relevancia de esta ponderación considere, como ejemplo, una región del país donde el sector productivo *X* tiene un valor absoluto de 10 en la región A y 100 en la región B. Si un entrevistado del sector A captura y expresa correctamente su sentimiento de que el sector *X* se contraerá 10%, en tanto que el entrevistado del sector B percibe correctamente que ese mismo sector *X* se expandirá 10%, ambas opiniones se cancelarían si no estuvieran ponderadas. Al ponderarse, se le da más peso a la opinión del entrevistado de la región B.

⁴¹ La Gráfica A1 y el Cuadro A8 del Apéndice muestran el comportamiento en el tiempo de los índices no ponderados y las correlaciones entre estos y los índices ponderados, respectivamente. Se puede apreciar, primero, que los distintos índices de sentimiento no ponderados siguen patrones muy similares en el tiempo; y, segundo, que las correlaciones de Pearson entre estos y los índices de sentimiento ponderados nacionales son superiores a 0.88; en tanto que en las regiones norte, centro norte y centro, las correlaciones son superiores a 0.81. En el sur, las correlaciones superan 0.70 en todos los casos, con

definiciones, mayores valores de los índices sugieren mayores niveles de sentimiento positivo por parte de los entrevistados.

Con base en lo anterior, tenemos dos índices generados con la clasificación de documentos derivados del SVM (sin neutrales y con neutrales), dos del RNN y dos del BERT, para cada región, así como en el nivel nacional. Ahora, para propósitos de identificar la calidad del ajuste de los distintos clasificadores (SVM, RNN y BERT), se obtuvieron también índices anotados excluyendo valores neutrales (Anotados) e incluyendo valores neutrales (Anotado c/n) por región y en el nivel nacional, también con las fórmulas (4) y (5), y en los que se utiliza el etiquetado manual realizado en este trabajo de los 76,895 documentos. Estos dos índices anotados son clave para determinar si los algoritmos elegidos clasifican los documentos de manera cercana a la realizada por los etiquetadores. Un buen algoritmo debería, en teoría, generar índices de sentimiento cercanos a los obtenidos con los respectivos índices anotados bajo el supuesto de que el etiquetado manual se realizó de manera adecuada.⁴²

4.1 Índices de Sentimiento Nacionales y Regionales Ponderados

Un elemento esencial en el aprendizaje de máquina es que los índices obtenidos a partir de los algoritmos ajustados por los investigadores se aproximen a los índices anotados. Dicho esto, pasamos a revisar en primera instancia el comportamiento de los índices obtenidos con los etiquetados de los algoritmos SVM, RNN y BERT, con los etiquetados anotados, en el nivel nacional. Un punto que conviene recordar nuevamente es que la información de los PED se obtiene solo en los meses de enero, abril, julio y octubre de cada año.

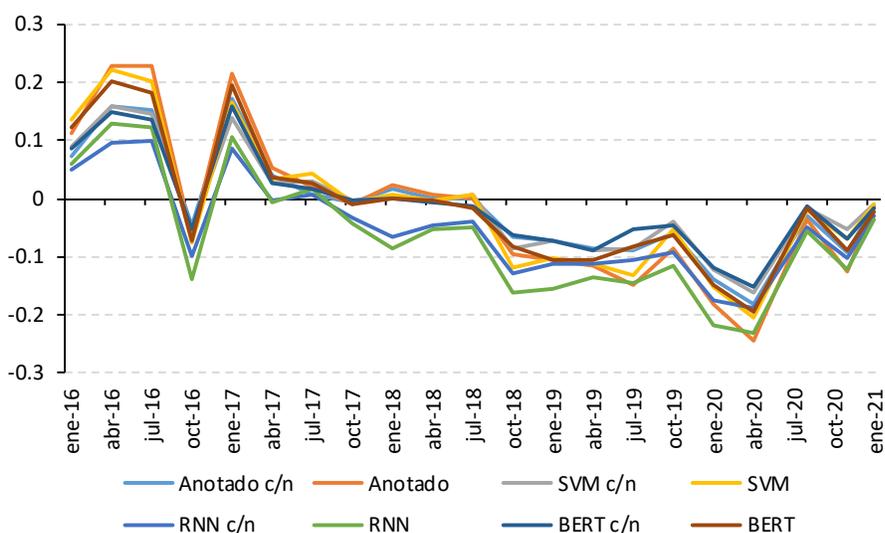
La Gráfica 2 muestra los seis índices de sentimiento obtenidos mediante los etiquetados derivados con los tres algoritmos de aprendizaje de máquina, y los dos índices anotados. Varios rasgos destacan de esta gráfica. En primer lugar, es evidente una asociación positiva entre los distintos índices con y sin datos neutrales. También resalta que los índices de

excepción de los índices BERT, donde son ligeramente superiores a 0.40. Esto sugiere que los índices ponderados y los no ponderados son similares bajo esta métrica.

⁴² Se debe tener presente que el etiquetado manual se realizará solo una ocasión. Una vez que surja nueva información, esta se alimentará y los algoritmos elegidos generarán la clasificación de los documentos (en nuestro caso, en 1, 0 y -1), y con esta información, pueden generarse los nuevos índices de sentimiento con base en las fórmulas (4) y (5).

sentimiento obtenidos mediante el etiquetado de los algoritmos RNN son, consistentemente, los que tienen los menores niveles. No obstante, la asociación entre los diferentes índices de sentimiento es confirmada con los coeficientes de correlación de Pearson que se muestran en el Cuadro 6. Allí se aprecia que los coeficientes de correlación bivariados superan, en todos los casos, el nivel de 0.9654. Estos cálculos sugieren robustez en cuanto a la clasificación generada con los distintos modelos de aprendizaje de máquina. Además, abonan a la posibilidad de que cuando estos índices se asocien con los diferentes indicadores de actividad, se obtengan resultados similares.

Gráfica 2. Índices de Sentimiento Nacionales Ponderados
Enero 2016 - Enero 2021



Nota: c/n indica que en el cálculo del índice respectivo se incluyeron las etiquetas “neutrales.”
Fuente: Elaboración propia con información de los PED del Banco de México.

Cuadro 6: Índices de Sentimiento Nacionales Ponderados
Enero 2016 - Enero 2021
Coeficientes de Correlación de Pearson

Indicador de Sentimiento*	Anotado c/n	Anotado	SVM c/n	SVM	RNN c/n	RNN	BERT c/n	Bert
Anotado c/n	1.0000							
Anotado	0.9972	1.0000						
SVM c/n	0.9866	0.9898	1.0000					
SVM	0.9779	0.9862	0.9973	1.0000				
RNN c/n	0.9722	0.9746	0.9872	0.9833	1.0000			
RNN	0.9654	0.9705	0.9850	0.9839	0.9980	1.0000		
BERT c/n	0.9909	0.9867	0.9893	0.9801	0.9800	0.9742	1.0000	
BERT	0.9898	0.9897	0.9922	0.9872	0.9819	0.9788	0.9982	1.0000

Nota: Todos los coeficientes de correlación son distintos de cero con un nivel de significancia del 1%.

Fuente: Elaboración propia con información del PED del Banco de México.

Por otro lado, los índices de sentimiento parecen reflejar, de manera general, una variedad de episodios relacionados que percibieron los entrevistados en el periodo analizado y que, de alguna manera, capturan diversos hechos asociados con el comportamiento de la economía mexicana. Por ejemplo, entre el 4T2015 y el 3T2016, la economía mexicana se caracterizó por un estancamiento del sector industrial y, en particular, de las manufacturas.⁴³ Ahora bien, en enero de 2016, los índices de sentimiento se ubican en un nivel bajo, suben en abril, pero retroceden ligeramente en julio de ese año. En octubre de 2016, los distintos índices de sentimiento caen de manera precipitada, coincidiendo con una marcada volatilidad en los mercados financieros vinculada al proceso electoral en Estados Unidos, así como con una elevada volatilidad y una depreciación significativa de la moneda nacional.⁴⁴

Con la llegada de Donald Trump a la presidencia de Estados Unidos surgieron preocupaciones relacionadas con la cancelación o renegociación del TLCAN, lo que se tradujo en volatilidad cambiaria y presiones adicionales a la baja de la inversión en México. En enero de 2017, la creciente expectativa de que la Administración Trump buscaría renegociar el TLCAN fue acompañada, a su vez, por un repunte en los índices de sentimiento. A partir de abril de 2017 y durante el 2018, la incertidumbre en torno al proceso electoral en México coincide, a su vez, con una baja en los índices de sentimiento.

El relativo estancamiento de estos índices a lo largo de 2019, por su parte, va acompañado de una moderada contracción de la economía mexicana durante ese año, en un contexto de desaceleración de la economía mundial derivada de tensiones comerciales y elevados riesgos geopolíticos.⁴⁵ Por otra parte, la reducción en los índices de sentimiento de enero de 2020 respecto de octubre 2019 se presenta en una etapa caracterizada por tensiones comerciales globales, riesgos geopolíticos y los riesgos de una propagación global del COVID-19.⁴⁶ En abril de 2020, los índices de sentimiento se redujeron con respecto de sus niveles de enero. La caída de ese mes coincide con el resguardo de las familias mexicanas para evitar contagiarse del virus, los cierres de actividades no esenciales por parte de las autoridades

⁴³ Ver Banco de México (2015), Banco de México (2016a, 2016b).

⁴⁴ Ver Banco de México (2016c).

⁴⁵ Sobre los factores detrás de la desaceleración, ver Banco de México (2019).

⁴⁶ Sobre estos eventos, ver Banco de México (2020a).

sanitarias mexicanas para contener el brote de COVID-19, el hecho de que medidas similares fueron implementadas en Estados Unidos -lo que detuvo las exportaciones hacia ese país-, además de que empezaron a notarse los primeros efectos negativos asociados con las disrupciones a las cadenas globales de valor.⁴⁷

En julio, mes en el que se alcanzó el pico de contagios y fallecidos de la llamada “primera ola” de COVID-19, los índices se recuperan, comportamiento que fue acompañado de una reclasificación de algunas actividades económicas “no esenciales” a “actividades esenciales.” Por su parte, las entrevistas realizadas durante octubre revelan una contracción en los índices de sentimiento respecto de los obtenidos en julio, comportamiento que coincide con preocupaciones crecientes de una segunda oleada de contagios.

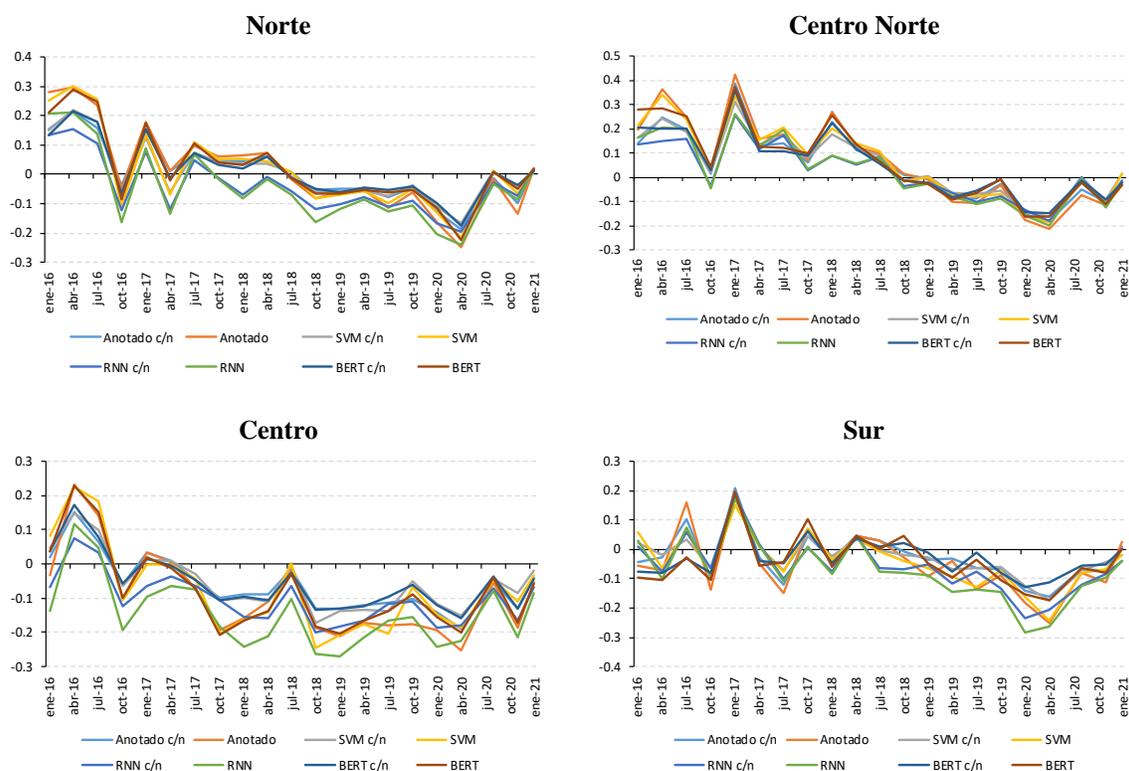
Finalmente, en enero de 2021, los índices de sentimiento se recuperan nuevamente respecto de los de octubre de 2020, si bien de manera moderada, ante la expectativa de un repunte en los niveles de consumo privado asociado a una mayor movilidad, y el buen comportamiento de las exportaciones manufactureras no automotrices derivado, principalmente, de la recuperación de la economía de Estados Unidos.⁴⁸ La reciente recuperación de los índices, no obstante, ha resultado insuficiente para recuperar los niveles que alcanzaron en enero de 2017.

La Gráfica 3 muestra, a su vez, los índices de sentimiento ponderados regionales (norte, centro norte, centro y sur), donde se aprecian patrones similares entre dichos índices *al interior de cada región*. Asimismo, puede apreciarse que es en la región centro norte donde los distintos índices de sentimiento son más parecidos entre sí (mismo patrón en el tiempo y menos dispersos), seguidos por los de la norte. En contraste, los índices de sentimiento de las regiones centro y sur lucen más dispersos y con patrones distintos para algunos periodos.

⁴⁷ Ver Banco de México (2020b).

⁴⁸ Ver Banco de México (2020c, 2021).

**Gráfica 3: Índices de Sentimiento Regionales Ponderados
Enero 2016 - Enero 2021**



Nota: c/n indica que en el cálculo del índice se incluyeron las etiquetas “neutrales.”

Fuente: Elaboración propia con información del PED del Banco de México.

Para revisar con mayor detalle los grados de asociación entre los índices de sentimiento *al interior de las regiones*, se construyó el Cuadro 7. Ahí se aprecian coeficientes de correlación positivos y superiores a 0.92 en todas las combinaciones de los índices de las regiones norte y centro norte; en tanto que en el centro todos son superiores a 0.91, con excepción de un coeficiente, que registra un nivel de 0.89. En el sur, en cambio, se observan correlaciones de hasta 0.74. Asimismo, en todas las regiones, con excepción de la sur, la mayor correlación de los índices anotados se obtiene con los índices BERT. Será conveniente tener presente estos grados de asociación lineal entre los indicadores en el nivel nacional, como al interior de las regiones, una vez que se relacionen con los distintos indicadores de actividad económica que se utilizarán en este trabajo.

Cuadro 7: Indicadores de Sentimiento Regionales Ponderados
Enero 2016 - Enero 2021
Coefficientes de Correlación de Pearson Intrarregionales

	Indicador de Sentimiento*	Anotado c/n	Anotado	SVM c/n	SVM	RNN c/n	RNN	BERT c/n	BERT
NORTE	Anotado c/n	1.0000							
	Anotado	0.9922	1.0000						
	SVM c/n	0.9701	0.9684	1.0000					
	SVM	0.9610	0.9720	0.9942	1.0000				
	RNN c/n	0.9341	0.9413	0.9710	0.9703	1.0000			
	RNN	0.9257	0.9413	0.9652	0.9722	0.9959	1.0000		
	BERT c/n	0.9795	0.9688	0.9860	0.9747	0.9598	0.9509	1.0000	
	BERT	0.9773	0.9772	0.9902	0.9882	0.9662	0.9633	0.9956	1.0000
	CENTRO NORTE	Anotado c/n	1.0000						
Anotado		0.9940	1.0000						
SVM c/n		0.9832	0.9846	1.0000					
SVM		0.9714	0.9839	0.9947	1.0000				
RNN c/n		0.9478	0.9453	0.9774	0.9682	1.0000			
RNN		0.9427	0.9496	0.9780	0.9780	0.9955	1.0000		
BERT c/n		0.9832	0.9767	0.9724	0.9608	0.9361	0.9327	1.0000	
BERT		0.9729	0.9786	0.9699	0.9697	0.9287	0.9353	0.9933	1.0000
CENTRO		Anotado c/n	1.0000						
	Anotado	0.9797	1.0000						
	SVM c/n	0.9536	0.9437	1.0000					
	SVM	0.9243	0.9369	0.9877	1.0000				
	RNN c/n	0.9336	0.9313	0.9437	0.9173	1.0000			
	RNN	0.8979	0.9281	0.9164	0.9107	0.9808	1.0000		
	Bert c/n	0.9802	0.9646	0.9676	0.9525	0.9460	0.9199	1.0000	
	Bert	0.9570	0.9712	0.9642	0.9722	0.9338	0.9320	0.9848	1.0000
	SUR	Anotado c/n	1.0000						
Anotado		0.9712	1.0000						
SVM c/n		0.9268	0.8895	1.0000					
SVM		0.8747	0.9028	0.9576	1.0000				
RNN c/n		0.8836	0.8476	0.9531	0.9208	1.0000			
RNN		0.8681	0.8669	0.9450	0.9565	0.9864	1.0000		
BERT c/n		0.8482	0.7668	0.8480	0.7403	0.7998	0.7516	1.0000	
BERT		0.8512	0.8069	0.8487	0.7818	0.7923	0.7663	0.9811	1.0000

Nota: c/n indica que en el cálculo del índice se incluyeron las etiquetas “neutrales.” Todos los coeficientes de correlación en este cuadro son distintos de cero.

Fuente: Elaboración propia con información del PED del Banco de México.

4.2 Índices de Sentimiento vs. Indicadores Suaves de Actividad Económica

Pasamos ahora a revisar las correlaciones entre los índices de sentimiento y un conjunto de indicadores suaves de actividad económica. Dada la definición de estos indicadores suaves, se espera que mayores niveles de estos se correlacionen positivamente con mayores niveles de los índices de sentimiento. El primer ejercicio consiste en correlacionar contemporáneamente los indicadores de sentimiento nacionales con dos indicadores oportunos (suaves) representativos en el nivel nacional publicados por el INEGI: el Indicador de Pedidos Manufactureros (IPM) y el Indicador de Confianza Empresarial del Sector Manufacturero (ICEMP).⁴⁹ De acuerdo con el INEGI, no existe un rezago en la publicación de estos indicadores oportunos; es decir, una vez que cierra el mes de referencia para el levantamiento de la información necesaria para elaborar dichos indicadores, su reporte del mes se publica el primer día hábil del siguiente mes.⁵⁰ En este sentido, los periodos de captura y procesamiento de información de estos indicadores mensuales son similares a los de nuestros índices de sentimiento. Dado que los indicadores de sentimiento se elaboran con información procedente de entrevistas realizadas en los meses de enero, abril, julio y octubre de cada año, como lo requiere el RER, las correlaciones contemporáneas implican que los meses del IPM e ICEMP correspondan con los cuatro meses señalados arriba. Dicho esto, pasamos a analizar correlaciones contemporáneas entre estos indicadores.

Dado que el IPM y el ICEMP son representativos en el nivel nacional, se correlacionaron con los índices de sentimiento nacionales ponderados. Los resultados se muestran en el Cuadro 8. Se puede apreciar que los coeficientes de correlación entre estos dos indicadores con cada uno de los diferentes índices de sentimiento nacionales resultaron positivos y estadísticamente distintos de cero a niveles de significancia del 1%. En el caso del IPM, los coeficientes de correlación fluctuaron entre un mínimo de 0.5781 y un máximo de 0.6373, en tanto que para el ICEMP fluctuaron entre 0.6512 y 0.7449. En otras palabras, las respuestas en formato de texto ofrecidas por los directivos entrevistados por el Banco de

⁴⁹ Como precaución, se realizaron pruebas de estacionariedad a las series de los índices de sentimiento (nacionales y regionales) y de los indicadores económicos. Estas aparecen en el Cuadro A9 del Apéndice. Se puede ver que todas las series son estacionarias. El cuadro muestra también estadísticos descriptivos de los distintos indicadores económicos.

⁵⁰ Información básica en torno a elementos metodológicos de los indicadores de actividad regionales y nacionales del INEGI puede consultarse en el Cuadro A10 del Apéndice.

México para la elaboración del RER proveen, en general, señales que van en la misma dirección que las obtenidas a partir de los dos indicadores suaves de actividad económica en el nivel nacional estimados por el INEGI con metodologías y muestras distintas.

Cuadro 8: Indicadores de Sentimiento Nacionales vs. Indicadores Nacionales de Opinión del Sector Manufacturero Enero 2016 - Enero 2021
Coeficientes de Correlación de Pearson

Indicador de Sentimiento	IPM		ICEMP	
	Coef. Corr.		Coef. Corr.	
Anotado c/n	0.6373	***	0.7449	***
Anotado	0.6291	***	0.7385	***
SVM c/n	0.6007	***	0.6980	***
SVM	0.5882	***	0.6863	***
RNN c/n	0.5958	***	0.6759	***
RNN	0.5781	***	0.6512	***
BERT c/n	0.5875	***	0.7089	***
BERT	0.5834	***	0.7059	***

Nota: *, ** y *** denotan niveles de significancia del 10%, 5% y 1%, respectivamente.

Fuente: Elaboración propia con información del PED del Banco de México e INEGI.

4.3. Índices de Sentimiento vs. Indicadores de Actividad Económica Nacionales

Esta sección analiza ahora en qué medida los índices de sentimiento nacionales se asocian con indicadores duros de actividad económica nacionales. Las variables son tasas de crecimiento entre el periodo t y el periodo $t-1$, y todas son estacionarias.⁵¹ Las tasas de crecimiento son trimestre a trimestre previo para todas las variables, con excepción del IGAE, en la que se toma la tasa de crecimiento mes a mes previo.

El Cuadro 9 presenta las correlaciones entre los indicadores de sentimiento nacionales y los cuatro indicadores de actividad económica nacional propuestos para el ejercicio: crecimiento trimestral del PIB (TCPIB), tasa de crecimiento mensual del IGAE (TCIGAE), tasa de crecimiento trimestral del indicador de actividad industrial (TCIMAI) y la tasa de crecimiento trimestral del indicador mensual de la actividad manufacturera (TCIMAIMAN).

⁵¹ Ver Cuadro A9 del Apéndice.

**Cuadro 9: Indicadores de Sentimiento Nacional vs. Indicadores de Actividad Económica Nacional
Enero 2016 - Enero 2021
Coeficientes de Correlación de Pearson**

Indicador de Sentimiento	TCPIB		TCIGAE		TCIMAI		TCIMAIMAN	
	Coef. Corr.		Coef. Corr.		Coef. Corr.		Coef. Corr.	
Anotado c/n	0.3655	*	0.3317	*	0.3618	*	0.3643	*
Anotado	0.3567	*	0.3209		0.3518	*	0.3533	*
SVM c/n	0.3696	**	0.3315	*	0.3550	*	0.3601	*
SVM	0.3562	*	0.3173		0.3409	*	0.3448	*
RNN c/n	0.3599	*	0.3232		0.3449	*	0.3531	*
RNN	0.3564	*	0.3167		0.3370	*	0.3456	*
BERT c/n	0.3587	*	0.3227		0.3450	*	0.3548	*
BERT	0.3528	*	0.3152		0.3379	*	0.3467	*

Nota: * y ** denotan niveles de significancia del 10% y 5%, respectivamente.

Fuente: Elaboración propia con información del PED del Banco de México e INEGI.

Los resultados muestran, conforme a lo anticipado, correlaciones positivas en todos los casos. Estas correlaciones, no obstante, son más bajas que las obtenidas con los indicadores suaves, y son distintas de cero, en la mayoría de los casos, a un nivel de significancia de solo 15% para las variables TCPIB, TCIMAI y TCIMAIMAN. En particular, los coeficientes de correlación se ubican entre 0.35 y 0.37 en el caso de TCPIB; entre 0.33 y 0.37 en el caso de TCIMAI, y entre 0.34 y 0.37 en el caso de TCIMAIMAN. En los casos del TCPIB y TCIMAIMAN, las correlaciones con los índices de sentimiento nacionales muestran niveles de significancia más cercanos al 10% que al 15%. En el caso de TCIGAE, los coeficientes de correlación se ubicaron entre 0.31 y 0.34, y si bien solo dos coeficientes alcanzaron el nivel de significancia de 15%, la mayoría de ellos está cerca de este. Aquí, la relevancia de encontrar asociaciones estadísticamente significativas se eleva ya que cuatro de los indicadores duros propuestos tienen un rezago en su publicación de aproximadamente 8 semanas, y dos de ellos tienen un rezago de 4 meses.

Esto indica, nuevamente, que la información en formato de texto provista por los entrevistados en las entrevistas concedidas para la elaboración del RER, ofrece señales en la dirección correcta, ya sea al alza o a la baja, de la actividad económica en su conjunto, de la actividad industrial en su totalidad, y de la industria manufacturera en particular. Esto puede ser de utilidad, dado el rezago de ocho semanas que caracteriza la publicación de estos cuatro indicadores.

También se exploró el grado de asociación entre los indicadores de sentimiento regional con dos indicadores de actividad económica regional; en este caso, las tasas de crecimiento respecto del trimestre previo de los ITAER y los ITAER del sector manufacturero (TCITAER y TCITAERMAN, respectivamente). El Cuadro 10 presenta las correlaciones entre los valores de los TCITAER y los valores de los índices de sentimiento obtenidos a partir de los distintos algoritmos. En esta ocasión, se aprecia una asociación entre estos de manera general solo en el norte, en tanto en el resto de las regiones las correlaciones escasamente alcanzan el 15% de significancia.

Cuadro 10: Indicadores de Sentimiento Regionales vs. TCITAER
Enero 2016 - Enero 2021
Coefficientes de Correlación de Pearson

Indicador de Sentimiento	TCITAER Norte		TCITAER Centro Norte		TCITAER Centro		TCITAER Sur	
	r		r		r		r	
Anotado c/n	0.3445	*	0.2772		0.3689	*	0.2517	
Anotado	0.3296	*	0.2718		0.3202		0.3219	
SVM c/n	0.3486	*	0.3264	*	0.2814		0.3316	*
SVM	0.3273	*	0.3141		0.2337		0.3975	**
RNN c/n	0.3295	*	0.3642	*	0.2907		0.2457	
RNN	0.3169		0.3552	*	0.2463		0.2948	
BERT c/n	0.3758	**	0.2884		0.2975		0.1906	
BERT	0.3628	*	0.2748		0.2572		0.2537	

Nota: * y ** denotan niveles de significancia del 10% y 5%, respectivamente.

Fuente: Elaboración propia con información del PED del Banco de México e INEGI.

El Cuadro 11 muestra, a su vez, las correlaciones entre TCITAERMAN con respecto de los índices de sentimiento regionales, donde resalta que estos son especialmente importantes para capturar la evolución del desempeño del sector manufacturero de la región norte. Es decir, los comentarios expresados por los informantes de esa región en torno al crecimiento de la actividad manufacturera se reflejan adecuadamente en el comportamiento de los índices de sentimiento en el norte, al obtenerse coeficientes de correlación estadísticamente distintos de cero a niveles de significancia de 5% y 10%. Esta región del país, conviene resaltar, es la que tiene también la mayor participación de dicha actividad en la estructura económica nacional. En cambio, en el resto de las regiones, los resultados sugieren que los índices de sentimiento regionales no muestran una asociación estadísticamente significativa con la actividad de sus respectivos sectores manufactureros.

**Cuadro 11: Indicadores de Sentimiento Ponderados Regionales vs. TCITAERMAN
Enero 2016 - Enero 2021**

Coefficientes de Correlación de Pearson^a

Indicador de Sentimiento	TCITAERMAN Norte		TCITAERMAN Centro Norte		TCITAERMAN Centro		TCITAERMAN Sur	
	r		r		r		r	
Anotado c/n	0.4174	*	0.2243		0.3350		0.2225	
Anotado	0.4111	*	0.2173		0.2898		0.3053	
SVM c/n	0.4040	*	0.2545		0.2383		0.3250	
SVM	0.3925	*	0.2420		0.1901		0.4142	*
RNN c/n	0.3920	*	0.2845		0.2579		0.2417	
RNN	0.3876	*	0.2726		0.2166		0.3121	
BERT c/n	0.4410	**	0.2278		0.2719		0.1121	
BERT	0.4351	**	0.2138		0.2284		0.1815	

Nota: * y ** denotan niveles de significancia del 10% y 5%, respectivamente.

Fuente: Elaboración propia con información del PED del Banco de México e INEGI.

El Cuadro 12 muestra, por su parte, que los índices de sentimiento nacionales tienen un mayor grado de asociación con los ciclos económicos de las regiones centrales. Esto se aprecia en el hecho de que los coeficientes de correlación entre los índices nacionales y los indicadores de actividad económica regional, son mayores y estadísticamente distintos de cero a niveles de significancia de 10% en todos los casos. El resultado podría responder a que estas regiones guardan relativamente una mayor similitud en su estructura productiva con la economía nacional. La actividad económica de la región norte también reporta una asociación positiva con los índices de sentimiento nacionales, si bien menos estrecha en comparación con las regiones centrales, y solo si se consideran niveles de significancia del 15%. Cabe recordar la vocación económica de la región norte, orientada mayormente hacia las actividades manufactureras, especialmente de exportación, característica que la hace relativamente distinta al resto de las regiones del país. En relación con la región sur, los indicadores de sentimiento nacionales no logran capturar el comportamiento de la actividad económica de esta región, lo cual podría atribuirse a que la estructura de su economía es, por mucho, relativamente más dependiente de la minería petrolera y de la actividad turística.

Cuadro 12. Índices de Sentimiento Nacionales Ponderados vs. TCITAER Regionales
Enero 2016 - Enero 2021
Coefficientes de Correlación de Pearson

Indicador de Sentimiento	TCITAER Norte		TCITAER Centro Norte		TCITAER Centro		TCITAER Sur	
	r		r		r		r	
Anotado c/n	0.3340	*	0.3780	**	0.3943	**	0.2711	
Anotado	0.3265	*	0.3713	**	0.3849	**	0.2608	
SVM c/n	0.3332	*	0.3902	**	0.3965	**	0.2824	
SVM	0.3203		0.3795	**	0.3828	**	0.2687	
RNN c/n	0.3283	*	0.3816	**	0.3881	**	0.2678	
RNN	0.3259	*	0.3801	**	0.3833	**	0.2637	
BERT c/n	0.3287	*	0.3768	**	0.3871	**	0.2661	
BERT	0.3237		0.3729	**	0.3810	**	0.2598	

Nota: * y ** denotan niveles de significancia del 10% y 5%, respectivamente.

Fuente: Elaboración propia con información del PED del Banco de México e INEGI.

5. Comentarios finales

Este trabajo presentó estimaciones de diversos índices de sentimiento regionales y nacionales obtenidos a partir de los Programas Trimestrales de Entrevistas a Directivos (PED) realizados en el periodo de enero 2016 a enero de 2021. Utilizando un total de 76,895 documentos (respuestas en formato de texto) obtenidos de un total de 9,802 entrevistas, los índices de sentimiento se correlacionaron con un conjunto de indicadores suaves (de opinión) y duros de actividad económica nacionales y regionales.

Los resultados obtenidos muestran que los índices de sentimiento se correlacionan contemporáneamente de manera positiva con diversos indicadores de actividad económica. En particular, las correlaciones son estadísticamente distintas de cero entre los índices de sentimiento nacionales y los Indicadores de Pedidos Manufactureros (IPM) y de Confianza Empresarial del Sector Manufacturero (ICEMP); en tanto que con las tasas de crecimiento trimestrales del PIB nacional (TCPIB), del Indicador Mensual de Actividad Industrial Total (TCIMAI) y de la Actividad Industrial Manufacturera (TCIMAIMAN), las correlaciones son significativas al 15%. Por otro lado, en el nivel regional, los índices de sentimiento registraron una correlación estadísticamente distinta de cero con las tasas de crecimiento del Indicador Trimestral de la Actividad Manufacturera Regional (TCITAERMAN) de la región norte. Finalmente, los índices de sentimiento nacionales registraron un mayor grado de asociación con las tasas de crecimiento trimestrales de los Índices de la Actividad Económica

Regional (TCITAER) de las regiones centrales y, en menor medida, con las de la región norte.

Hacia delante, los algoritmos entrenados en este trabajo para la detección de tópicos y clasificación del sentimiento nacional y regional podrán continuar usándose para procesar, de manera automática, la información en formato de texto que se capture en los PED a fin de estimar los índices de sentimiento y obtener información más oportuna de la actividad económica nacional y regional. Otra de las aportaciones de este trabajo es haber generado etiquetas para el entrenamiento de futuros modelos de predicción basados en aprendizaje de máquina, frente a la alternativa de generar las etiquetas haciendo uso, por ejemplo, de lexicones preentrenados, que son más rígidos a los cambios de la estructura temática y a la evolución del idioma.

Este trabajo, no obstante, aún ofrece áreas de oportunidad. Por ejemplo, dado que en la asignación manual de las etiquetas se clasificaron como neutrales comentarios que carecían de sentido por sí mismos, podría considerarse en futuros proyectos incorporar a la clasificación otros elementos de contexto, como el sector al que pertenecen las empresas y/o la entidad federativa en la que se ubican. Adicionalmente, el análisis de correlación que se implementó aquí entre los distintos índices de sentimiento y los indicadores económicos no consideró correlaciones entre los primeros y los componentes cíclicos o de tendencia de los segundos. Asimismo, no se exploraron medidas alternativas de correlación, las cuales pudieran utilizarse como medidas de robustez de los resultados obtenidos aquí. También, la utilidad de los índices de sentimiento que aquí se obtuvieron pudiera evaluarse de manera más formal al incluir a estos, por ejemplo, como variables explicativas en análisis econométricos donde la variable dependiente sea algún indicador duro de actividad económica. Otra área de oportunidad es explorar las asociaciones de estos índices con indicadores económicos que no se consideraron aquí. En específico, es posible que los índices de sentimiento de la región sur se asocien con indicadores de actividad turística o del ramo energético, y no tanto con indicadores de actividad manufacturera de esa región.

Una limitante que debemos reconocer, no obstante, es que los modelos de aprendizaje de máquina, que son la base para la construcción de nuestros índices de sentimiento, no se

recomiendan para situaciones en las que se tienen “pocos” datos, ya que los algoritmos pueden aprender patrones o hacer inferencias basados, principal o exclusivamente, en las características del grupo mayoritario, lo que pudiera llevar a ignorar información relevante.⁵²

El trabajo que aquí se presenta debe interpretarse, por tanto, solo como una primera aproximación en cuanto a la capacidad de los índices de sentimiento basados en la información en formato de texto contenida en el PED del Banco de México para complementar, de manera más oportuna, la información provista por datos duros y suaves de actividad económica en México.

⁵² Al respecto, ver Taboada et al. (2011); Barocas (2014); Boyd y Crawford (2012); Sweeney (2013); Miller y Dwork (2015); Moritz (2014); Kay et al. (2015); Wallach (2014).

Referencias

Algaba, A., Gardia, D., Bluteau, K., Borms, S. y K. Boudt. 2020. “Econometrics Meets Sentiment: An Overview of Methodology and Applications.” *Journal of Economic Surveys*. Vol. 34, No. 3, pp. 512-547.

Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A. y Aljaaf, A.J. 2020. “A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science.” En: Berry, M., Mohamed, A., Yap, B. (eds): *Supervised and Unsupervised Learning for Data Science*. Springer.

Anandarajan, M., Hill C., y T. Nolan. 2019. “Semantic Space Representation and Latent Semantic Analysis.” En: *Practical Text Analytics. Advances in Analytics and Data Science*, Vol. 2. Springer, Cham. En: https://doi.org/10.1007/978-3-319-95663-3_6.

Aragón, M., Carmona, M., Montes, M., Escalante, H., Villaseñor, L. y D. Moctezuma. 2019. “Overview of MEX-A3T at IberLEF 2019: Authorship and Aggressiveness Analysis in Mexican Spanish Tweets.”

Arora, S. 2021. “Data Mining vs. Machine Learning: What is the Difference.” En: www.simplilearn.com/data-mining-vs-machine-learning-article.

Athey, S. 2018. “The Impact of Machine Learning on Economics.” University of Stanford Manuscript. January.

Azqueta-Gavaldón, A., Hirschbühl, D., Onorante, L. y L. Saiz. 2020. “Economic Policy Uncertainty in the Euro Area: an Unsupervised Machine Learning Approach.” *European Central Bank Working Paper Series No. 2359* / Enero.

Banco de México. 2015. Informe Trimestral Octubre-Diciembre.

Banco de México. 2016a. Informe Trimestral Enero-Marzo.

Banco de México. 2016b. Informe Trimestral Abril-Junio.

Banco de México. 2016c. Informe Trimestral Julio-Septiembre.

Banco de México. 2019. Informe Trimestral Octubre-Diciembre.

Banco de México. 2020a. Informe Trimestral Enero-Marzo.

Banco de México. 2020b. Reporte sobre las Economías Regionales Julio-Septiembre.

Banco de México. 2020c. Informe Trimestral Octubre-Diciembre.

Banco de México. 2021. Informe Trimestral Enero-Marzo.

Barocas, S. 2014. "Data Mining and the Discourse on Discrimination." En: <https://dataethics.github.io/proceedings/DataMiningandtheDiscourseOnDiscrimination.pdf>

Benhabib, J. y M. Spiegel. 2017. "Sentiment and Economic Activity: Evidence from U.S. States." National Bureau of Economic Research Working Paper No. 23899.

Blei, D. 2011. "Introduction to Probabilistic Topic Models." *Communications of the ACM*, Vol. 55, No. 4, pp. 77-84.

Bojanowski, P., Grave, E., Joulin, A. y Mikolov, T. 2017. "Enriching Word Vectors with Subword Information". *Transactions of the Association for Computational Linguistics*. En: <https://aclanthology.org/Q17-1010.pdf>.

Boyd, D. y K. Crawford. 2012. "Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon." *Information, Communication & Society*. Vol. 15, No. 5, pp. 662-679.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. e Y. Bengio. 2014. "Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation." En: [arXiv preprint arXiv:1406.1078](https://arxiv.org/abs/1406.1078).

D'Andrea, A., Ferri, F., Grifoni, P. y T. Guzzo. 2015. "Approaches, Tools and Applications for Sentiment Analysis Implementation." *International Journal of Computer Applications*. Vol.125, No.3, pp. 26-33. September.

De Bondt, G. y S. Schiaffi. 2015. "Confidence Matters for Current Economic Growth: Empirical Evidence for the Euro-area and the United States." *Social Science Quarterly*. Vol. 96, No. 4, pp. 1027-1040.

Devlin, J., Chang, M., Lee, K., y K. Toutanova. 2018. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." En: [arXiv preprint arXiv:1810.04805](https://arxiv.org/abs/1810.04805).

Díaz, M. y J. Huerta. 2020. "Co-movimiento entre los Índices de Confianza del Consumidor de México y Estados Unidos 2001-2018." *Economía, Sociedad y Territorio*, Vol. 20, No. 62, pp. 123-150.

Doerr, S., Gambacorta, L. y J. Serena. 2021. "Big Data and Machine Learning in Central Banking." BIS Working Paper No. 930. Monetary and Economic Department. Marzo.

Garrett, T., Hernández-Murillo, R. y M. Owyang. 2004. "Does Consumer Sentiment Predict Regional Consumption?" *Federal Reserve Bank of St. Louis Review*, Marzo/Abril, Vol. 87, No. 2, pp. 123-35.

Gentzkow, M., Bryan, K. y M. Taddy. 2019. "Text as Data." *Journal of Economic Literature*. Vol. 57, No. 3, pp. 535-574.

Hatzivassiloglou, V. y K. McKeown. 2002. "Predicting the Semantic Orientation of Adjectives." In 35th Annual Meeting of the Association for Computational Linguistics, pp. 174-181, Madrid, Spain. Association for Computational Linguistics.

Kay, M, C. Matuszek y S. Munson. 2015. "Unequal Representation and Gender Stereotypes in Image Search Results for Occupations." En: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. ACM, pp. 3819-3828.

Kim, S. y E. Hovy. 2004. "Determining the Sentiment of Opinions." En: Proceedings of International Conference on Computational Linguistics."

Korab, P. 2021. "Use of Machine Learning in Economic Research: What the Literature Tells Us." En: <https://towardsdatascience.com/use-of-machine-learning-in-economic-research>

Jurafsky, D. y J. Martin. 2020. "Speech and Language Processing." Third Edition Draft. Disponible en Internet: <https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf>.

Medhat, W., Hassan, A. y H. Korashy. 2014. "Sentiment Analysis Algorithms and Applications: A Survey." *Ain Shams Engineering Journal*. Vol 5, No. 4, pp.1093-1113.

Miller, C. y C. Dwork. 2015. "Algorithms and Bias: Q. and A." New York Times. En: <http://www.nytimes.com/2015/08/11/upshot/algorithms-and-bias-q-and-a-with-cynthia-dwork.html>

Moritz, H. 2014. "How Big Data is Unfair." En: <https://medium.com/@mrtz/how-big-data-is-unfair-9aa544d739de>.

Muñiz, V. 2020. "Una Consulta sobre Técnicas de Aprendizaje de Máquina." Entrevista Presencial. Centro de Investigación en Matemáticas A.C. Monterrey. Noviembre.

Nasukawa, T. y J. Yi. 2003. "Sentiment Analysis: Capturing Favorability using Natural Language Processing." En: Proceedings of the 2nd International Conference on Knowledge Capture, Octubre.

Pang, B., Lee, L. y S. Vaithyanathan. 2002. "Thumbs up? Sentiment Classification using Machine Learning Techniques." En: Proc. Of the 7th EMNLP, pp.79-86.

Pinto, S. 2019. "Sentiment Analysis of the Fifth District Manufacturing and Service Surveys." *Economic Quarterly*. Vol. 105, No. 9, pp. 133-164.

Prabowo, R. y M. Thelwall. 2009. "Sentiment Analysis: A Combined Approach." *Journal of Informetrics*, Vol.3, No. 2, pp. 143-157.

- Rho, C., R. Fernández y B. Palma. 2021. “A Sentiment Based Indicator for the Mexican Financial Sector.” Banco de México, Documento de Investigación No. 2021-04. Mayo.
- Rosenbaum, P. y D. Rubin. 1983. “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” *Biometrika*, Vol. 70, No. 1, pp. 41-55.
- Salhin, A., M. Sherif y E. Jones. 2016. “Managerial Sentiment, Consumer Confidence and Sector Returns.” *International Review of Financial Analysis*. Vol. 47, pp. 24-38.
- Santero, T. y N. Westerlund. 1996. “Confidence Indicators and Their Relationship with Changes in Economic Activity.” OECD Economic Department Working Papers No. 170.
- Stsiopkina, M. 2022. “Hard vs. Soft Data: The Difference.” En: <https://oxylabs.io/blog/ha>.
- Suss, J. y H. Treitel. 2019. “Predicting Bank Distress in the UK with Machine Learning.” Bank of England Staff Working Paper No. 831.
- Sutton, R. y A. Barto. 2018. “Reinforcement Learning” 2nd Ed. The MIT Press.
- Sweeney, L. 2013. “Discrimination in On-line Ad Delivery.” *Queue*. Vol.11, No.3, pp. 1-20.
- Taboada, M., Brooke, J., Tofiloski, T., Voll, K. y M. Stede (2011), “Lexicon-Based Methods for Sentiment Analysis.” *Computational Linguistics*, Vol. 37, No. 2.
- Turney, P. 2002. “Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews.” En: Proceedings of the 40th Meeting of the Association for Computational Linguistics, Philadelphia, PA, pp. 417-424.
- Uno, Y. y K. Adachi. 2019. “Calculating Non-Response Bias in Firms’ Inflation Expectations Using Machine Learning Techniques.” No.19-E-17. Diciembre.
- Wallach, H. 2014. “Big Data, Machine Learning, and the Social Sciences: Fairness, Accountability, and Transparency.” En: <https://medium.com/@hannawallach/big-data-machine-learning-and-the-social-sciences-927a8e20460d>.
- Weng, L. 2021. “Learning with not Enough Data Part 1: Semi-Supervised Learning.” En: <https://lilianweng.github.io/posts/2021-12-05-semi-supervised/>
- Wiesalla, L. 2021. “The Machine Learning Workflow. Concepts and Applications.” En: <https://www.nextlytics.com/blog/machine-learning-workflow>.
- Witten, I., F. Eibe, M. Hall y C. Pal. 2017. “Data Mining: Practical Machine Learning Tools and Techniques.” 4th Ed. Morgan Kaufmann Publishers. Elsevier.

Apéndice

Cuadro A1. Pasos para Implementar Modelos de Análisis de Sentimientos

1	Seleccionar fuente de información a analizar, o corpus.	Se obtiene normalmente de blogs, foros, redes sociales, entrevistas. Los datos a analizar suelen estar desorganizados, utilizar vocabularios distintos, slangs, etc.
		Se refiere a limpiar los datos antes de realizar al análisis. Esto conlleva, principalmente: a. Identificar y extraer palabras individuales (“tokenización”). b. Remover cifras o palabras que no proveen significado al texto (eliminar “stop words”, como “la”, “el”, “en”, “con”, etc.). c. Reducir palabras a sus raíces (“word stemming”). d. Agrupar palabras de acuerdo con su significado (lematizar). e. Mantener el anonimato de las fuentes.
2	Preprocesar (Preparar) la Información	Una vez que concluye la “limpieza” de la base datos, el texto retenido es examinado, conservándose aquellas oraciones con expresiones subjetivas (opiniones, creencias, puntos de vista) y eliminándose aquellas con información factual. En ocasiones, la preparación del texto se lleva a cabo manualmente. Sin embargo, cuando la información es abundante, en esta etapa se utilizan técnicas de análisis de texto y procesamiento de lenguaje. En la mayoría de los algoritmos de aprendizaje de máquina, sin embargo, sí es necesario llevar a cabo un trabajo manual para al etiquetado de los datos a fin de entrenar los modelos. El etiquetado, conviene señalar, consiste en asignar valores numéricos a frases textuales o palabras de los documentos.
3	Elegir modelo a utilizar	La elección se apoya en la extracción de elementos o características de los documentos.
4	Entrenar el modelo	El entrenamiento se realiza con una sub-muestra de los datos, normalmente un 80% de los datos disponibles. Los restantes se reservan para llevar a cabo la “predicción.”
5	Evaluar el modelo (algoritmo)	Para evaluar el desempeño de un modelo se utilizan datos internos no utilizados en la construcción del algoritmo, aplicándose diversos criterios para determinar qué tan bien el modelo ajusta esos datos. En relación con estos criterios, ver Cuadro A6 de este Apéndice.
6	Predicción	Comparar los datos que genera el modelo, con los datos reservados para la prueba.

Fuente: Con información de Wiesalla (2021).

**Cuadro A2. Documentos por Sección, Región y Tipo de Entrevistado
Identificados en Función del PED en el que fueron Obtenidos**

Sección	Región	Universo 1T2011-4T2020			Muestra 4T2015-4T2020		
		Directivos	Stakeholders	Total	Directivos	Stakeholders	Total
Actividad Económica	Norte	10,100	683	10,783	6,172	601	6,773
	Centro Norte	7,197	531	7,728	4,965	476	5,441
	Centro	8,314	664	8,978	6,487	609	7,096
	Sur	7,391	437	7,828	4,655	403	5,058
	Nacional	33,002	2,316	35,317	22,279	2,089	24,368
Perspectivas	Norte	10,530	728	11,258	6,488	639	7,127
	Centro Norte	7,101	536	7,637	4,909	481	5,390
	Centro	8,731	684	9,415	6,731	623	7,354
	Sur	7,620	436	8,056	4,754	398	5,152
	Nacional	33,982	2,384	36,366	22,882	2,141	25,023
Riesgos	Norte	13,119	825	13,944	7,424	719	8,143
	Centro Norte	9,294	637	9,931	5,855	565	6,420
	Centro	9,165	727	9,892	6,223	644	6,867
	Sur	10,196	551	10,747	5,951	484	6,435
	Nacional	41,774	2,742	44,514	25,453	2,412	27,865
Totales	Norte	33,749	2,236	35,985	20,084	1,959	22,043
	Centro Norte	23,592	1,704	25,296	15,729	1,522	17,251
	Centro	26,210	2,075	28,285	19,441	1,876	21,317
	Sur	25,207	1,424	26,631	15,360	1,285	16,645
	Nacional	108,758	7,442	116,197	70,614	6,642	77,256

Fuente: Elaboración propia con información del PED del Banco de México.

...Continúa

Entrevistas por Región y Tipo de Entrevistados*
Directivos

Trimestre	Nacional	Norte	Centro Norte	Centro	Sur	Trimestre	Nacional	Norte	Centro Norte	Centro	Sur
1T2011	369	150	79	58	82	2T2016	441	127	102	111	101
2T2011	373	149	79	59	86	3T2016	448	121	103	110	114
3T2011	471	161	82	122	106	4T2016	444	117	102	118	107
4T2011	476	150	88	123	115	1T2017	432	115	105	112	100
1T2012	471	155	91	122	103	2T2017	428	113	100	112	103
2T2012	481	163	90	97	131	3T2017	441	118	105	113	105
3T2012	435	142	84	75	134	4T2017	440	121	103	111	105
4T2012	484	153	90	90	151	1T2018	426	109	102	111	104
1T2013	497	150	100	79	168	2T2018	441	118	102	114	107
2T2013	502	152	98	82	170	3T2018	444	114	106	112	112
3T2013	483	162	101	83	137	4T2018	432	114	103	113	102
4T2013	455	158	102	72	123	1T2019	430	107	104	111	108
1T2014	473	164	100	79	130	2T2019	437	112	104	113	108
2T2014	466	151	100	82	133	3T2019	430	107	102	117	104
3T2014	492	162	100	91	139	4T2019	424	108	100	111	105
4T2014	468	156	100	81	131	1T2020	427	113	102	110	102
1T2015	453	134	102	87	130	2T2020	422	105	103	112	102
2T2015	463	143	101	93	126	3T2020	417	105	101	107	104
3T2015	464	147	100	92	125	4T2020	423	105	101	115	102
4T2015	480	151	100	100	129	1T2021	422	109	103	108	102
1T2016	444	124	104	109	107	Total	18,349	5,435	4,044	4,117	4,753

Stakeholders

Trimestre	Nacional	Norte	Centro Norte	Centro	Sur	Trimestre	Nacional	Norte	Centro Norte	Centro	Sur
1T2015	44	12	10	11	11	2T2018	42	11	10	12	9
2T2015	47	17	10	15	5	3T2018	40	10	10	9	11
3T2015	35	10	10	10	5	4T2018	45	13	11	10	11
4T2015	42	11	10	10	11	1T2019	41	11	10	10	10
1T2016	39	10	10	11	8	2T2019	33	7	8	10	8
2T2016	40	10	9	12	9	3T2019	39	10	10	9	10
3T2016	34	7	9	9	9	4T2019	40	10	10	12	8
4T2016	46	10	10	20	6	1T2020	37	10	10	11	6
1T2017	41	13	11	10	7	2T2020	38	10	10	11	7
2T2017	43	13	9	12	9	3T2020	42	11	11	10	10
3T2017	43	10	10	14	9	4T2020	43	11	11	11	10
4T2017	42	12	11	11	8	1T2021	38	10	10	12	6
1T2018	41	13	10	11	7	Total	537	148	129	156	104

Nota: En total, durante los 41 trimestres del periodo 2011-2021 se han realizado 19,364 entrevistas en el nivel nacional.

Fuente: Elaboración propia con información del PED del Banco de México.

Cuadro A3. Preguntas de la Sección de Actividad Económica

Trimestre	Preguntas
1T2011 al 4T2011	Durante los últimos 3 meses, ¿Cómo cambió la demanda de sus productos/servicios? ¿A qué lo atribuye?
1T2012 al 3T2013	Durante el X TRIM: ¿Cómo cambió la demanda de sus productos/servicios? ¿A qué lo atribuye?
4T2013 al 3T2016	Durante el X TRIM, y respecto del TRIM previo, cómo cambió: La demanda por sus productos o servicios (o sus ventas). ¿A qué lo atribuye? Factores externos Factores internos
4T2016 al 1T2021	Especifique los factores externos que impulsaron el nivel de actividad en su empresa durante el TRIM: Especifique los factores internos que impulsaron el nivel de actividad en su empresa durante el TRIM: Especifique los factores externos que limitaron el nivel de actividad en su empresa durante el TRIM: Especifique los factores internos que limitaron el nivel de actividad en su empresa durante el TRIM: Observaciones de actividad

...Continúa

Cuadro A3. Preguntas de la Sección de Riesgos

Trimestre	Preguntas
1T2011	<p>¿Qué aspectos considera usted que son un factor de riesgo en la actividad económica en su sector en el corto y mediano plazo?</p> <p>Observaciones adicionales de riesgo</p>
2T2011	<p>¿Qué elementos considera, que de materializarse, podrían acelerar la actividad económica en la región?</p> <p>¿Qué elementos considera, que de materializarse, podrían desacelerar la actividad económica en la región?</p> <p>Observaciones adicionales riesgos</p>
3T2011 al 2T2012	<p>¿Cuáles son los principales riesgos que enfrenta el desempeño económico de su estado o entidad federativa en los próximos meses?</p>
3T2012 al 1T2013	<p>¿Cuáles son los 2 principales riesgos que enfrenta el desempeño económico de su estado o entidad federativa en los próximos 6 meses?</p>
3T2013	<p>¿Cuáles son los 2 principales riesgos que enfrenta el desempeño económico de su estado o entidad federativa en los próximos 6 meses? (<i>Considere tanto los riesgos al alza como los riesgos a la baja</i>)</p>
4T2013	<p>Riesgos que pudieran impulsar el desempeño económico de la entidad en los próximos 6 meses</p> <p>Riesgos que pudieran limitar el desempeño económico de la entidad en los próximos 6 meses</p>
1T2014 al 3T2015	<p>Riesgos al alza (aquéllos que de materializarse impulsarían el desempeño económico) en los próximos 6 meses</p> <p>Riesgos a la baja (aquéllos que de materializarse limitarían el desempeño económico) en los próximos 6 meses</p>
4T2015 al 3T2016	<p>Riesgos al alza (aquéllos que de materializarse impulsarían el desempeño económico) en los próximos 12 meses</p> <p>Riesgos a la baja (aquéllos que de materializarse limitarían el desempeño económico) en los próximos 12 meses</p>
4T2016 al 1T2021	<p>¿Cuáles son los 2 principales eventos/situaciones/hechos que de llegar a ocurrir impulsarían el desempeño económico de su entidad federativa en los próximos 12 meses?</p> <p>¿Cuáles son los 2 principales eventos/situaciones/hechos que de llegar a ocurrir limitarían el desempeño económico de su entidad federativa en los próximos 12 meses?</p>

...Continúa

Cuadro A3. Preguntas de la Sección de Perspectivas

Trimestre	Preguntas
1T2011 al 2T2011	Respecto del momento actual, cómo espera usted que cambie la demanda por su producto en los próximos 3 y 6 meses , ¿A qué lo atribuye? Observaciones adicionales
3T2011 al 4T2011	Respecto del momento actual, cómo espera usted que cambie la demanda de su producto en los próximos 3 y 6 meses , ¿A qué factores externos lo atribuye? ¿A qué factores internos lo atribuye?
1T2012 al 4T2012	Respecto de la situación en el X TRIM, cómo anticipa usted que cambie: La demanda por sus productos o servicios en los próximos 6 y 12 meses , ¿A qué factores externos lo atribuye? ¿A qué factores internos lo atribuye? Observaciones adicionales
1T2013 al 3T2015	Respecto de la situación en el X TRIM, cómo anticipa usted que cambie: La demanda por sus productos o servicios, ¿A qué factor externo atribuye el cambio esperado en los próximos 6 meses ? ¿A qué factor interno atribuye el cambio esperado en los próximos 6 meses ?
4T2015 al 3T2016	Respecto de la situación en el X TRIM, cómo anticipa usted que cambie la demanda por sus productos o servicios ¿A qué factor externo atribuye el cambio esperado en los próximos 12 meses ? ¿A qué factor interno atribuye el cambio esperado en los próximos 12 meses ?
4T2016 al 1T2021	Respecto de la situación en el X TRIM, cómo anticipa usted que cambie la DEMANDA por sus productos o servicios para los próximos 12 meses : Especifique los factores externos que impulsarán el nivel de actividad en su empresa para los próximos 12 meses : Especifique los factores internos que impulsarán el nivel de actividad en su empresa para los próximos 12 meses : Especifique los factores externos que limitarán el nivel de actividad en su empresa para los próximos 12 meses : Especifique los factores internos que limitarán el nivel de actividad en su empresa para los próximos 12 meses : Observaciones adicionales

...Continúa

Cuadro A3. Preguntas Adicionales

Trimestre	Preguntas
1T2011	En general, ¿Qué factores espera que impulsen o limiten el crecimiento de la actividad en su sector en el corto y mediano plazo? Observaciones adicionales de factores
2T2011	¿Qué factores están impulsando el crecimiento de la actividad económica en su región? ¿Qué factores están limitando el crecimiento de la actividad económica en su región? Observaciones adicionales factores
3T2011	¿Qué factores internos considera usted que están impulsando el crecimiento económico en su estado o entidad federativa? ¿Qué factores externos considera usted que están impulsando el crecimiento económico en su estado o entidad federativa?
2T2013	Nota importante: Si el directivo espera que las tasas de crecimiento sean mayores ya sea en el 3er Trimestre o en el 4to Trimestre de 2013, preguntar: ¿Hay algún riesgo que pudiera afectar este repunte? ¿Cómo espera que sea la tasa de crecimiento anual en su entidad en el 3er Trimestre de 2013 comparada con la del 2do Trimestre? ¿A qué factores atribuye esas expectativas? Y la tasa de crecimiento anual en su entidad en el 4to Trimestre, ¿cómo se compara con la del 3er Trimestre? ¿Por las mismas razones que para el tercer trimestre o por cuáles razones?
3T2013	¿Cuáles considera son las fortalezas con las que cuenta su entidad para crecer en el contexto económico esperado en el 2014? ¿Cuáles considera son las debilidades con las que cuenta su entidad para crecer en el contexto económico esperado en el 2014?

Nota: Las respuestas que conforman la base de datos son las que son abiertas. En ese sentido, antes del 4T2016 las respuestas se refieren a factores a los que atribuyen la tendencia observada en el trimestre de referencia. A partir del 4T2016, se clasifican entre factores de impulso y limitantes. Las preguntas a *stakeholders* son las mismas pero se pregunta por el desempeño del sector en general, en lugar de la actividad de una empresa. Las preguntas a *stakeholders* se realizaron de forma diferenciada a partir de 2015.

Fuente: Elaboración propia con información del PED del Banco de México e INEGI.

Cuadro A4. Correspondencia Sector RER con Sector y Subsector SCIAN

Sector (RER-Banxico)	Sector (SCIAN) 2 dígitos	Sub sector (SCIAN) 3 dígitos
1) Agropecuario-Industria Alimentaria	11 Agricultura, ganadería, aprovechamiento forestal, pesca y caza	111 Agricultura 112 Cría y explotación de animales 112 Ganadería
	31-33 Industrias manufactureras	311 Industria alimentaria
2) Comercio	43 Comercio al por mayor	431-436
	46 Comercio al por menor	461-468
3) Construcción e inmuebles	23 Construcción	236 Edificación 237 Construcción de obras de ingeniería civil u obra pesada 238 Trabajos especializados para la construcción
	53 Servicios inmobiliarios y de alquiler de bienes muebles e intangibles	531 Servicios inmobiliarios 532 Servicios de alquiler de bienes muebles
		312-315 321-327 331-336 339
5) Minería, electricidad, gas y agua	21 Minería	211 Extracción de petróleo y gas 212 Minería de minerales metálicos y no metálicos excepto petróleo y gas 213 Servicios relacionados con la minería
	22 Electricidad, agua y suministro de gas por ductos al consumidor final	221 Generación, transmisión y suministro de energía eléctrica 222 Agua y suministro de gas productos cons. final
6) Otros servicios	54 Servicios profs., científicos y técnicos	541 Servicios profesionales, científicos y técnicos
	56 Servicios de apoyo a los negocios y manejo de desechos y servicios de remediación	561 Servicios de apoyo a los negocios
	61 Servicios educativos	611 Servicios educativos
	62 Servicios de salud y asistencia social	621 Servicios médicos de consulta externa y servicios relacionados 622 Hospitales
	72 Servicios de alojamiento temporal y de preparación de alimentos y bebidas	722 Servicios de preparación de alimentos y bebidas
7) Servicios Financieros y Seguros	52 Servicios financieros y de seguros	522 Insts .de intermed. crediticia y financiera no bursátil 523 Acts. bursátiles cambiarias y de inversión financiera 524 Compañías de fianzas, seguros y pensiones
	48-49 Transportes, correos y almacenamiento	483 Transporte por agua 484 Autotransporte de carga 488 Servicios relacionados con el transporte 492 Servicios de mensajería y paquetería 493 Servicios de almacenamiento
	51 Información en medios masivos	511 Edición de publicaciones y de software, excepto mediante de Internet 515 Radio y televisión, excepto mediante de Internet 517 Otras telecomunicaciones 518 Proveedores de acceso a Internet, servs. de búsqueda en la red y servicios de procesamiento de información 519 Otros servicios de información
9) Turismo	48-49 Transp, correos y almacenamiento	487 Transporte Turístico
	72 Servicios de alojamiento temporal y de preparación de alimentos y bebidas	721 Servicios de alojamiento temporal

Fuente: Elaboración propia con información de INEGI.

Cuadro A5. Tópicos por Trimestre y por Región

Trimestre	Tópico	Norte	Centro Norte	Centro	Sur
1T2016	1	Inseguridad	Tipo cambio	Inseguridad	Inseguridad
	2	EUA	Inseguridad	Tipo cambio	Precio petróleo
	3	Crecimiento industria	Gasto público	EUA	Actividad petrolera
	4	Inversión			Turismo
	5	Mayor demanda			
	6	Depreciación peso			
2T2016	1	Inseguridad	Tipo cambio	Tipo cambio	Inseguridad
	2	EUA	Inseguridad	Mercado interno	Precio petróleo
	3	Tipo cambio	EUA	Inseguridad	Inestabilidad social
	4	Depreciación peso			
3T2016	1	Depreciación peso	Tipo cambio	Tipo cambio	Inseguridad
	2	Inseguridad	Inseguridad	Inseguridad	Precio petróleo
	3	EUA		Inversión	Gasto público
	4	Efecto calendario		Precio petróleo	
4T2016	1	Incertidumbre EUA	Tipo cambio	Tipo cambio	Tipo cambio
	2	Aumento precio combustible	EUA	Incremento precios	Inestabilidad social
	3	Inseguridad	Incremento precios	Inseguridad	Inseguridad
	4		Inseguridad	Incertidumbre	Precio petróleo
1T2017	1	EUA	EUA	Tipo cambio	Inseguridad
	2	Tipo cambio	Inseguridad	Inseguridad	Precio petróleo
	3	Renegociación TLCAN			EUA
	4	Inseguridad			Actividad económica
2T2017	1	Inseguridad	Inseguridad	Tipo cambio	Inseguridad
	2	EUA	Tasa interés	Inseguridad	Precio petróleo
	3	Tipo cambio		Renegociación TLCAN	Gobierno estatal
	4	Inversión			EUA
	5	Incertidumbre			
3T2017	1	EUA	Renegociación TLCAN	Renegociación TLCAN	Actividad económica
	2	Inseguridad	Inseguridad	Elección 2018	Inseguridad
	3	Renegociación TLCAN		Tipo cambio	Precio petróleo
	4	Incremento precios		Demanda servicios	Renegociación TLCAN
	5	Obra pública			
4T2017	1	Incertidumbre TLCAN	Renegociación TLCAN	Renegociación TLCAN	Proceso electoral
	2	Inseguridad	Inseguridad	Inseguridad	Inseguridad
	3	Tipo cambio	Tipo cambio	Elección 2018	Tipo cambio
	4	Renegociación TLCAN			Alza costos
	5	EUA			
	6	Elecciones			

...Continúa

Trimestre	Tópico	Norte	Centro Norte	Centro	Sur
1T2018	1	Inseguridad	Inseguridad	Tipo cambio	Proceso electoral
	2	EUA	Incertidumbre elecciones	Renegociación TLCAN	Inseguridad
	3	Renegociación TLCAN	Renegociación TLCAN	Elecciones	Zonas económicas especiales
	4	Incremento precios		Inseguridad	Tasa interés
	5			Tasa interés	
2T2018	1	EUA	Inseguridad	Tipo cambio	Renegociación TLCAN
	2	Renegociación TLCAN	Tipo cambio	Inseguridad	Inseguridad
	3	Inseguridad	Renegociación TLCAN	Renegociación TLCAN	Tasa interés
	4	Tipo cambio			Incertidumbre
	5	Incertidumbre política			
3T2018	1	Inseguridad	Inseguridad	Tipo cambio	Nuevo gobierno
	2	Nuevo gobierno	Dinamismo sector	Inseguridad	Inseguridad
	3	EUA	Cambio gobierno	Nuevo gobierno	Cambio gobierno
	4	USMCA		Proyectos construcción	Tasa interés
	5	Obra pública		TLCAN	Mayor actividad económica
	6	Encarecimiento combustible			
4T2018	1	Inseguridad	Incertidumbre gobierno	Inseguridad	Inseguridad
	2	EUA	Inseguridad	Tipo cambio	Cambio gobierno
	3	Tipo cambio	Tipo cambio	Desabasto gasolina	Tasa interés
	4	Inversión		EUA	Gasto público
	5	Obra pública		Inversión	
1T2019	1	EUA	Inseguridad	Inseguridad	Inseguridad
	2	Inversión	Gobierno federal	Tipo cambio	EUA
	3	Incertidumbre gobierno	EUA	EUA	Gasto público
	4	Cierre frontera			Tasa interés
	5	Inseguridad			
2T2019	1	EUA	Inseguridad	Inseguridad	Inversión pública
	2	Incertidumbre gobierno	Gasto público	Tipo cambio	Inseguridad
	3	Tipo cambio		Incertidumbre EUA	EUA
	4	Inversión			Precio petróleo
	5	Inseguridad			Gasto público
	6				Puerto Veracruz
3T2019	1	Inseguridad	inseguridad	Tipo cambio	Inseguridad
	2	Obra pública	Gobierno federal	Inseguridad	Precio petróleo
	3	Ratificación TMEC	EUA	EUA	Aumento gasto público
	4		Vivienda		

...Continúa

Tópicos por Trimestre y por Región

Trimestre	Tópico	Norte	Centro Norte	Centro	Sur
4T2019	1	Inseguridad	Inseguridad	Inseguridad	Inversión pública
	2	Obra pública	Gobierno federal	Tipo cambio	Inseguridad
	3	Incertidumbre		EUA	Precio petróleo
	4	Obra pública			Tasa interés
	5	Aprobación TMEC			EUA
1T2020	1	COVID 19	COVID 19	Tipo cambio	Precio petróleo
	2	Apoyo gobierno	Apoyo gobierno	Apoyo gobierno	COVID 19
	3	EUA	Industria automotriz		Tipo cambio
	4	Tipo cambio	Pandemia		Emergencia sanitaria
2T2020	1	Recuperación económica	COVID 19	Tipo cambio	Pandemia
	2	COVID 19	Reactivación industria	Inseguridad	Precio petróleo
	3	Actividad económica	Inversión gobierno	Apoyo empresas	Confinamiento
	4	Apoyo empresas	Inseguridad	COVID 19	Tipo cambio
	5	Inseguridad	Exportaciones EUA		
3T2020	1	EUA	Gobierno federal	Inseguridad	Pandemia
	2	COVID 19	COVID 19	Tipo cambio	Reactivación económica
	3	Reactivación económica		Obra pública	Inversión
	4	Inseguridad		Rebote COVID 19	
	5			Desempleo	
	6			Cierre empresas	
4T2020	1	EUA	EUA	Tipo cambio	Pandemia
	2	Vacunación	Deterioro seguridad	Inseguridad	Precio petróleo
	3	Inseguridad	Vacunación	Vacunación	COVID 19
	4	Reactivación económica	Pandemia	Control pandemia	Turismo
	5				Obra pública
	6				EUA
1T2021	1	COVID 19	Vacunación	Tipo cambio	Vacunación
	2	Vacunación	Inversión gobierno federal	Inseguridad	Precio petróleo
	3	incremento precios	EUA	COVID 19	Inversión
	4		inseguridad	Vacunación	Reactivación económica
	5				Pandemia

Fuente: Elaboración propia con información del PED del Banco de México.

Cuadro A6. Criterios para Evaluación de Algoritmos

Medida	Definición		Observaciones
Exactitud (Accuracy)	$(VP + VN)/TOTAL$	<p>VP = número de predicciones positivas correctamente clasificadas</p> <p>VN = número de predicciones negativas correctamente clasificadas</p> <p>TOTAL = número total de casos u observaciones.</p>	Primera métrica que se revisa al evaluar un clasificador. No obstante, si los datos están desbalanceados o si se está más interesado en detectar una de las clases, la Exactitud no captura realmente la eficacia de un clasificador.
Precisión (Precision)	$VP / (VP + FP)$	<p>FP = número de predicciones positivas calificadas incorrectamente (falsos positivos).</p>	Cuantifica la fracción de positivos verdaderos entre el total de los clasificados como positivos. Se utiliza para medir qué tan efectivo es el modelo en detectar la categoría de interés, es decir, la categoría positiva.
Recuperación (Recall)	$VP / (VP + FN)$	<p>FN = número de predicciones negativas calificadas incorrectamente</p>	Mientras más cercano a 1, indica que más datos de la categoría verdadera fueron bien clasificados.
F1 (F1 Score)	$\frac{2 * (Precisión * Recuperación)}{Precisión + Recuperación}$		Media armónica de Precisión y Recuperación. Se ubica entre 0 y 1. Entre más cercano a 1, mejor el desempeño del modelo.

Nota: Existe una relación entre “Precisión” y “Recuperación.” Un modelo que predice todo como positivo tendrá un recuperación de 1, pero una precisión muy baja ya que tendría muchos falsos positivos; mientras que un modelo que solo predijera un positivo y el resto negativos tendría una recuperación muy baja, pero una precisión muy alta. Es por ello que se recurre a la medida F1, ya que esta mitiga el impacto de las tasas altas y acentúa el de las tasas bajas.

Fuente: Elaboración propia.

Cuadro A7. Procedimiento para Entrenar las Redes: RNN y BERT

Para entrenar las RNN se probaron distintas configuraciones de hiperparámetros para determinar el mejor tamaño del estado oculto y la capa de embeddings (entrada). Se observó que variar estos hiperparámetros modifica bastante el rendimiento del modelo obtenido. El mejor resultado se obtuvo con un tamaño de estado oculto de 40 y embedding de 10. Todo esto se probó con una red recurrente simple. Una vez obtenidos estos parámetros, se probaron distintas celdas recurrentes LSTM y GRU con un número distinto de capas: 1, 3, 5 y 10. Los mejores resultados para este tipo de redes se obtuvieron con una celda GRU y 10 capas. Adicionalmente, se probó utilizar vectores preentrenados fasttext en español, como capa de entrada. Esta modificación mejoró el F1 score. Cabe mencionar que cuatro épocas de entrenamiento fueron suficientes para obtener el mejor F1 en el conjunto de pruebas. Entrenar por más épocas sobreajusta y no mejora en el conjunto de pruebas. La tasa de aprendizaje fue de $1e-3$.

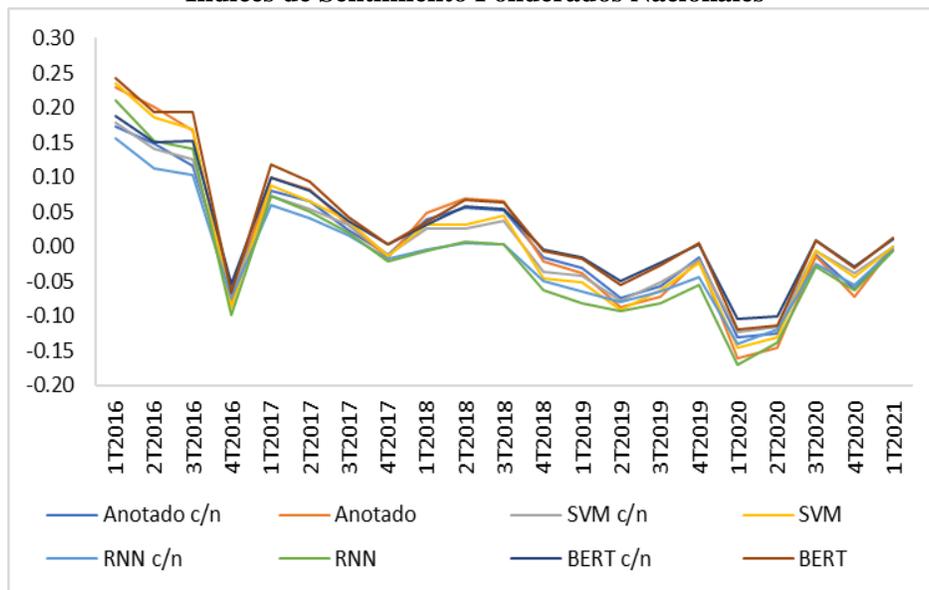
Por su parte, el entrenamiento de las redes BERT fue más sencillo. Se utilizó el modelo preentrenado en un corpus en español BETO, y adicionalmente se utilizó una versión de BETO ya ajustada en la tarea de análisis de sentimiento. Se agregó una capa de clasificación después de la última capa del codificador. La intención fue ajustar finamente (fine tuning) el modelo preentrenado y la capa de clasificación a la tarea de análisis de sentimiento y, en particular, al dominio de la entrevista regional. Por esta razón se utiliza una tasa de aprendizaje muy baja, $2e-5$, y evitar destruir los pesos ya aprendidos durante el preentrenamiento de BETO. En este modelo fue necesario ajustar por 2 épocas para obtener los mejores resultados. El Cuadro muestra los resultados obtenidos.

Evaluación de Diferentes Especificaciones RNN y BERT

Modelo	Accuracy	F1-Macro	F1-Micro
LSTM 1L	0.7373	0.7295	0.7668
LSTM 3L	0.5666	0.3789	0.4536
LSTM 5L	0.4901	0.7138	0.7451
GRU 1L	0.7781	0.7422	0.7746
GRU 3L	0.7803	0.7436	0.7741
GRU 5L	0.7813	0.7403	0.774
GRU 10L	0.7622	0.7221	0.7559
GRU 3L + FastText	0.7702	0.7288	0.7694
GRU 10L + FastText	0.7847	0.7462	0.7786
BERT (beto-base)	0.8553	0.8254	0.8566
BERT (beto-sentiment-analysis)	0.866	0.8387	0.8651

Fuente: Elaboración propia con información del PED del Banco de México e INEGI.

Cuadro A8: Índices de Sentimiento No Ponderados y su Correlación con Índices de Sentimiento Ponderados Nacionales



Nota: c/n indica que en el cálculo del índice respectivo se incluyeron las etiquetas “neutrales.”
Fuente: Estimaciones propias.

**Índices de Sentimiento Nacionales y Regionales
Enero 2016 - Enero 2021**

Coefficientes de Correlación de Pearson entre Índices Ponderados y No Ponderados

Índices	Nacional	Norte	Centro Norte	Centro	Sur
Anotado c/n	0.8874	0.8583	0.9083	0.9174	0.7061
Anotado	0.8938	0.8848	0.9201	0.8683	0.7338
SVM c/n	0.8944	0.8603	0.9359	0.9202	0.7201
SVM	0.8990	0.8915	0.9383	0.9002	0.7697
RNN c/n	0.9064	0.8710	0.8920	0.8648	0.7141
RNN	0.9022	0.8917	0.9064	0.8104	0.7565
BERT c/n	0.8922	0.8356	0.9267	0.8937	0.4063
BERT	0.8949	0.8612	0.9410	0.8725	0.4011

Nota: c/n indica que en el cálculo del índice respectivo se incluyeron las etiquetas “neutrales.” Todos los coeficientes resultaron estadísticamente distintos de zeros a un nivel de significancia de 0.01%.
Fuente: Estimaciones propias con información de INEGI y Banco de México.

Cuadro A9. Estadísticas Descriptivas y Pruebas de Estacionariedad de las Series

Estadísticas Descriptivas						
	Media	Mediana	Máximo	Mínimo	Desv. Est.	Obs.
ICEMP	52.768	52.892	56.599	43.972	2.926	21
IPM	51.094	51.637	52.516	43.882	1.900	21
TCPIB	-0.001	0.004	0.111	-0.203	0.053	21
TCIGAE	-0.004	0.002	0.057	-0.171	0.041	21
TCIMAI	-0.009	0.002	0.063	-0.260	0.060	21
TCIMAIMAN	-0.009	0.001	0.105	-0.307	0.073	21
TCITAER_NT	0.001	0.005	0.144	-0.226	0.061	21
TCITAER_CN	0.002	0.003	0.123	-0.199	0.054	21
TCITAER_CE	0.000	0.006	0.105	-0.208	0.053	21
TCITAER_SR	-0.005	0.001	0.079	-0.193	0.048	21
TCITAERMAN_NT	-0.011	0.005	0.085	-0.339	0.079	21
TCITAERMAN_CN	-0.004	-0.002	0.132	-0.263	0.067	21
TCITAERMAN_CE	-0.008	0.005	0.154	-0.349	0.087	21
TCITAERMAN_SR	0.001	0.001	0.041	-0.117	0.033	21

Fuente: Elaboración propia con información de INEGI.

Pruebas de Raíces Unitarias para las Series Económicas

Ho: La serie tiene una raíz unitaria

Series	t-stat	Prob.	Especificación	Rechazo Ho:
ICEMP	-4.3109	0.0144	Constante y tendencia	Sí
IPM	-5.9193	0.001	Constante y tendencia	Sí
TCPIB	-5.715	0.000	Constante	Sí
TCIGAE	-5.869	0.000	Constante	Sí
TCIMAI	-5.630	0.000	Constante	Sí
TCIMAIMAN	-5.933	0.000	Constante	Sí
TCITAER_NT	-6.257	0.000	Constante	Sí
TCITAER_CN	-5.729	0.000	Constante	Sí
TCITAER_CE	-5.781	0.000	Constante	Sí
TCITAER_SR	-5.730	0.000	Constante	Sí
TCITAERMAN_NT	-5.433	0.000	Constante	Sí
TCITAERMAN_CN	-6.518	0.000	Constante	Sí
TCITAERMAN_CE	-6.672	0.000	Constante	Sí
TCITAERMAN_SR	-3.179	0.038	Constante	Sí

Nota: Corresponden a pruebas Dickey-Fuller Aumentadas. La columna "Prob" presenta los "valores p" de una cola de McKinnon.

Fuente: Elaboración propia con información de INEGI.

...Continúa

Pruebas de Raíces Unitarias para los Índices de Sentimiento

Ho: La serie tiene una raíz unitaria

	Series	t-Stat	Prob.	Especificación:	Rechazo Ho:
Nacional	Anotado_c/n	-4.5722	0.0086	Constante y tendencia	Sí
	Anotado	-4.402	0.0121	Constante y tendencia	Sí
	SVM_c/n	-3.988	0.027	Constante	Sí
	SVM	-3.778	0.040	Constante	Sí
	RNN_c/n	-3.802	0.038	Constante	Sí
	RNN	-3.803	0.038	Constante	Sí
	BERT_c/n	-4.476	0.011	Constante	Sí
	BERT	-4.314	0.014	Constante	Sí
Norte	Anotado_c/n	-4.232	0.0168	Constante y tendencia	Sí
	Anotado	-3.876	0.0333	Constante y tendencia	Sí
	SVM_c/n	-4.218	0.017	Constante y tendencia	No
	SVM	-3.962	0.028	Constante y tendencia	No
	RNN_c/n	-3.821	0.037	Constante y tendencia	Sí
	RNN	-3.696	0.047	Constante y tendencia	Sí
	BERT_c/n	-4.418	0.012	Constante y tendencia	Sí
	BERT	-4.128	0.021	Constante y tendencia	Sí
Centro Norte	Anotado_c/n	-5.0667	0.0033	Constante y tendencia	Sí
	Anotado	-5.1432	0.0028	Constante y tendencia	Sí
	SVM_c/n	-2.001	0.564		No
	SVM	-2.015	0.557		No
	RNN_c/n	-4.904	0.005	Constante y tendencia	Sí
	RNN	-4.927	0.004	Constante y tendencia	Sí
	BERT_c/n	-4.854	0.005	Constante y tendencia	Sí
	BERT	-4.714	0.007	Constante y tendencia	Sí
Centro	Anotado_c/n	-2.3622	0.164		No
	Anotado	-2.4567	0.1402		No
	SVM_c/n	-2.475	0.136		No
	SVM	-2.522	0.125		No
	RNN_c/n	-2.748	0.084	Constante	Sí
	RNN	-2.923	0.060	Constante	Sí
	BERT_c/n	-2.439	0.145		No
	BERT	-2.475	0.136		No
Sur	Anotado_c/n	-4.2882	0.0036	Constante	Sí
	Anotado	-4.3391	0.0032	Constante	Sí
	SVM_c/n	-3.414	0.023	Constante	Sí
	SVM	-3.484	0.020	Constante	Sí
	RNN_c/n	-3.036	0.049	Constante	Sí
	RNN	-3.094	0.043	Constante	Sí
	BERT_c/n	-4.166	0.005	Constante	Sí
	BERT	-3.999	0.007	Constante	Sí

Nota: Corresponden a pruebas Dickey-Fuller Aumentadas. La columna "Prob" presenta los "valores p" de una cola de McKinnon.

Fuente: Elaboración propia con información del PED del Banco de México.

Cuadro A10

Indicadores de Opinión Nacionales

IPM	Indicador de Pedidos Manufactureros	<p>Índice compuesto que integra expectativas de directivos empresariales respecto del: Volumen esperado de pedidos, Producción esperada, Niveles esperados de personal ocupado, Oportunidad en la entrega de insumos por parte de los proveedores e Inventarios de insumos. Varía entre 0 y 100 puntos. A medida que el optimismo se generaliza entre los informantes, el valor del indicador crece, y viceversa. Este indicador es útil para adelantar tendencias de la actividad económica.</p> <p>Unidad de medida: Puntos. Cifras ajustadas por estacionalidad.</p> <p>Periodicidad: Mensual.</p> <p>Rezago en publicación: No.</p>
ICEMP	Indicador de Confianza Empresarial del Sector Manufacturero	<p>Índice compuesto que agrega cinco variables respecto de la percepción que tienen los directivos empresariales del sector manufacturero sobre la situación económica presente y futura en el país y en su empresa. Varía entre 0 y 100 puntos A medida que el optimismo se generaliza, el indicador crece, y viceversa. Este indicador es útil para adelantar tendencias de la actividad económica.</p> <p>Unidad de medida: Puntos. Cifras ajustadas por estacionalidad.</p> <p>Periodicidad: Mensual.</p> <p>Rezago en publicación: No.</p>

Indicadores Nacionales

TCPIB	Crecimiento del PIB Real Trim. vs. Trim. previo	<p>Rezago en publicación: La estimación “oportuna” del PIB se pone a disposición del público el último día hábil del mes siguiente al trimestre de referencia, y son sustituidos cuando se publican los resultados “completos” del PIB Trimestral, es decir, a los 52 días de concluido el trimestre.</p> <p>Periodicidad: Trimestral.</p>
TCIGAE	Crecimiento del IGAE Mes vs. Mes previo	<p>Periodicidad: Mensual.</p> <p>Rezago en publicación: 1 mes.</p>
TCIMAI	Crecimiento Trimestral del Indicador Mensual de Actividad Industrial Total	<p>Los datos mensuales del Índice Mensual de la Actividad Industrial (IMAI) están disponibles desde enero de 1993 y se expresan en índices de volumen físico con base fija en el año 2013=100, los cuales son de tipo Laspeyres, publicándose índices mensuales, índices acumulados y sus respectivas variaciones anuales. Su cobertura geográfica es nacional e incorpora a los sectores económicos: 21. Minería; 22. Generación, transmisión y distribución de energía eléctrica, suministro de agua y de gas por ductos al consumidor final; 23. Construcción y 31-33. Industrias manufactureras y sus subsectores de acuerdo con el Sistema de Clasificación Industrial de América del Norte 2013 (SCIAN), alcanzando una representatividad del 97% del valor agregado bruto del año 2013, año base de productos del SCNM.</p> <p>Unidad de medida: Índice de volumen físico 2013=100.</p> <p>Cifras ajustadas por estacionalidad.</p> <p>Periodicidad: Mensual.</p> <p>Rezago en publicación: 52 días hábiles de concluido el trimestre de referencia.</p>

		Componente del IMAI.
		Unidad de medida: Índice de volumen físico 2013=100.
TCIMAIMAN	Crecimiento Trimestral del Indicador de Actividad Industria Manufacturera	Cifras ajustadas por estacionalidad.
		Periodicidad: Mensual
		Rezago en publicación: 52 días hábiles de concluido el trimestre de referencia.

Indicadores Regionales

		Banco de México en base a las cifras ajustadas por estacionalidad del Indicador Trimestral de la Actividad Económica Estatal del INEGI.
		Unidad de medida: Índice IT-2013=100
TCITAER	Crecimiento Trimestral del Indicador de Actividad Económica Regional	Regiones: Norte, Centro Norte, Centro y Sur.
		Cifras ajustadas por estacionalidad.
		Periodicidad: Trimestral.
		Rezago en publicación: 4 meses.
		Regiones: Norte, Centro Norte, Centro y Sur.
		Cifras ajustadas por estacionalidad.
TCITAERMAN	Crecimiento Trimestral del Indicador Regional de Actividad Manufacturera	Periodicidad: Trimestral.
		Rezago en publicación: 4 meses.
		Fuente: Banco de México con base en series ajustadas por estacionalidad del Indicador Mensual de Actividad Manufacturera por Entidad Federativa del INEGI.

Fuente: Elaboración propia con información de INEGI y Banco de México.