ÍNDICES DE SENTIMIENTO REGIONALES Y SU ASOCIACIÓN CON INDICADORES OPORTUNOS DE ACTIVIDAD ECONÓMICA EN MÉXICO, 2016-2021

REGIONAL SENTIMENT INDEXES AND THEIR ASSOCIATION WITH TIMELY INDICATORS OF ECONOMIC ACTIVITY IN MEXICO, 2016-2021

Leonardo E. Torre

Facultad de Economía, Universidad Autónoma de Nuevo León
Banco de México

Eva E. González

Banco de México

Luis R. Casillas

Banco de México

Jorge A. Alvarado

Banco de México

Resumen: Estimamos índices de sentimiento a nivel regional y nacional utilizando información en formato de texto del Programa Trimestral de Entrevistas a Directivos, empleada para la elaboración del Reporte sobre las Economías Regionales del Banco de México, referente a los factores que los entrevistados consideran que afectaron, afectan o pudieran afectar la actividad económica en su sector o entidad federativa. Estos índices, estimados con información de los programas de entrevistas trimestrales llevados a cabo entre enero de 2016 y enero de 2021, son posteriormente asociados con diferentes indicadores de actividad económica regional y nacional publicados por el INEGI, obteniéndose correlaciones positivas y estadísticamente significativas entre los índices de sentimiento y algunos indicadores de actividad económica. Dado que estos índices de sentimiento pueden obtenerse con mayor rapidez que la mayoría de los indicadores tradicionales de actividad económica aquí analizados, el trabajo destaca la relevancia de la información en formato de texto contenida en el Programa Trimestral de Entrevistas a Directivos para complementar la obtenida con indicadores tradicionales.

Abstract: We estimate sentiment indexes at the regional and national level using text data obtained from the Programa Trimestral de Entrevistas a Directivos del Banco de México, used to elaborate the Report on the Regional Economies regarding the factors that the interviewed consider affected, affect, or could affect economic activity in their sector or state. Using text data from quarterly interviews performed from January 2016 to January 2021, we associate these indexes with different indicators of regional and national economic activity published by INEGI. The estimates indicate positive and statistically significant correlations among the sentiment indexes and some economic activity indicators. Since the sentiment indexes can be estimated relatively faster than most of the traditional economic indicators analyzed, this paper outlines the relevance of the text data contained in the Programa Trimestral de Entrevistas a Directivos to supplement the information obtained from traditional indicators.

Clasificación JEL/JEL Classification: C45, R11, R15

Palabras clave/keywords: análisis de sentimientos; aprendizaje automático; análisis regional; México

Fecha de recepción: 14 III 2023 *Fecha de aceptación: 26 X 2023*

https://doi.org/10.24201/ee.v39i2.455

 https://doi.org/10.24201/ee.v39i2.455

1. Introducción

El Banco de México realiza encuestas y entrevistas a diversos agentes económicos, a nivel nacional, 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 2011 en los meses de enero, abril, julio y octubre de cada año, capturan información cualitativa y en formato de texto sobre: 1) actividad económica, 2) perspectivas y 3) riesgos para la actividad económica de las regiones en las que 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

[1] La información de las fuentes entrevistadas en los PED es tratada de manera estrictamente confidencial por el Banco de México. Por ello, la información recibida para este trabajo se encuentra codificada.

[2] De acuerdo con la información obtenida, 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 a lo largo del documento haremos referencia a estos meses para indicar el periodo de realización del PED utilizado para elaborar el RER correspondiente. Un porcentaje de las fuentes entrevistadas un trimestre no se consulta el siguiente trimestre, entrevistándose en su lugar a otros directivos. Adicionalmente, el número total de entrevistados por trimestre varía, si bien deben entrevistarse al menos a 440 fuentes, 110 por cada región.

[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 sur está compuesto por: Campeche, Chiapas, Guerrero, Oaxaca, Quintana Roo, Tabasco, Veracruz y Yucatán.

de respuesta a 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 provista por el RER, también existe abundante literatura, identificada como análisis de sentimientos, que muestra que la información en formato de texto refleja sentimientos o actitudes de individuos, empresas y grupos especializados, entre otros, en torno a condiciones económicas recientes, 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 y aplicándose metodologías de análisis de sentimientos encaminadas a utilizar la información en formato de texto tan pronto esté disponible, como la que ofrece un PED, para generar 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 que pudiera utilizarse en su elaboración, esos índices numéricos -también llamados "índices de sentimiento"- pueden generarse de manera más oportuna que los indicadores económicos tradicionales basados en datos duros o suaves. Esto es de relevancia para los formuladores de política económica, quienes normalmente toman decisiones en tiempo real utilizando información incompleta, rezagada y sujeta a revisiones frecuentes. Hasta cierto grado, lo que proponemos en este trabajo es estimar índices de sentimiento dentro un marco similar al *nowcasting*.

El nowcasting se refiere a la predicción del pasado, presente o futuro muy cercano de una variable objetivo, permitiendo mitigar algunas de las dificultades derivadas de los rezagos que caracterizan la publicación de los indicadores económicos. El nowcasting suele realizarse con técnicas como el aprendizaje automático, las cuales

[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 a través de encuestas o entrevistas para medir, por ejemplo, la satisfacción de consumidores en relación con un bien o servicio, o revisar su calidad (Stsiopkina, 2022).

permiten procesar grandes volúmenes de información, numérica y en formato de texto, en tiempo real, rezagado, o tan pronto ésta se encuentre disponible, y que pudiera estar relacionada con la variable objetivo. Un ejemplo de nowcasting empleando algoritmos de aprendizaje automático para el caso mexicano es Campos-Vázquez y López-Araiza (2020), quienes aplican dos modelos de aprendizaje automático: LASSO (*Least Absolute Shrinkage and Selection Operator*) y bosques aleatorios (*Random Forest*), para generar un pronóstico de la tasa de desempleo nacional, antes de que el Instituto Nacional de Estadística y Geografía (INEGI) publique la información, utilizando datos en tiempo real de un índice de Google Trends. Este índice se genera como el cociente del número de búsquedas del término especificado por el usuario (empleo) sobre el total de búsquedas en Google en ese momento. Otro ejemplo de nowcasting para el caso mexicano es el que ofrecen González y Herman (2020), quienes utilizan información cuantitativa con frecuencia diaria o semanal y mensual para pronosticar el nivel del tipo de cambio peso mexicano/dólar estadounidense para el corto plazo (diaria, semanal o mensual) empleando modelos de aprendizaje automático. En específico, los autores utilizan cinco modelos para generar los pronósticos: regresión logística/lineal, regresión logística/lineal regularizada, máquina de soporte vectorial (VSM), potenciación del gradiente/regresión (GBC/GBR5) y redes neuronales. El estudio concluye que, entre los modelos estimados, los VSM y de potenciación del gradiente son los que producen los mejores resultados en términos de precisión y beneficios acumulados.

Este documento ofrece una contribución más al análisis económico para el caso mexicano, empleando técnicas de aprendizaje automático para obtener índices de sentimiento, los cuales son el objetivo principal de este trabajo. En particular, y de acuerdo con nuestro conocimiento, se presentan las primeras estimaciones de índices de sentimiento a partir de la información en formato de texto contenida en el PED. Los índices de sentimiento se calculan 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

[5] Los índices se ponderan con base en la participación del PIB de un sector en una región (o a nivel país) 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.

texto se clasificó por dos de los autores en una de tres categorías, dependiendo de si trasmitían un sentimiento positivo, un sentimiento negativo o un sentimiento neutral (o no definido). Esta clasificación, o etiquetado, se realizó de manera manual, con el apoyo de tres algoritmos de aprendizaje automático (máquina de soporte vectorial, redes neuronales recurrentes y representación de codificador bidireccional de transformadores). Un aspecto relevante de la metodología propuesta aquí es que, si bien la información utilizada no está disponible en tiempo real, sí puede procesarse de manera inmediata tan pronto esté disponible para generar índices de sentimiento. Los indicadores buscan determinar si éstos proveen información más oportuna para señalar la dirección que tomarán algunos indicadores de actividad económica del INEGI antes de que esta institución los publique.

El trabajo utiliza la información en formato de texto de 9,802 entrevistas realizadas trimestralmente 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 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 se asocian con diferentes indicadores de actividad económica regional y nacional publicados por el INEGI.

La relación entre los índices de sentimiento ponderados por región y nacionales con los indicadores del INEGI se cuantificó mediante coeficientes de correlación de Pearson. Así, se obtuvieron correlaciones entre los índices de sentimiento ponderados, con los siguientes tipos de indicadores: 1) dos indicadores mensuales suaves de actividad económica nacional: el Indicador de Pedidos Manufactureros y el Indicador de Confianza Empresarial del Sector Manufacturero; 2) cuatro indicadores duros nacionales: las tasas de crecimiento trimestral del producto interno bruto (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 (IGAE); y 3) 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 sobresalen que: 1) los índices de sentimiento ponderados nacionales obtenidos con diferentes métodos muestran patrones en el tiempo muy similares entre ellos; 2) 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; 3) los índices de sentimiento nacionales muestran correlaciones positivas con los dos indicadores oportunos nacionales de actividad económica; 4) 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; 5) 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; y 6) que las correlaciones entre los índices de sentimiento con los indicadores duros son menores a las correlaciones entre los indicadores suaves y duros.

La principal aportación de este trabajo es mostrar que la información en formato de texto contenida en las respuestas de los directivos entrevistados trimestralmente por el Banco de México en diferentes regiones del país para la elaboración del RER, al ser procesada con técnicas de aprendizaje automático, da lugar a índices de sentimiento que están correlacionados positivamente con diversos indicadores de coyuntura económica del INEGI. Más aún, estos índices de sentimiento podrían generarse -y, por tanto, publicarse- tan pronto se capture la información de un PED, a diferencia de los indicadores duros del INEGI, cuya publicación puede retrasarse desde cuatro semanas hasta cuatro meses.[6]

El trabajo se organiza como se indica a continuación. La segunda sección presenta una breve reseña del enfoque de análisis de sentimientos y de la utilización de técnicas de aprendizaje automático para generar índices de sentimiento. La tercera sección revisa la metodología utilizada para estimar los índices de sentimiento que se presentan en este trabajo. La cuarta sección muestra los índices de sentimiento estimados y sus respectivas correlaciones con los ocho indicadores de actividad económica regional y nacional. La quinta sección presenta los 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

[6] Para los rezagos en la publicación de los indicadores de INEGI utilizados en este trabajo, véase el cuadro A10 en el apéndice.

conservar y procesar esa información, generaron incentivos para que empresas, proveedores de bienes y servicios, políticos, investigadores, entre otros actores, intentaran aprovecharlas para su toma de decisiones (D'Andrea *et al.*, 2015). Un porcentaje significativo de esta información, no obstante, se caracterizó por ser presentada en formato de texto, poco sistematizada y sobre la cual las metodologías econométricas tradicionales no podían aplicarse. Por ello, se ha impulsado la adopción de técnicas alternativas, capaces de aprovechar esta información. Una de ellas es el análisis de sentimientos,[7] el cual hace referencia a procesos o métodos que permiten detectar el sentimiento positivo o negativo contenido en un texto, ya sea 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, de tal forma que actualmente 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 (D'Andrea *et al.*, 2015). 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):

[7] El análisis de sentimientos surge 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).

[8] Es importante distinguir entre un "índice de sentimiento" y un "indicador de confianza". De acuerdo con Santero y Westerlund (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,

 https://doi.org/10.24201/ee.v39i2.455

$$Indice\ de\ Sentimiento\ A = \frac{Positivos - Negativos}{Positivos + Negativos} \tag{1}$$

$$Indice\ de\ Sentimiento\ B = \frac{Positivos - Negativos}{Positivos + Negativos + Neutrales} \tag{2}$$

Estas definiciones se basan en los conteos de los documentos obtenidos mediante algoritmos, donde *positivos* se refiere a la cuenta total de documentos que, de acuerdo con los criterios de clasificación adoptados, trasmiten un sentimiento positivo; *negativos* se refiere a la 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: 1) el análisis de texto basado en diccionarios,[9] 2) el aprendizaje automático,[10] y 3) los enfoques híbridos.[11] Enseguida se explica brevemente en qué consiste el segundo enfoque, ya que es éste al que pertenecen las técnicas utilizadas en este trabajo para clasificar los documentos. Es conveniente mencionar que, una vez obtenidos los índices de sentimiento, se les comparará con índices que se desprenden de conteos parciales o totales de los documentos clasificados de

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, véanse Garrett et al. (2004), De Bondt y Schiaffi (2015), Salhin et al. (2016) y Benhabib y Spiegel (2017). Para el caso de México, véase Díaz y Huerta (2020).

[9] Turney (2002) y Hatzivassiloglou y McKeown (2002) presentan revisiones de literatura relacionada con este tema.

[10] Véase Medhat et al. (2014) para una exposición sobre aprendizaje automático.

[11] Véase Prabowo y Thelwall (2009) para una exposición sobre el enfoque híbrido.

manera manual por los investigadores interesados (índices anotados), como referencia a 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 "anotados".

2.1 *Algoritmos de aprendizaje automático*

El aprendizaje automático hace referencia a algoritmos enfocados principalmente en predecir, clasificar, agrupar o generar clústeres de datos. Son 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] Estos algoritmos se clasifican en cuatro tipos: aprendizaje supervisado, aprendizaje no supervisado, semi-supervisado y por refuerzo.[14] En este trabajo solo utilizamos algoritmos pertenecientes a las primeras dos categorías. De acuerdo con Athey (2018), 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 $\mu(x) = E[Y|X = x]$ y que haga un buen trabajo para

[12] Es oportuno mencionar que el aprendizaje automático 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 (Arora, 2021; 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] Véase 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.

[15] Se supone que las observaciones son independientes y 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 automático funcionen (Athey, 2018).

predecir los verdaderos valores de Y en un conjunto independiente de datos (Athey, 2018).[16]

Sobre los algoritmos de aprendizaje no supervisado, el mismo autor indica que éstos 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 utilizado principalmente 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 se conocen como 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).[17]

2.2 *Algunas aplicaciones en bancos centrales*

El uso de estas herramientas en los bancos centrales se ha extendido en años recientes. Entre los realizados en algunos de los principales bancos centrales del mundo puede mencionarse, por ejemplo, el estudio de Suss y Treitel (2019) sobre el Banco de Inglaterra. Estos autores utilizan un algoritmo de aprendizaje automático supervisado (árboles de decisión aleatorios) para desarrollar un sistema de alerta temprana en torno a vulnerabilidades en el sistema bancario de Inglaterra el cual, de acuerdo con su evaluación, genera mejores resultados que los obtenidos con 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 Sistema de la Reserva Federal 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 automático para identificar a directivos de empresas que afirman no tener una expectativa cuantitativa de inflación,

[16] El párrafo original, en idioma inglés, es autoría de Athey (2018: 4).

[17] El párrafo original, en idioma inglés, es autoría de Athey (2018: 3).

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.[18]

Por su parte, Azqueta-Gavaldón *et al.* (2020), del Banco Central Europeo, utilizan técnicas de aprendizaje automático 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 *et al.* (2021) aplican técnicas de análisis de texto a mensajes de Twitter en español correspondientes al 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 correlaciona positivamente con medidas de riesgo financiero, volatilidad del mercado accionario, riesgo de default soberano y volatilidad cambiaria.[19]

---

[18] Una exposición de la técnica de "propensity score matching" está disponible en Rosenbaum y Rubin (1983).

[19] Si bien la revisión de trabajos que aplican técnicas de aprendizaje automático 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 automático 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 automático 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

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 automático 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 automático 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 automático 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 pre-procesamiento 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.[20] Enseguida describimos, a la luz de estos pasos, el proceso para identificar los modelos de aprendizaje automático adoptados en este trabajo para generar el etiquetado necesario para calcular los índices de sentimiento.

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 capital humano para darle servicio.

[20] El cuadro A1 del apéndice muestra 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 mediante entrevistas en persona o por teléfono, 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.[21,22] La primera recaba información sobre el desempeño de la empresa o sector económico para 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 que impulsarían o afectarían el desempeño, en el corto plazo, de la entidad federativa en la que opera su empresa. Las entrevistas constan de preguntas 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.[23] Entre enero de 2011 y 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.[24]

La información recabada en las entrevistas ha sido utilizada para: 1) generar índices de difusión de actividad económica; y 2) como complemento descriptivo en los análisis sectoriales y regionales publica-

[21] 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.

[22] Conviene tener presente que el RER del cuarto trimestre (4T) de un año determinado utiliza información del PED realizado en enero del año entrante; el RER del primer trimestre (1T) utiliza información del PED de abril, el RER del segundo trimestre (2T) se apoya en las entrevistas del PED de julio, y el RER correspondiente al tercer trimestre (3T) utiliza la información del PED realizado en octubre.

[23] Las preguntas utilizadas en los PED a lo largo del tiempo están disponibles a solicitud del interesado.

[24] Véanse la nota al pie 3 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.

dos en el RER.[25] Las respuestas en formato de texto, o documentos, obtenidos a partir de las preguntas abiertas relacionadas con factores que 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 3. La intención de este trabajo es utilizar estos documentos para construir índices de sentimiento.[26] Descartando la "no respuesta", el universo de información disponible para la elaboración de este trabajo (enero de 2011 a enero de 2021) está integrado por 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 serán utilizados (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) y se obtuvo 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 de forma empírica, buscando identificarlos con base en palabras o frases utilizadas en la disciplina de Economía.[27] De esta manera, como primer paso del

[25] 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 A3 del apéndice presenta la relación de esta clasificación con el Sistema de Clasificación Industrial de América del Norte.

[26] En los casos de preguntas abiertas puede ocurrir que el entrevistado no ofrezca comentarios.

[27] El algoritmo LDA es un modelo estadístico que define una distribución pro-

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, 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 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
Las 25 palabras relevantes más frecuentes en los PED
Enero 2011 - enero 2021

a) Unigrama			
Palabra	*Frecuencia*	*Palabra*	*Frecuencia*
actividad	12,002	inversión	5,792
Año	9,563	mayor	5,709
cambio	9,203	mercado	5,581
demanda	8,897	nuevo	5,182
económico	8,727	poder	5,179
empresa	8,343	precio	5,171
esperar	7,950	público	5,159
estado	7,666	sector	5,142
estados	7,354	ser	5,124

babilística 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 (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 (Anandarajan et al., 2019).

 https://doi.org/10.24201/ee.v39i2.455

Cuadro 1
(Continuación)

a) Unigrama			
Palabra	*Frecuencia*	*Palabra*	*Frecuencia*
gobierno	7,023	tener	5,101
incertidumbre	6,946	unidos	5,067
incremento	6,593	venta	5,043
inseguridad	5,821		
b) Bigrama			
Palabra	*Frecuencia*	*Palabra*	*Frecuencia*
actividad económico	5,095	materia prima	728
cambio gobierno	4,485	mayor demanda	725
covid 19	1,563	nuevo gobierno	712
demanda producto	1,446	nuevo proyecto	689
demanda servicio	1,412	obra público	642
estados unidos	1,141	parte gobierno	632
gasto público	1,016	poder adquisitivo	631
gobierno estatal	1,005	reforma fiscal	621
gobierno federal	845	sector automotriz	617
incremento precio	829	tasa interés	602
industria automotriz	814	tipo cambio	591
inversión extranjero	778	volatilidad tipo	570
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

Cuadro 1
(Continuación)

c) Trigrama			
Palabra	*Frecuencia*	*Palabra*	*Frecuencia*
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 pre-procesamiento 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 tópicos representados como una mezcla de palabras y con esto, los investigadores determinan cuáles son los más relevantes. El cuadro 2 muestra los siete tópicos detectados, así como los términos más recurrentes en cada uno. 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.[28] Este análisis permitió determinar aquellos 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 estos pueden brindar contexto a un indicador, en este caso, al IGAE. La gráfica destaca, por ejemplo, cómo la relevancia de los tópicos cambia en el tiempo. Así, la reforma fiscal es el más relevante del 3T-2013;

[28] 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.

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 del 3T-2016.[29] En el 2T-2018 el proceso electoral en México y la renegociación del Tratado de Libre Comercio de América del Norte (TLCAN) son los que resaltan; mientras que en el 1T-2020 el tipo de cambio y el COVID-19 son los relevantes. Finalmente, en el 2T-2020, cuando se contrae la actividad económica en México, el COVID-19 es el que destaca; en tanto que la vacunación en México y la economía de Estados Unidos aparecen como los dominantes durante el 1T-2021, acompañando así a la recuperación económica. Debe destacarse 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 laborar el RER del 4T-2015, lo que elevó el número de respuestas por sección, permitiendo visualizar mejor tópicos regionales.[30,31]

[29] Las expresiones 1T, 2T, 3T y 4T se refieren a los trimestres 1, 2, 3 y 4 de un determinado año.

[30] 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 4T-2016 en la sección de actividad económica es posible identificar cuatro respuestas divididas en: 1) factores externos de impulso, 2) factores internos de impulso, 3) factores externos limitantes y 4) factores internos limitantes. Estas preguntas están disponibles a solicitud del interesado.

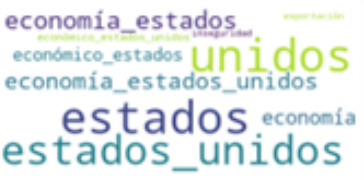
[31] El cuadro A4 del apéndice presenta los tópicos relevantes por región en los PED de enero de 2015 a enero de 2021. Este no incluye los términos más recurrentes, pero están disponibles a solicitud del interesado.

Cuadro 2
Tópicos enero 2011 - enero 2021

Tópico	*Términos más recurrentes*
1. Ninguno	
2. Seguridad pública:	
3. Mercado cambiario	
4. Obras públicas/inversión pública	

 https://doi.org/10.24201/ee.v39i2.455

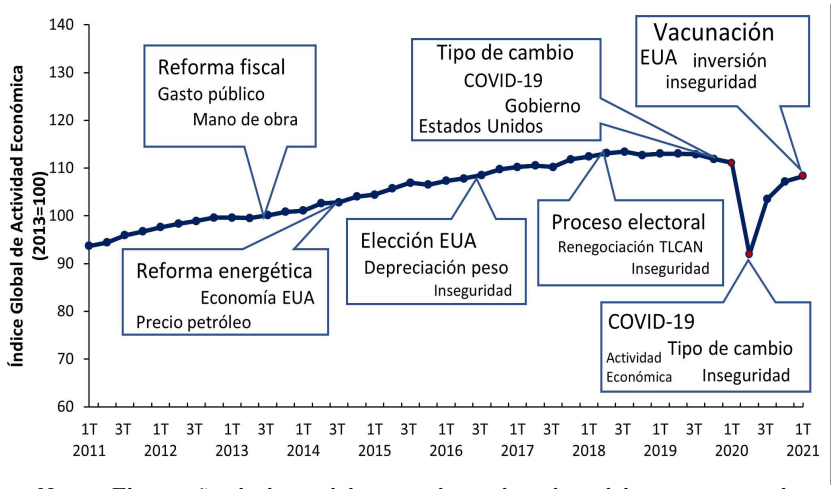
Cuadro 2
(Continuación)

Tópico	*Términos más recurrentes*
5. Inversión	
6. Actividad económica de Estados Unidos	
7. Incertidumbre	

Notas: 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.

Gráfica 1
IGAE y tópicos relevantes



Nota: El tamaño de las palabras en las nubes de palabras corresponde a la frecuencia de la palabra en el tópico.

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

3.3 *Pre-procesamiento de la información*

Antes de entrenar un modelo de aprendizaje automático se requiere realizar el pre-procesamiento 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 cuarta sección utilizarán solo los documentos obtenidos de los PED de enero de 2016 a enero de 2021. Para este trabajo se llevaron a cabo las siguientes acciones:

1. Se realizó una corrección ortográfica de los textos.
2. Se identificaron palabras en inglés que se tradujeron al idioma español.
3. 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*).
4. Se eliminaron números y signos de puntuación.

5. Se lematizaron las palabras, lo que significa agruparlas de acuerdo con su significado.
6. Todas las letras mayúsculas se cambiaron a minúsculas.
7. 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 pre-procesar la información, procedimos a revisar diversas metodologías de aprendizaje automático 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 las etiquetas "positivo", "negativo" y "neutral" a cada uno de los documentos obtenidos en las entrevistas de los PED de enero de 2016 a enero de 2021.[32] Si bien los PED del RER iniciaron en enero de 2011, optamos por etiquetar solo los documentos de las entrevistas 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 automático. En relación con esta asignación de etiquetas, algunos

[32] Existen modelos que entrenan diccionarios 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.

trabajos de investigación reportan haber etiquetado todos los documentos, o bien, una fracción de estos, mediante una clasificación manual de dos etiquetadores de manera independiente.[33] En ese trabajo, dos de los autores se encargaron de revisar y etiquetar, por separado, cada uno de los 76,895 documentos del periodo enero de 2016 a enero de 2021 en positivo, negativo y neutral, de acuerdo con criterios como los que se muestran en el cuadro 3. Es conveniente señalar que estos 76,895 documentos se dividieron, a su vez, en dos subconjuntos. El primero, con 61,516 documentos (80% del total), el cual será utilizado en la etapa de "entrenamiento" de los modelos. El segundo, con los 15,379 documentos restantes (20% del total), el cual se usará en la etapa de "evaluación" del desempeño de dichos modelos. Las etapas de entrenamiento y evaluación se explican más adelante.

Para el etiquetado 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.

[33] Para una referencia en la que el etiquetado es realizado de manera independiente por dos investigadores, véase Aragón et al. (2019). Sobre el tema también consultamos, en enero de 2020, al Dr. Víctor Muñiz, catedrático del Centro de Investigación en Matemáticas A.C. en Monterrey (CIMAT Monterrey) y especialista en ciencia de datos. Debemos reconocer que los etiquetados "manuales" son susceptibles a errores. Además, emplear solo dos "etiquetadores" y asignar "neutral" a los casos en los que discrepan ambos investigadores genera puede generar pérdida de información. No obstante, en la literatura se encuentran ejemplos en los que se utilizan solo dos etiquetadores (Aragón et al., 2019).

 https://doi.org/10.24201/ee.v39i2.455

Cuadro 3
(Continuación)

	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.

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 automático, 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, del total de los documentos revisados y clasificados por los etiquetadores (76,895), 80% se seleccionó 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.

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.

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: máquinas de vectores soporte (SVM, por sus siglas en inglés), Naive Bayes, clasificador Ridge y árbol de decisión. Cada clasificador se probó con las siguientes representaciones del texto: bolsa de palabras (*bag of words*); frecuencia de término-frecuencia inversa de documento (TF-IDF); y los LSA 100, 300, 500 y 1000.[34] Como resultado de estas pruebas se seleccionaron varios modelos. El algoritmo SVM, con la representación de texto TF-IDF (SVM+TF-IDF), mostró el mejor desempeño, indicado por el mayor

[34] En la literatura de aprendizaje automático, 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, véase Jurafsky y Martin (2020).

valor del estadístico F1-Macro, por lo cual se utilizará como referencia para comparar otros modelos (cuadro 4).[35] El F1-Macro se utiliza cuando se tienen datos desbalanceados, como ocurre en nuestra base de datos, ya que ésta se integra con 30,715 documentos positivos, 31,156 negativos y 15,024 neutrales (76,895 documentos en total). El F1-Micro es otro indicador de ajuste; éste, sin embargo, otorga el mismo peso a cada uno de los documentos clasificados y que aquí se presenta solo para propósitos de comparación. Sin embargo, ambos F1 arrojan el mismo resultado.

Finalmente, se probaron los modelos de RNN y BERT.[36] El mejor modelo de RNN fue el que utilizó celdas GRU y vectores pre-entrenados FastText.[37] En el caso de los modelos basados en BERT, se evaluaron dos modelos pre-entrenados 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 del que se dispone (cuadro 5).[38]

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.[39]

---

[35] El F1 es una medida diseñada para evaluar qué tan bien predice un modelo y se define como: $F1 = 2 * \left(\frac{Presición * Recuperación}{Presición + Recuperación}\right)$, 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 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 A5 del apéndice presenta este y otros criterios de evaluación de modelos.

[36] El cuadro A6 del apéndice muestra una explicación en torno al entrenamiento de los modelos RNN y BERT.

[37] Véase Cho et al. (2014) para información sobre las celdas GRU. Y Bojanowski et al. (2017) para una revisión de los vectores pre-entrenados FastText

[38] Para consultar sobre modelos BERT, véase Devlin et al. (2018).

[39] 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 con 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

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 suaves e indicadores duros de actividad económica.

Los índices de sentimiento a nivel regional y nacional que aquí se presentan son los que se desprenden de los clasificadores SVM, RNN y BERT. Éstos se obtienen como se describe a continuación: una vez que se tienen etiquetados todos los documentos, se toman como base

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 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.

los conteos de cada categoría en la que estos 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).[40] Los índices de sentimiento regional (ISR) se construyen, a su vez, con las formulaciones que se definen en las expresiones 3 y 4 utilizando la clasificación derivada de los algoritmos:

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

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

Al índice (3) le llamamos "índice 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 $j(t)$.[41] Por ejemplo, si la entrevista se realizó en enero de 2017, la

[40] 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).

[41] Las estimaciones de los ponderadores sectoriales a 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

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.[42,43] De acuerdo con estas 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 a nivel país. 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 a nivel nacional, también con las fórmulas (3) y (4), 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

año se apoyan en los datos del año correspondiente. Otro punto por 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 por COVID-19.

[42] 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 lo opinión del entrevistado de la región B.

[43] Se estimaron correlaciones entre los distintos índices ponderados y no ponderados. A nivel nacional, estas resultaron superiores a 0.88, en tanto que, en las regiones norte, centro norte y centro, resultaron superiores a 0.81. En el sur las correlaciones superan 0.70, en todos los casos, con excepción de los índices BERT, donde resultaron ligeramente superiores a 0.40. Esto sugiere que las correlaciones entre los índices ponderados y los no ponderados son similares bajo esta métrica.

de sentimiento cercanos a los obtenidos con los respectivos índices anotados bajo el supuesto de que el etiquetado manual se realizó de manera adecuada.[44]

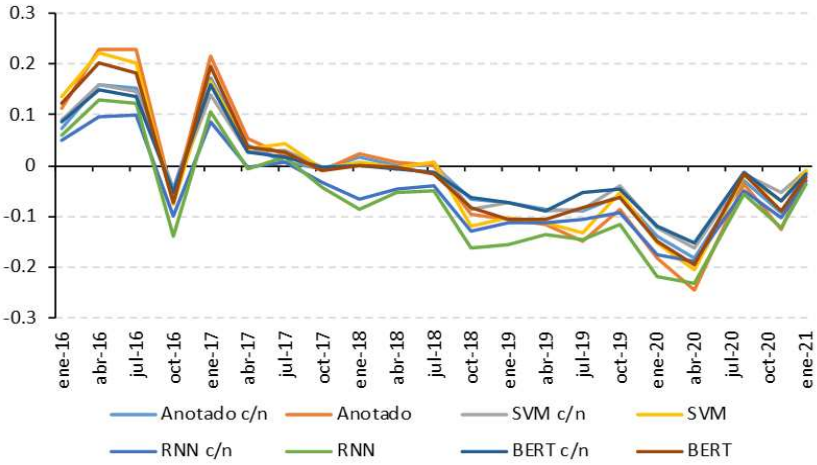
4.1 *Índices de sentimiento nacionales y regionales ponderados*

Un elemento esencial en el aprendizaje automático 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, a 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 automático, 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 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. Ahí 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 automático. Además, abonan a la posibilidad de que cuando estos índices se asocien con los diferentes indicadores de actividad, se obtengan resultados similares.

[44] Se debe tener presente que el etiquetado manual se realizará sólo 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 (3) y (4).

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	0.985	0.9839	0.998	1.0000		

 https://doi.org/10.24201/ee.v39i2.455

Cuadro 6
(Continuación)

*Indicador de Sentimiento**	*Anotado c/n*	*Anotado*	*SVM c/n*	*SVM*	*RNN c/n*	*RNN*	*BERT c/n*	*BERT*
BERT c/n	0.9909	0.9867	0.9893	0.9801	0.98	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 4T-2015 y el 3T-2016, ésta se caracterizó por un estancamiento del sector industrial y, en particular, de las manufacturas (Banco de México, 2015, 2016a, 2016b). Ahora bien, en enero de 2016, los índices de sentimiento se ubicaron en un nivel bajo, subieron en abril, pero retrocedieron ligeramente en julio de ese año. En octubre de 2016, los distintos índices de sentimiento cayeron de manera precipitada, coincidiendo con una marcada volatilidad en los mercados financieros atribuible al proceso electoral en Estados Unidos, así como a una elevada volatilidad y una depreciación significativa de la moneda nacional (Banco de México, 2016c).

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 privada en México. En enero de 2017, la creciente expectativa de que la Administración Trump buscaría renegociar y no cancelar 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 coincidió, a su vez, con una baja en los índices de sentimiento.

El relativo estancamiento de estos índices a lo largo de 2019 estuvo 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 en enero de 2020, respecto de octubre 2019, se presentó en una etapa caracterizada por tensiones comerciales globales, riesgos geopolíticos y los riesgos de una propagación global de 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 coincidió con el resguardo de las familias mexicanas para evitar contagiarse del virus, los cierres de actividades no esenciales por parte de las autoridades 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 (Banco de México, 2020b).

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 recuperaron, 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 recuperaron 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 (Banco de México, 2020c, 2021). La reciente recuperación de los índices, no obstante, ha resultado insuficiente para recuperar los niveles que alcanzaron en enero de 2017.

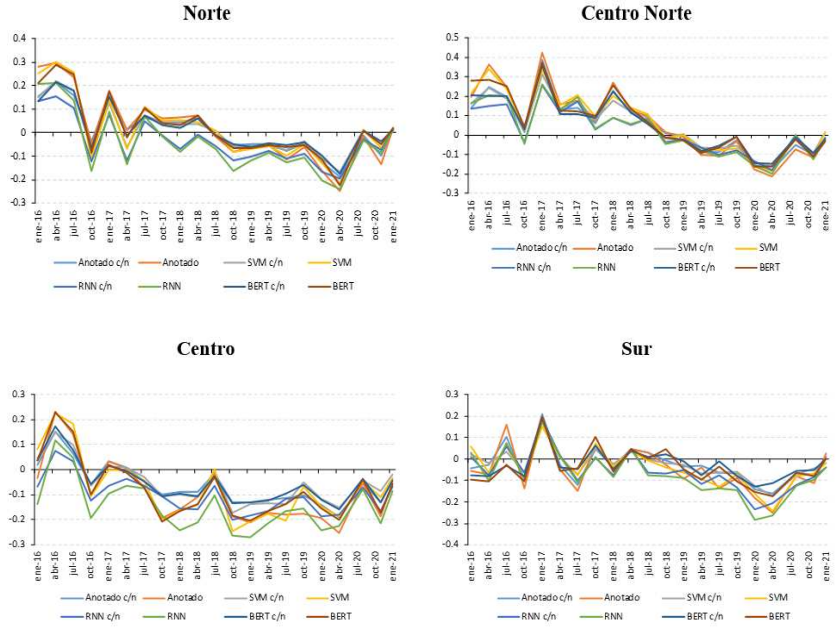
La gráfica 3 muestra, a su vez, los índices de sentimiento ponderados regionales (norte, centro norte, centro y sur), donde se aprecian patrones similares entre dichos índices al interior de cada región. Asimismo, puede apreciarse que es en la región centro norte donde los distintos índices de sentimiento 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.

[45] Sobre los factores detrás de la desaceleración, véase Banco de México (2019).

[46] Sobre estos eventos, véase Banco de México (2020a).

 https://doi.org/10.24201/ee.v39i2.455

Gráfica 3
Índices de sentimiento regionales 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.

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 a 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

Region	*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	0.961	0.972	0.9942	1.0000				
	RNN c/n	0.9341	0.9413	0.971	0.9703	1.0000			
	RNN	0.9257	0.9413	0.9652	0.9722	0.9959	1.0000		
	BERT c/n	0.9795	0.9688	0.986	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	0.994	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	0.978	0.978	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

Cuadro 7
(Continuación)

Region	*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	0.9797	1.0000						
	SVM c/n	0.9536	0.9437	1.0000					
Centro	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	0.946	0.9199	1.0000	
	Bert	0.957	0.9712	0.9642	0.9722	0.9338	0.932	0.9848	1.0000
	Anotado c/n	1.0000							
	Anotado	0.9712	1.0000						
	SVM c/n	0.9268	0.8895	1.0000					
Sur	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	0.945	0.9565	0.9864	1.0000		
	BERT c/n	0.8482	0.7668	0.848	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

Notas: 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 con un nivel de significancia de 95%.

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

4.2 *Índices de sentimiento versus 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 a nivel nacional publicados por el INEGI: el Indicador de Pedidos Manufactureros (IPM) y el Indicador de Confianza Empresarial del Sector Manufacturero (ICEMP).[47] En estos indicadores no existe un rezago en su publicación; es decir, una vez que cierra el mes de referencia para el levantamiento de la información necesaria para elaborar dicho indicadores, su reporte del mes se publica el primer día hábil del siguiente mes.[48] 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 a 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

[47] Se realizaron pruebas Dickey-Fuller Aumentadas a las series de los índices de sentimiento (nacionales y regionales) y de los indicadores económicos. Al ser el tamaño de muestra de solo 20 observaciones en cada serie, reconocemos que éstas enfrentan una crítica sobre su robustez. El cuadro A7 del apéndice presenta los resultados, donde puede verse que todas son estacionarias. El cuadro muestra también estadísticos descriptivos de los distintos indicadores económicos.

[48] Información básica en torno a elementos metodológicos de a los indicadores de actividad regionales y nacionales del INEGI puede consultarse en el cuadro A10 del apéndice.

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 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 a 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	***	0.698	***
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 versus indicadores duros de actividad económica nacionales*

Esta sección analiza en qué medida los índices de sentimiento nacionales se asocian con indicadores duros de actividad económica. Las variables son tasas de crecimiento entre el periodo t y el periodo

$t-1$, y todas son estacionarias (cuadro A7 del apéndice). 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).

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	*	0.355	*	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		0.337	*	0.3456	*
BERT c/n	0.3587	*	0.3227		0.345	*	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, nuevamente, 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

 https://doi.org/10.24201/ee.v39i2.455

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 éste. 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 ocho semanas, y dos de ellos tienen un rezago de cuatro meses.

Cuadro 10
Indicadores de sentimiento regionales vs. TCITAER
Enero 2016 - enero 2021
Coeficientes 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.

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 11
Indicadores de sentimiento ponderados regionales vs. TCITAERMAN
Enero 2016 - enero 2021
Coeficientes de Correlación de Pearson

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 11 muestra 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 alcanzan a capturar la actividad de sus respectivos sectores manufactureros, con la excepción del sur, donde solo un índice de sentimiento (SVM) presentó una correlación significativa, quizá como resultado del poco peso que este sector recibe en las entrevistas.

Cuadro 12
Índices de sentimiento regionales ponderados vs. TCITAER regionales
Enero 2016 - enero 2021
Coeficientes 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.

En cambio, es posible que los índices de sentimiento de, por ejemplo, la región sur, estén más relacionados con indicadores duros del sector de servicios de restaurantes y hoteles, dado el peso que este sector recibe en el reporte. El RER, conviene destacar, distribuye las entrevistas al interior de cada región en función de la participación relativa de los diferentes sectores en el PIB total de su respectiva región.

Esto mismo explicaría por qué en el cuadro 12, la región norte, donde las entrevistas del sector manufacturero reciben mayor peso relativo que en el resto de las regiones, registra una correlación positiva y estadísticamente significativa entre sus índices de sentimiento regionales y el ITAERMAN de su región, algo que no se observa en el resto de las regiones.

Cuadro 13
Índices de sentimiento nacionales ponderados vs. indicadores de actividad económica regionales
Enero 2016 - enero 2021
Coeficientes de Correlación de Pearson

Indicador de sentimiento	*TCITAER Norte*		*TCITAER Centro Norte*		*TCITAER Centro*		*TCITAER Sur*
	r		*r*		*r*		*r*
Anotado c/n	0.334	*	0.378	**	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	**	0.381	**	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.

Un último punto por resaltar, respecto de las asociaciones entre los índices de sentimiento y los indicadores duros de actividad económica, se presenta en el cuadro 13. Este cuadro sugiere que los índices de sentimiento nacionales capturan relativamente mejor los ciclos económicos de las regiones centrales y, en menor medida, del norte. Esto se aprecia en el hecho de que los coeficientes de correlación entre los índices de sentimiento nacionales y los indicadores de actividad económica de las regiones centro norte y centro son mayores y estadísticamente distintos de cero a niveles de significancia de 5% en todos los casos. Esto podría responder a que las regiones guardan

 https://doi.org/10.24201/ee.v39i2.455

relativamente una mayor similitud en su estructura económica con la economía nacional. La actividad económica de la región norte reporta una asociación positiva con los índices de sentimiento nacionales, si bien menos estrecha en comparación con las regiones centrales, y con niveles de significancia de 10%. 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 relativamente más dependiente de la minería petrolera y de la actividad turística.

4.4 *Índices de sentimiento versus indicadores duros de actividad económica nacionales*

Finalmente, es relevante mencionar que, si bien los coeficientes de correlación entre los indicadores de sentimiento y los indicadores duros son positivos, la evidencia señala que los dos indicadores suaves que aquí se presentan (IPM e ICEMP) muestran una correlación positiva y estadísticamente significativa con tres indicadores duros de actividad (TCPIB, TCIMAI y TCIMAIMAN), la correlación entre los dos indicadores suaves y todos los indicadores duros (TCPIB, TCIGAE, TCIMAI y TCIMAIMAN) es notoriamente más alta (cuadro 14). Esto indica que los índices de sentimiento aquí estimados son pertinentes para anticipar la dirección de diversos indicadores de actividad económica; sin embargo, es necesario explorar con más detalle otros algoritmos de aprendizaje automático y continuar actualizando la base de datos para aplicar los algoritmos necesarios en la generación de los índices de sentimiento.

Cuadro 14
Índices de sentimiento nacionales ponderados vs. indicadores de actividad económica regionales
Enero 2016 - enero 2021
Coeficientes de Correlación de Pearson

Indicador 'suave' de act. económica	*TCPIB*		*TCIGAE*		*TCIMAI*		*TCIMAIMAN*	
	Coef. Corr.		*Coef. Corr.*		*Coef. Corr.*		*Coef. Corr.*	
IPM	0.7179	***	0.7515	*	0.8029	***	0.7732	***
ICEMP	0.5123	**	0.5495	**	0.5931	***	0.0863	*

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 PED realizados para la elaboración del Reporte sobre las Economías Regionales del Banco de México, entre enero de 2016 y 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 y de Confianza Empresarial del Sector Manufacturero; en tanto que, con las tasas de crecimiento trimestral del PIB nacional, del Indicador Mensual de Actividad Industrial Total y de la Actividad Industrial Manufacturera, las correlaciones son significativas al 15%. A nivel regional, los índices de sentimiento solo resultaron relevantes para capturar el comportamiento de la tasa de crecimiento del Indicador

Trimestral de la Actividad Económica Regional y de la tasa de crecimiento del Indicador Trimestral de la Actividad Manufacturera Regional en el norte.

Una aportación de este trabajo consiste en haber generado etiquetas para entrenar futuros modelos de predicción basados en aprendizaje automático, frente a la alternativa de generar las etiquetas haciendo uso, por ejemplo, de diccionarios pre-entrenados, que son más rígidos a los cambios de la estructura temática y a la evolución del idioma. Asimismo, los algoritmos entrenados en este documento para la detección de tópicos y clasificación de los sentimientos nacional y regional podrán continuar usándose para procesar, de manera automática, la información en formato de texto capturada en los PED, a fin de estimar índices de sentimiento y obtener información más oportuna de la actividad económica nacional y regional.

No obstante, este trabajo aún ofrece áreas de oportunidad. Por ejemplo, dado que en la asignación manual de las etiquetas se clasificaron como "neutrales" los 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 o la entidad federativa en la que se ubican. Asimismo, no se exploraron medidas alternativas de correlación, las cuales pudieran utilizarse como medidas de robustez de los resultados obtenidos en este trabajo. También, la utilidad de los índices de sentimiento que aquí se obtuvieron pudiera evaluarse de manera más formal al incluirlos, 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 es que los modelos de aprendizaje automático, 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.[49] El trabajo que aquí se presenta debe

[49] De acuerdo con expertos en el tema, una cantidad mínima de datos debe alcanzar, aproximadamente, 80,000 respuestas.

interpretarse, por tanto, como una primera aproximación en cuanto a la capacidad de los índices de sentimiento basados en información en formato de texto del PED del Banco de México para complementar, de manera más oportuna, la información provista por datos duros y suaves de actividad económica en México.

Agradecimientos

Agradecemos a Alejandro Noriega, Diana Martínez, Ramsés Franco y Eduardo Sandoval por todo su apoyo a lo largo de este proyecto, así como a Víctor Muñiz, Hairo Miranda, Tatiana Rueda y Víctor López, de CIMAT-Monterrey, por sus consejos y asesoría durante la elaboración de este documento. Los autores agradecen también los comentarios de Caterina Rho, Raúl Fernández, Alejandrina Salcedo y Juan Carlos Chávez, así como el apoyo de Vanessa Gutiérrez y Luis Fernando Colunga. Todos los errores son responsabilidad de los autores.

Leonardo E. Torre: leonardo.torre@banxico.org.mx

Eva E. González: egonzalezg@banxico.org.mx

Luis R. Casillas: lcasillas@banxico.org.mx

Jorge A. Alvarado: jorge.alvarado@banxico.org.mx

Referencias

Algaba, A., D. Gardia, K. Bluteau, S. Borms y K. Boudt. 2020. Econometrics meets sentiment: An overview of methodology and applications, *Journal of Economic Surveys*, 34(3): 512-547.

Alloghani, M., D. Al-Jumeily, J. Mustafina, A. Hussain y A.J. Aljaaf. 2020. A systematic review on supervised and unsupervised machine learning algorithms for data science, en M. Berry, A. Mohamed y B. Yap (eds.), *Supervised and Unsupervised Learning for Data Science*, Springer.

Anandarajan, M., C. Hill y T. Nolan. 2019. Semantic space representation and latent semantic analysis, en *Practical Text Analytics, Advances in Analytics and Data Science*, Vol. 2, Springer.

Aragón, M., M. Carmona, M. Montes, H. Escalante, L. Villaseñor y D. Moctezuma. 2019. Overview of MEX-A3T at IberLEF 2019: Authorship and aggressiveness analysis in Mexican Spanish tweets, documento presentado en SEPLN Workshop on Iberian Languages Evaluation Forum (IberLEF), Bilbao.

Arora, S. 2021. Data mining vs. machine learning: The key difference, www.simplilearn.com/data-mining-vs-machine-learning-article.

Athey, S. 2018. The impact of machine learning on Economics, en A. Agrawal, J. Gans y A. Goldfarb (eds.), *The Economics of Artificial Intelligence: An Agenda*, National Bureau of Economic Research.

Azqueta-Gavaldón, A., D. Hirschbühl, L. Onorante y L. Saiz. 2020. Economic policy uncertainty in the Euro area: An unsupervised machine learning approach, Working Paper Series No. 2359, European Central Bank.

Banco de México. 2015. Informe trimestral octubre-diciembre, https://www.banxico.org.mx/publicaciones-y-prensa/informes-trimestrales/%7B94CE88E5-3F13-4707-8038-30A0F49D6E47%7D.pdf.

Banco de México. 2016a. Informe trimestral enero-marzo, https://www.banxico.org.mx/publicaciones-y-prensa/informes-trimestrales/%7BA3AA2471-B70C-DAA2-01DF-EA06C6546B6A%7D.pdf

Banco de México. 2016b. Informe trimestral abril-junio, https://www.banxico.org.mx/publicaciones-y-prensa/informes-trimestrales/%7BAD156BB0-60B7-947E-A9EF-9C96AB882667%7D.pdf.

Banco de México. 2016c. Informe trimestral julio-septiembre, https://www.banxico.org.mx/publicaciones-y-prensa/informes-trimestrales/%7BD093DF85-0D83-3DD2-A533-3431AFAFE3A1%7D.pdf.

Banco de México. 2019. Informe trimestral octubre-diciembre, https://www.banxico.org.mx/publicaciones-y-prensa/informes-trimestrales/%7B0DED33B2-FF70-345D-53BE-77EA35A0D743%7D.pdf.

Banco de México. 2020a. Informe trimestral enero-marzo, https://www.banxico.org.mx/publicaciones-y-prensa/informes-trimestrales/%7B23C2DCA8-4AD3-FBE0-B0BF-4D30C8066B84%7D.pdf.

Banco de México. 2020b. Reporte sobre las economías regionales julio-septiembre, https://www.banxico.org.mx/publicaciones-y-prensa/reportes-sobre-las-economias-regionales/%7B8427BCB2-D8F2-C28A-8DD4-EB8DD9770681%7D.pdf.

Banco de México. 2020c. Informe trimestral octubre-diciembre, https://www.banxico.org.mx/publicaciones-y-prensa/informes-trimestrales/%7B81BD569D-DD6E-885A-A67F-5664A37B4148%7D.pdf.

Banco de México. 2021. Informe trimestral enero-marzo, https://www.banxico.org.mx/publicaciones-y-prensa/informes-trimestrales/%7B49D9C039-CE93-FC5A-59A6-DFF7579FDB26%7D.pdf.

Barocas, S. 2014. Data mining and the discourse on discrimination, documento presentado en KDD '23: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, California.

Benhabib, J. y M. Spiegel. 2017. Sentiment and economic activity: Evidence from U.S. states, NBER Working Paper No. 23899.

Blei, D. 2011. Introduction to probabilistic topic models, *Communications of the ACM*, 55(4): 77-84.

Bojanowski, P., E. Grave, A. Joulin y T. Mikolov. 2017. Enriching word vectors with subword information, *Transactions of the Association for Computational Linguistics*, 5: 135-146.

Boyd, D. y K. Crawford. 2012. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon, *Information, Communication and Society*, 15(5): 662-679.

Campos-Vázquez, R.M., B. López-Araiza y E. Sergio. 2020. Grandes datos, Google y desempleo, *Estudios Económicos*, 35(1): 125-151.

Cho, K., B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk y Y. Bengio. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation, ArXiv preprint, arXiv:14 06.1078.

D'Andrea, A., F. Ferri, P. Grifoni y T. Guzzo. 2015. Approaches, tools and applications for sentiment analysis implementation, *International Journal of Computer Applications*, 125(3): 26-33.

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, 96(4): 1027-1040.

Devlin, J., M. Chang, K. Lee y K. Toutanova. 2018. BERT: Pre-training of deep bidirectional transformers for language understanding, ArXiv preprint, arXiv: 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*, 20(62): 123-150.

Doerr, S., L. Gambacorta y J. Serena. 2021. Big data and machine learning in central banking, BIS Working Paper No. 930, Bank for International Settlements.

Garrett, T., R. Hernández-Murillo y M. Owyang. 2004. Does consumer sentiment predict regional consumption?, *Federal Reserve Bank of St. Louis Review*, 87(2): 123-35.

Gentzkow, M., K. Bryan y M. Taddy. 2019. Text as data, *Journal of Economic Literature*, 57(3): 535-574.

González, C. y M. Herman. 2020. Foreign exchange forecasting via machine learning, https://cs229.stanford.edu/proj2018/poster/76.pdf.

Hatzivassiloglou, V. y K. McKeown. 2002. Predicting the semantic orientation of adjectives, en *35th Annual Meeting of the Association for Computational Linguistics*, Madrid, Association for Computational Linguistics.

Jurafsky, D. y J. Martin. 2020. Speech and language processing, tercera edición (borrador), https://web.stanford.edu/ jurafsky/slp3/.

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*, Nueva York, Association for Computational Linguistics.

Kim, S. y E. Hovy. 2004. Determining the sentiment of opinions, en *Proceedings of International Conference on Computational Linguistics*, Ginebra, Coling.

Korab, P. 2021. Use of machine learning in economic research: What the literature tells us, https://towardsdatascience.com/use-of-machine-learning-in-economic-research.

Medhat, W., A. Hassan y H. Korashy. 2014. Sentiment analysis algorithms and applications: A survey, *Ain Shams Engineering Journal*, 5(4): 1093-1113.

Miller, C. y C. Dwork. 2015. Algorithms and bias: Q. and A., New York Times, 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, https://medium.com/@mrtz/how-big-data-is-unfair-9aa544d739de.

Muñiz, V. 2020. Una consulta sobre técnicas de aprendizaje automático, Entrevista presencial, Centro de Investigación en Matemáticas A.C. Monterrey.

Nasukawa, T. y J. Yi. 2003. *Sentiment Analysis: Capturing Favorability using Natural Language Processing*, Nueva York, Association for Computing Machinery.

Pang, B., L. Lee y S. Vaithyanathan. 2002. Thumbs up? Sentiment classification using machine learning techniques, en *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, Nueva York, Association for Computational Linguistics.

Pinto, S. 2019. Sentiment analysis of the fifth district manufacturing and service surveys, *Economic Quarterly*, 105(9): 133-164.

Prabowo, R. y M. Thelwall. 2009. Sentiment analysis: A combined approach, *Journal of Informetrics*, 3(2): 143-157.

Rho, C., R. Fernández y B. Palma. 2021. A sentiment based indicator for the Mexican financial sector, Documento de Investigación No. 2021-04, Banco de México.

Rosenbaum, P. y D. Rubin. 1983. The central role of the propensity score in observational studies for causal effects, *Biometrika*, 70(1): 41-55.

Salhin, A., M. Sherif y E. Jones. 2016. Managerial sentiment, consumer confidence and sector returns, *International Review of Financial Analysis*, 47: 24-38.

Santero, T. y N. Westerlund. 1996. Confidence indicators and their relationship with changes in economic activity, OECD Working Papers No. 170.

Stsiopkina, M. 2022. Hard vs. soft data: The difference, https://oxylabs.io/blog/hard-data-vs-soft-data.

Suss, J. y H. Treitel. 2019. Predicting bank distress in the UK with machine learning, Staff Working Paper No. 831, Bank of England.

Sutton, R. y A. Barto. 2018. *Reinforcement Learning*, Estados Unidos, The MIT Press.

Sweeney, L. 2013. Discrimination in on-line ad delivery, *Queue*, 11(3): 1-20.

Taboada, M., J. Brooke, T. Tofiloski, K. Voll y M. Stede. 2011. Lexicon-based methods for sentiment analysis, *Computational Linguistics*, 37(2): 267-307.

Turney, P. 2002. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews, ArXiv preprint, arXiv:cs/0212032.

Uno, Y. y K. Adachi. 2019. Calculating non-response bias in firms inflation expectations using machine learning techniques, Working Paper Series No.19-E-17, Bank of Japan.

Wallach, H. 2014. Big data, machine learning, and the social sciences: Fairness, accountability, and transparency, 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, https://lilianweng.github.io/posts/2021-12-05-semi-supervised/

Wiesalla, L. 2021. The machine learning workflow. Concepts and applications, https://www.nextlytics.com/blog/machine-learning-workflow.

Witten, I., E. Frank, M. Hall y C. Pall. 2017. *Data Mining: Practical Machine Learning Tools and Techniques*, Elsevier, Morgan Kaufmann Publishers.

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	*Pre-procesar (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 automático, 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.

 https://doi.org/10.24201/ee.v39i2.455

Cuadro A1
(Continación)

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 A5 de este apéndice.
6	*Predicción*	Comparar los datos que genera el modelo, con los datos reservados para la prueba.

Fuente: Elaboración propia 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.

Cuadro A3

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
4) Manufacturas	31-33 Industrias manufactureras	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

Cuadro A3
(Continuación)

Sector (RER-Banxico)	*Sector (SCIAN) 2 dígitos*	*Sub sector (SCIAN) 3 dígitos*
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
8) Transporte y comunicaciones	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 A4
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

Cuadro A4
(Continuación)

Trimestre	*Tópico*	*Norte*	*Centro Norte*	*Centro*	*Sur*
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			
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	

Cuadro A4
(Continuación)

Trimestre	*Tópico*	*Norte*	*Centro Norte*	*Centro*	*Sur*
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			

Cuadro A4
(Continuación)

Trimestre	*Tópico*	*Norte*	*Centro Norte*	*Centro*	*Sur*
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		
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		

Cuadro A4
(Continuación)

Trimestre	*Tópico*	*Norte*	*Centro Norte*	*Centro*	*Sur*
3T2020	1	EUA	Gobierno federal	Inseguridad	Pandemia
	2	COVID 19	COVID 19	Tipo cambio	Reactivación económica
	3	Reactivación económica		Obra pública	Inversión
	4	Inseguridad		Rebrote 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 A5
Criterios para evaluación de algoritmos

Medida	*Definición*		*Observaciones*
Exactitud (Accuracy)	(VP + VN)/TOTAL	VP = número de predicciones positivas correctamente clasificadas VN = número de predicciones negativas correctamente clasificadas TOTAL = número total de casos u observaciones.	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)	FP = número de predicciones positivas calificadas incorrectamente (falsos positivos).	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.

 https://doi.org/10.24201/ee.v39i2.455

Cuadro A5
(Continuación)

Medida	*Definición*		*Observaciones*
Recuperación (Recall)	VP/(VP + FN)	FN = número de predicciones negativas calificadas incorrectamente	Mientras más cercano a 1, indica que más datos de la categoría verdadera fueron bien clasificados.
F1 (F1 Score)	2*(Precisión*Recuperación) Precision+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 una 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. En una tarea de clasificación multiclase es posible medir el F1 general de la asignación de clases, es decir, el F1 micro. No obstante, cuando se trata con un base datos que contiene una cantidad desbalanceada de ejemplos por clase, resulta más conveniente obtener el F1 para cada clase, para posteriormente promediar (sin ponderación) los scores resultantes de cada una de las clases. Este caso se trata del F1 macro. En este trabajo, dada la naturaleza de los datos, se prefiere utilizar el F1 macro.

Fuente: Elaboración propia.

Procedimiento para entrenar las redes: RNN y BERT

Para entrenar las RNN se probaron distintas configuraciones de hiperparámetros con el fin de 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 pre-entrenados 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 pre-entrenado 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 tunning) el modelo pre-entrenado 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 A6 muestra los resultados obtenidos.

Cuadro A6
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

Cuadro A6
(Continuación)

Modelo	*Accuracy*	*F1-Macro*	*F1-Micro*
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 A7
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.

Cuadro A8
Pruebas de raíces unitarias para las series económicas

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	-5.63	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	-5.73	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

Notas: Corresponden a pruebas Dickey-Fuller aumentadas. La columna "Prob" presenta los "valores p" de una cola de McKinnon. Ho: La serie tiene una raíz unitaria.

Fuente: Elaboración propia con información de INEGI.

 https://doi.org/10.24201/ee.v39i2.455

Cuadro A9
Pruebas de raíces unitarias para los índices de sentimiento

	Series	*t-Stat*	*Prob.*	*Especificación:*	*Rechazo Ho:*
	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
Nacional	SVM	-3.778	0.04	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
	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
Norte	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
	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
Centro Norte	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

Cuadro A9
(Continuación)

	Series	*t-Stat*	*Prob.*	*Especificación:*	*Rechazo Ho:*
	Anotado c/n	-2.3622	0.164		No
	Anotado	-2.4567	0.1402		No
	SVM c/n	-2.475	0.136		No
Centro	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
	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
Sur	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

Notas: Corresponden a pruebas Dickey-Fuller aumentadas. La columna "Prob" presenta los "valores p" de una cola de McKinnon. Ho: La serie tiene una raíz unitaria.

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

 https://doi.org/10.24201/ee.v39i2.455

Cuadro A10
Indicadores de opinión nacionales, indicadores nacionales e indicadores regionales

Indicadores de opinión nacionales		
IPM	Indicador de Pedidos Manufactureros	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. Unidad de medida: Puntos. Cifras ajustadas por estacionalidad. Periodicidad: Mensual. Rezago en publicación: No.
ICEMP	Indicador de Confianza Empresarial del Sector Manufacturero	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. Unidad de medida: Puntos. Cifras ajustadas por estacionalidad. Periodicidad: Mensual. Rezago en publicación: No.

Cuadro A10
(Continuación)

Indicadores nacionales		
TCPIB	Crecimiento del PIB Real	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.
	Trim. vs. Trim. previo	Periodicidad: Trimestral.
TCIGAE	Crecimiento del IGAE	Periodicidad: Mensual.
	Mes vs. Mes previo	Rezago en publicación: 1 mes.
TCIMAI	Crecimiento Trimestral del Indicador Mensual de Actividad Industrial Total	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

 https://doi.org/10.24201/ee.v39i2.455

Cuadro A10
(Continuación)

Indicadores nacionales		
TCIMAI (cont.)		(SCIAN), alcanzando una representatividad del 97% del valor agregado bruto del año 2013, año base de productos del SCNM. Unidad de medida: Índice de volumen físico 2013=100. Cifras ajustadas por estacionalidad. Periodicidad: Mensual. Periodicidad: Mensual. Periodicidad: Mensual. Rezago en publicación: 52 días hábiles de concluido el trimestre de referencia.
TCIMAIMAN	Crecimiento Trimestral del Indicador de Actividad Industria Manufacturera	Componente del IMAI. Unidad de medida: Índice de volumen físico 2013=100. Cifras ajustadas por estacionalidad. Periodicidad: Mensual Rezago en publicación: 52 días hábiles de concluido el trimestre de referencia.
Indicadores regionales		
TCITAER	Crecimiento Trimestral del Indicador de Actividad Económica Regional	Banco de México en base a las cifras ajustadas por estacionalidad del Indicador Trimestral de la Actividad Económica Estatal del INEGI.

Cuadro A10
(Continuación)

Indicadores regionales		
TCITAER (cont.)		Unidad de medida: Índice IT-2013=100 Regiones: Norte, Centro Norte, Centro y Sur. Cifras ajustadas por estacionalidad. Periodicidad: Trimestral. Rezago en publicación: 4 meses.
TCITAERMAN	Crecimiento Trimestral del Indicador Regional de Actividad Manufacturera	Regiones: Norte, Centro Norte, Centro y Sur. Cifras ajustadas por estacionalidad. 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.

Leonardo E. Torre, Eva E. González, Luis R. Casillas, Jorge A. Alvarado

Índices de sentimiento regionales y su asociación con indicadores oportunos dé actividad económica en México, 2016-2021
Regional sentiment indexes and their association with timely indicators of economic activity in Mexico, 2016-2021

Estudios Económicos (México, D.F.)
vol. 39, núm. 2, p. 349 - 419, 2024
El Colegio de México A.C.,
ISSN: 0188-6916

DOI: https://doi.org/10.24201/ee.v39i2.455