



Munich Personal RePEc Archive

Entrepreneurship and remote working: Analysis with X Data

Gutierrez-Lythgoe, Antonio

15 December 2023

Online at <https://mpra.ub.uni-muenchen.de/119490/>
MPRA Paper No. 119490, posted 16 Dec 2023 16:44 UTC

Emprendimiento y teletrabajo: Análisis con datos de X

Gutiérrez-Lythgoe, Antonio

15 de diciembre de 2023

Resumen

Este estudio aborda la relación entre la identificación como emprendedor y las opiniones sobre el teletrabajo, utilizando datos de la plataforma X, anteriormente Twitter. Los resultados indican correlaciones significativas entre ser emprendedor y una menor probabilidad de expresar opiniones negativas sobre el teletrabajo. Además, se documenta una asociación positiva entre esta identificación y métricas de actividad en la plataforma, como seguidores, seguidos, retuits y favoritos. En términos de connotaciones, ser emprendedor muestra correlaciones significativas y positivas con la probabilidad de ser clasificado como positivo. La interacción entre esta condición y el género masculino presenta una correlación negativa y significativa en la probabilidad de ser clasificado como positivo, indicando posible variación según el género. Estos hallazgos evidencian diferencias en las actitudes hacia el teletrabajo en distintos contextos laborales, proporcionando resultados relevantes para la comprensión de las perspectivas en el entorno de trabajo y la toma de decisiones en el ámbito empresarial.

Abstract

This study explores the relationship between the identification as an entrepreneur and opinions on working from home, utilizing data from platform X, formerly Twitter. The results indicate significant correlations between being an entrepreneur and a lower probability of expressing negative opinions on working from home. Furthermore, a positive association is documented between this identification and activity metrics on the platform, such as followers, followings, retweets, and favorites. In terms of sentiment analysis, being an entrepreneur shows significant and positive correlations with the probability of being classified as positive. The interaction between this condition and male gender exhibits a negative and significant correlation in the probability of being classified as positive, indicating potential variation based on gender. These findings highlight differences in attitudes toward working from home in various work contexts, offering relevant insights into perspectives in the workplace and decision-making in the business environment.

Keywords: Entrepreneurship, Working from home, Social Media, Statistics

JEL Classification: L26, J24

Introducción

Desde el inicio de la pandemia del COVID-19 (Andrade et al., 2022), el teletrabajo se ha integrado en la rutina laboral con la intención de contribuir al equilibrio entre trabajo y familia (Giménez et al., 2012; Molina, 2021; Daneshfar et al., 2022). Investigaciones recientes sugieren que este cambio en el paradigma laboral no es efímero, sino que se ha convertido en una práctica habitual (Barrero et al., 2021). El teletrabajo, en la nueva normalidad, tiene el potencial de transformar diversos aspectos de nuestras rutinas, incluyendo la gestión del tiempo laboral (Gimenez-Nadal et al., 2020), la movilidad diaria (Mouratidis et al., 2021; Mouratidis & Papagiannakis, 2021; Ozbilen et al., 2021) y el bienestar y la salud (Gimenez-Nadal et al., 2018, 2023; Charalampous et al., 2019). No obstante, también plantea desafíos significativos para las empresas, generando incertidumbre en relación con el rendimiento de los trabajadores (Golden et al., 2008; Bloom et al., 2015; Golden & Gajendran, 2019; Karl et al., 2022). En este contexto, resulta crucial explorar la opinión pública sobre el teletrabajo en la era post-pandémica (Saura et al., 2022), especialmente desde la perspectiva de los emprendedores, quienes desempeñan un papel fundamental en la innovación, el crecimiento económico y la generación de empleo, impactando directamente en la economía (Acs & Armington, 2006; Campaña et al., 2020; Gimenez-Nadal et al., 2022; Molina et al., 2016; Velilla et al., 2018).

Con el propósito de explorar la percepción de emprendedores sobre el teletrabajo, utilizamos datos de redes sociales, adoptando una perspectiva novedosa en ciencias sociales mediante la aplicación de Big Data (Einav & Levin, 2014). A diferencia de investigaciones anteriores centradas en la pandemia, nos enfocamos en un marco temporal posterior, analizando tuits en español durante octubre y noviembre de 2022. Un aspecto relevante es que examinamos tuits en español, ya que las investigaciones previas se han centrado principalmente en la comunidad angloparlante. Para llevar a cabo este análisis, aplicamos un modelo multilingüe preentrenado, XLM-T, basado en roBERTa (Barbieri et al., 2021), que proporciona probabilidades de connotaciones positivas, negativas o neutras en cada tuit. Este enfoque específico aporta nuevas perspectivas sobre la opinión de emprendedores respecto al teletrabajo en el contexto actual.

En la muestra general, la identificación como emprendedor correlaciona negativa y significativamente con la probabilidad de que los tuits sobre teletrabajo sean percibidos de manera negativa. Esta identificación se asocia positivamente con indicadores de actividad en la plataforma, sugiriendo mayor visibilidad. En la muestra restringida, ser emprendedor se vincula con una mayor propensión a seguir a otros usuarios. El análisis revela diferencias significativas en la probabilidad de tuits negativos entre emprendedores y no emprendedores, respaldando la conexión entre ser emprendedor y menor probabilidad de expresar opiniones negativas sobre el teletrabajo. Estos hallazgos, respaldados estadísticamente, indican una sutil variación en la identificación como emprendedor en las percepciones y expresiones en Twitter sobre el teletrabajo.

Este trabajo realiza distintas contribuciones en tres dimensiones diferentes. En primer lugar, contribuimos a la integración de Big Data en la investigación de ciencias sociales, alineándose con la evolución de las tecnologías de información y comunicación (Einav & Levin, 2014). En segundo término, aportamos al campo de las ciencias sociales mediante la aplicación de técnicas avanzadas de inteligencia artificial para el procesamiento de lenguaje natural en tuits en español (Lansley & Longley, 2016; Dubey & Tripathi, 2020; Osorio-Arjona et al., 2021; Zhang et al., 2021; Daneshfar et al., 2022; Karl et al., 2022; Saura et al., 2022). En tercer lugar, contribuimos al ámbito de investigación sobre teletrabajo, el cual ha experimentado un renovado interés en la era post-pandemia, siendo un terreno relativamente menos explorado en periodos anteriores (Olson et al., 1998). En cuarto lugar, contribuimos al análisis del emprendimiento. La estructura del artículo se organiza de la siguiente manera. La Sección 2 presenta los datos empleados y el proceso de limpieza de estos. La Sección 3, detalla la metodología estadística empleada, la Sección 4 detalla los principales resultados obtenidos y en último lugar, la Sección 6 presenta la conclusión del Trabajo.

Datos

Los datos empleados en este análisis provienen de la red social X, anteriormente conocida como Twitter. Para el estudio del teletrabajo recogemos todos los tuits publicados entre el 30 de septiembre de 2022 hasta el 1 de diciembre de 2022. Es decir, establecemos un marco temporal de dos meses, con el objetivo de obtener una muestra homogénea (veasé, Osorio-Arjona et al., 2021). Los tuits analizados contienen al menos una de las tres palabras siguientes: teletrabajar, teletrabajo o teletrabajando. Se seleccionan estas palabras porque son empleadas con frecuencia en castellano para referirse a la práctica del teletrabajo.

Seleccionados los tuits, aplicamos filtros con el objetivo de obtener una muestra de tuits que hayan sido publicados por individuos. En otras palabras, eliminamos aquellos tuits sospechosos de ser publicados por cuentas bot o cuentas correspondientes a instituciones o empresas. En primer lugar, eliminamos tuits duplicados con la finalidad de mitigar la presencia de cuentas bot (Stieglitz et al., 2017). En segundo lugar, eliminamos aquellos tuits que contengan enlaces URL (Saura et al., 2022). En tercer lugar, eliminamos los tuits duplicados.

Para el análisis textual mediante técnicas de lenguaje natural procesamos los tuits publicados mediante las siguientes acciones:

1. Eliminamos caracteres raros
2. Eliminamos palabras con un solo carácter
3. Sustituimos los espacios en blanco
4. Convertimos el texto en minúsculas
5. Eliminamos palabras vacías¹ (stopwords) y dígitos
6. Eliminamos nombres de usuario que aparezcan en el tuit

Tras la aplicación de los filtros, obtenemos una muestra total con 10288 tuits². Depurada la muestra, procedemos a la implementación un modelo de lenguaje multilingüe XLM-T de Barbieri et al. (2021)³. Este modelo, es un modelo que pertenece a los modelos roBERTa (Liu et al., 2019). Estos modelos mejoran los enfoques utilizados previamente en el procesamiento de lenguaje natural, conocidos como BERT (Devlin et al., 2018). El modelo empleado está específicamente entrenado para analizar tuits, y permite el procesamiento de texto en 8 lenguajes distintos, entre ellos el español. Concretamente, este modelo ha sido confeccionado y afinado a través de una muestra de 198 millones de tuits en distintos idiomas. Una vez aplicado el modelo obtendremos los resultados en valores reales. Aplicando la función Softmax convertimos los resultados en probabilidades. De esta manera, obtendremos la probabilidad de 0 a 1 de que el tuit analizado tenga una probabilidad positiva, negativa o neutra.

Además del contenido de los tuits, disponemos de información adicional relacionada con cada uno de ellos como: descripción del usuario, nombre de usuario, nombre visible, localización de usuario, número de retuits del tuit, verificación de la cuenta, fecha del tuit, número de seguidores del usuario y número de perfiles que sigue el usuario. Hacemos uso de estos campos para inferir algunas características sociodemográficas básicas a partir de la información disponible.

Para identificar aquellos individuos emprendedores recurrimos a la información contenida en la descripción de su perfil. Establecemos una serie de palabras claves relacionadas con la actividad

¹ Preposiciones, conjunciones, artículos, pronombres y otros.

² Cada tuit, corresponde a un usuario distinto. Por tanto, disponemos de información de 10288 usuarios distintos.

³ Este modelo está disponible para su descarga en: <https://huggingface.co/cardiffnlp/twitter-xlm-roberta-base-sentiment>

empresaria, emprendimiento, start-up, business owner, fundador, fundadora, co-fundador, co-fundadora, empresario. En el caso de encontrar coincidencias en la descripción de los usuarios con alguna de estas palabras, categorizamos al usuario como emprendedor. Utilizamos una variable binaria que representa con valor 1 si emprendedor o emprendedora, y 0 en otro caso. Identificamos así a un total de 121 emprendedores en nuestra muestra original. Es decir, un 1,17% de los usuarios de la muestra.

Para categorizar el género de los usuarios comparamos el nombre visible en X de cada usuario con los nombres con frecuencia igual o mayor a 20 personas en España. Obtenemos así una lista con 55762 nombres diferenciada por géneros, encontramos dicha información a través del Instituto Nacional de Estadística⁴ (INE). En el caso en el que el nombre disponible coincida con alguno presente en la lista, asignamos el género correspondiente. Utilizamos una variable binaria que representa con valor 1 si el género es masculino, y 0 femenino. Obtenemos coincidencias en los nombres para 5173 usuarios, el 50% de la muestra original. Esto es debido a que un elevado número de usuarios no presentan su nombre real como nombre visible en X.

Por último, identificamos si el usuario reside en España o en otra parte del mundo. Recordemos que nuestra muestra contiene tuits en español, por lo que los usuarios son hispanohablantes y pueden residir en distintas localizaciones de España o Latinoamérica entre otras. Para ello, depuramos de nuestra muestra original aquellos individuos que no indican su localización en su perfil, nos quedamos con una muestra de 3485 usuarios. Establecemos una serie de palabras clave para buscar coincidencias en este campo y determinar su lugar de residencia: España, Madrid, Barcelona, Valencia, Sevilla, Zaragoza, Málaga, Murcia, Palma, Bilbao, Alicante, Toledo, Valladolid, Córdoba, Vigo, Gijón, Hospitalet, Vitoria, Coruña, Granada, Elche, Oviedo, Badalona, Cartagena. Si se obtiene coincidencia, indicamos en una variable binaria con valor 1 indicando residencia en España y valor 0, en caso contrario.

En la Tabla 1, mostramos los estadísticos descriptivos para la muestra restringida y no restringida. En la muestra sin restricciones, la media de seguidores es de 2,166.590, mientras que en la muestra restringida esta media es ligeramente mayor, alcanzando 2,560.184. La cantidad de cuentas seguidas, retuits y favoritos exhiben niveles similares en ambas muestras. Respecto a las probabilidades de que un tuit sea negativo, neutral o positivo, se observa que varían entre las dos muestras. Respecto a la proporción de emprendedores identificados, esta cantidad es relativamente en ambas muestras, representando el 1,2% y el 1,8% respectivamente. En cuanto a las variables categóricas, en la muestra restringida, alrededor del 68.7% de los usuarios se identifican como hombres, e identificamos al el 32.9% de la muestra restringida como residentes en España. Como observamos, existe una mayor proporción de usuarios varones este aspecto está presente en la literatura empírica con datos de X/Twitter (Mislove et al., 2011).

Metodología

El objetivo principal del trabajo es analizar la opinión de los emprendedores sobre la práctica del teletrabajo. Para ello, con los datos existentes, podemos hacer uso de distintas metodologías estadísticas. Primeramente, realizamos un análisis de correlación en ambas muestras, restringida y no restringida, para analizar la existencia de posibles relaciones de correlación estadísticamente significativas.

⁴ Enlace:

https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177009&menu=ultiDatos&idp=1254734710990

En segundo lugar, planteamos un análisis de homogeneidad para observar si la distribución de tuits positivos, negativos y neutros, se distribuye por igual entre emprendedores y no emprendedores. Para ello, recurriremos al contraste de Pearson de Chi-Cuadrado sobre los tuits categorizados en función de la probabilidad más alta que muestren. Es decir, categorizamos un tuit como positivo si su probabilidad de ser positivo es mayor a la probabilidad de ser neutro o negativo. En segundo lugar, realizamos una comparación entre dos grupos de nuestra muestra, emprendedores y no emprendedores. Es decir, comparamos el comportamiento de variables continuas (las probabilidades de que un tuit sea positivo, negativo o neutro) en distintos grupos. Para ello recurrimos al análisis estadístico descriptivo e inferencial.

Por otro lado, estudiaremos las distribuciones de cada probabilidad entre emprendedores y no emprendedores. Para ello recurrimos a contrastes de normalidad de Jarque & Bera (1987) y Kolmogorov-Smirnov (Kolmogorov, 1933; Smirnov, 1936) con la corrección de Lilliefors (1967). En función del resultado obtenido del contraste de normalidad, optamos por contrastes paramétricos, robustos o no paramétricos para contrastar si existen diferencias estadísticas entre ambos grupos. En caso de asumir que la distribución en ambos grupos se asimila a una distribución normal, optamos por el contraste de T de Student (Student, 1908) o el contraste de Welch (1951) en caso de heterocedasticidad en la muestra. En caso de rechazar la hipótesis de normalidad, aplicamos un contraste no paramétrico de Mann-Whitney (Wilcoxon, 1992). Como podemos apreciar en el histograma y Q-Q plot de los emprendedores y en los contrastes de normalidad mencionados, rechazamos la hipótesis nula de normalidad. Por tanto, aplicamos el contraste no paramétrico.

Este contraste de rangos asignados de Wilcoxon, también conocido como la prueba U de Mann-Whitney permite comparar dos muestras independientes y determinar si hay evidencia de que una de las muestras tiene valores significativamente mayores o menores que la otra. Definimos el estadístico de la prueba como:

$$Z = \frac{V - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \sim N(0,1),$$

dónde $V = \min\{R_+, R_-\}$ siendo R_+ los rangos de las diferencias entre ambos grupos que son positivas y R_- de las diferencias negativas. La hipótesis nula contrastada es que la mediana de la diferencia entre ambos grupos es igual a 0.

Por último, planteamos un modelo de regresión lineal para la tres probabilidades de que un tuit sea negativo, neutro o positivo. Estimamos el modelo lineal mediante Mínimos Cuadrados Ordinarios:

$$\text{Probabilidad} = \alpha + \beta_1 X + \beta_2 \text{Emprendedor} + \beta_3 \text{Hombre} + \beta_4 \text{España} + \beta_5 \text{Género} * \text{Emprendedor} + \beta_6 \text{España} * \text{Emprendedor} + \varepsilon,$$

En la ecuación propuesta, la variable dependiente es la probabilidad, y las variables independientes incluyen un conjunto de predictores, representados por X, como el número de seguidores, seguidos, favoritos y retuits, junto con variables categóricas, como si el usuario se identifica como emprendedor (Emprendedor), su género (Hombre), y si reside en España (España). Los coeficientes $\alpha, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6$, representan respectivamente la constante, la relación lineal entre las variables predictivas y la probabilidad, y las correlaciones de las variables categóricas y sus interacciones.

Resultados

En el análisis de correlación entre variables de la Tabla 3, que corresponde a la muestra sin restricciones, se destaca que la variable "ser emprendedor" muestra una correlación negativa y estadísticamente significativa con la probabilidad de que un tuit sobre teletrabajo tenga una connotación negativa. Este hallazgo sugiere que la identificación como emprendedor está asociada a una menor probabilidad de que los tuits sobre teletrabajo sean percibidos de manera negativa. Además, notamos que esta variable presenta correlaciones positivas y significativas con el número de seguidores, seguidos, retuits y favoritos, indicando una posible relación entre la identificación como emprendedor y una mayor actividad y visibilidad en la plataforma.

En la Tabla 4, que corresponde a la muestra restringida, observamos únicamente una correlación positiva y estadísticamente significativa entre la variable "ser emprendedor" y el número de usuarios que sigue el emprendedor. Este resultado sugiere que, en la muestra restringida, la identificación como emprendedor está asociada a una mayor propensión a seguir a otros usuarios, lo cual podría indicar una dinámica específica en el comportamiento de esta categoría de usuarios en el contexto del teletrabajo. Estos hallazgos resaltan la importancia de considerar las restricciones de la muestra al interpretar las correlaciones y proporcionan una visión más detallada de las relaciones entre las variables analizadas en el estudio.

En la Figura 1, se observa que la mayor proporción de tuits de emprendedores presenta una connotación neutra (41.32%), seguida por tuits con connotación negativa (32.23%) y positiva (26.45%). En contraste, los usuarios no emprendedores exhiben una distribución diferente, con una mayor proporción de tuits negativos (44.10%) en comparación con tuits neutros (34.96%) y positivos (20.94%). Sin distinguir entre emprendedores o no, la mayor proporción de tuits presenta una connotación negativa (43.96%), seguida por la connotación neutra (35.01%) y la connotación positiva (21.01%). Para corroborar estadísticamente estas observaciones, se realizó un análisis de homogeneidad mediante el contraste de Pearson de Chi-Cuadrado, obteniendo un p-valor inferior a 0.05, lo que indica una diferencia significativa en las distribuciones de tuits entre emprendedores y no emprendedores.

$$P(\chi_2^2 > 6.9428) = 0.03107$$

Como podemos comprobar previamente el valor obtenido para el estadístico de contraste es de 6,9428. En otras palabras, rechazamos la hipótesis nula que establece que las distribuciones de los tuits son iguales para ambos grupos emprendedores y no emprendedores. En la Tabla 2, se presentan los residuos estandarizados calculados para evaluar las diferencias entre las frecuencias observadas y esperadas.

$$Residuo_{ij} = \frac{n_{ij} - \frac{n_i \cdot n_j}{N}}{\sqrt{\frac{n_i \cdot n_j}{N}}}$$

Dónde el término $\frac{n_i \cdot n_j}{N}$ es la frecuencia conjunta esperada y n_{ij} es la frecuencia conjunta esperada. Observamos un valor absoluto mayor a 1.5 para los tuits con connotación negativa, indicando que la frecuencia esperada es mayor a la observada. Esto sugiere que las distribuciones de los tuits por connotación son estadísticamente diferentes entre emprendedores y no emprendedores, con una menor frecuencia observada para tuits con connotación negativa sobre el teletrabajo.

La estadística descriptiva revela una diferencia en la probabilidad de que un tuit sobre teletrabajo emitido por un emprendedor sea distinto al de otros individuos. Específicamente, en el gráfico de medias de la Figura 2, se observa una mayor variabilidad en las probabilidades de que un tuit sea negativo para los emprendedores, sugiriendo que la probabilidad de que un emprendedor opine

negativamente sobre el trabajo es menor que la de otros individuos. En los gráficos restantes (Figura 4 y Figura 6), se observa un patrón similar de mayor variabilidad en las probabilidades de que el tuit sea positivo o neutro, aunque no parece existir una diferencia significativa.

Mediante el análisis inferencial utilizando el contraste no paramétrico de Mann-Whitney para las probabilidades mencionadas, se rechaza la hipótesis nula en el caso de la probabilidad negativa, indicando diferencias significativas en la probabilidad de que los tuits sobre teletrabajo contengan connotaciones negativas entre emprendedores y no emprendedores (ver Tabla 6). Específicamente, el intervalo de confianza para la diferencia en la mediana señala que la probabilidad de un tuit negativo en no emprendedores es estadísticamente mayor que en emprendedores, con una estimación de la diferencia de la mediana de 0.03 y un intervalo de confianza al 95% que abarca valores entre 0.004 y 0.075. En cuanto a las probabilidades de que el tuit sea positivo o neutral, los resultados no evidencian validez estadística contra la hipótesis nula, sugiriendo que la diferencia no es estadísticamente distinta de 0.

La Tabla 7 presenta los resultados del modelo de regresión lineal que explora las relaciones entre las variables predictoras y las probabilidades de clasificación de tuits sobre teletrabajo como negativos, neutros o positivos. En el análisis de las variables cuantitativas, se observa que el número de retuits tiene una correlación estadísticamente significativa y positiva con la probabilidad de que un tuit sea clasificado como negativo y una correlación negativa significativa con la probabilidad de que sea clasificado como positivo. Además, el número de favoritos muestra una correlación positiva y estadísticamente significativa con la probabilidad de que un tuit sea clasificado como neutro. No se encuentra una asociación estadísticamente significativa entre el número de seguidores y seguidos y las clasificaciones de sentimiento.

En relación con las variables categóricas, destaca que la variable "Emprendedor" presenta correlaciones estadísticamente significativas en una de las tres clasificaciones de sentimiento. Ser emprendedor se correlaciona positivamente con la probabilidad de que sea clasificado como positivo. La variable "Hombre" también muestra correlaciones significativas, siendo positiva en la probabilidad de que un tuit sea clasificado como neutro. La interacción entre "Emprendedor" y "Hombre" presenta una correlación negativa y estadísticamente significativa en la probabilidad de que un tuit sea clasificado como positivo, indicando que la relación entre emprendimiento y connotación podría variar según el género. No encontramos evidencia estadística de una relación entre la localización del usuario y su interacción con la variable "Emprendedor" y las tres probabilidades.

En resumen, estos resultados sugieren que existen correlaciones estadísticamente significativas entre variables cuantitativas y categóricas, como retuits, favoritos, ser emprendedor y género, y las probabilidades de que un tuit sobre teletrabajo tenga una connotación negativa, neutra o positiva. No obstante, es esencial reconocer las limitaciones asociadas con el bajo nivel de R^2 , indicando que las variables incluidas en el modelo explican solo una pequeña proporción de la variabilidad observada en las clasificaciones de sentimiento.

Conclusión

En este Trabajo hemos analizado las opiniones compartidas en español por los usuarios de la red social X, previamente Twitter, sobre el teletrabajo. Concretamente, hemos centrado nuestra atención sobre aquellos individuos que identificamos como emprendedores. Para ello hemos utilizado técnicas de Inteligencia Artificial para procesar las connotaciones de cada tuit en la muestra. Para profundizar en el análisis por variables sociodemográficas identificamos el género del usuario mediante la búsqueda de coincidencias entre el nombre del perfil de usuario y los nombres frecuentes en España según el Instituto Nacional De Estadística. Además, mediante coincidencias en las palabras clave de la descripción identificamos la localización del usuario.

Para el análisis estadístico planteamos varios análisis incluyendo: análisis de correlación, análisis de homogeneidad entre grupos, contraste no paramétrico entre grupos y regresión lineal. En la muestra sin restricciones, la identificación como emprendedor muestra una correlación negativa y significativa con la probabilidad de que los tuits sobre teletrabajo sean percibidos de manera negativa. Además, se asocia positivamente con el número de seguidores, seguidos, retuits y favoritos, indicando mayor actividad y visibilidad. En la muestra restringida, ser emprendedor está relacionado con una mayor propensión a seguir a otros usuarios. El análisis inferencial confirma diferencias significativas en la probabilidad de tuits negativos entre emprendedores y no emprendedores, respaldando la asociación entre ser emprendedor y menor probabilidad de expresar opiniones negativas sobre el teletrabajo. Estos hallazgos, respaldados por pruebas estadísticas, resaltan una ligera diferencia de la identificación como emprendedor en las percepciones y expresiones en Twitter sobre el teletrabajo. Sin embargo, se reconoce la limitación asociada con el bajo nivel de R^2 , sugiriendo cautela en la generalización de los resultados y destacando la necesidad de investigaciones futuras que aborden estas limitaciones para una comprensión más completa.

Es esencial abordar las limitaciones inherentes al estudio. La variabilidad en la presentación de datos por parte de los usuarios, la existencia potencial de perfiles privados, cuentas duplicadas o bots, y la falta de precisión en la información demográfica introducen sesgos y errores en la identificación de usuarios. Al interpretar los resultados, se debe considerar cuidadosamente estas limitaciones.

Este análisis de las percepciones sobre el teletrabajo entre usuarios identificados como emprendedores en la red social X revela una asociación estadísticamente significativa entre la identificación como emprendedor y una menor probabilidad de expresar opiniones negativas sobre el teletrabajo. Este hallazgo sugiere que aquellos que desempeñan roles clave en la generación de empleo tienden a tener una perspectiva menos crítica hacia esta modalidad laboral. La comprensión de esta dinámica adquiere especial relevancia en el contexto de la formulación de políticas públicas, ya que ofrece información valiosa sobre cómo los líderes empresariales visualizan el teletrabajo.

Referencias

- Acs, Z. J., & Armington, C. (2006). *Entrepreneurship, geography, and American economic growth*. Cambridge University Press.
- Andrade, C., Gillen, M., Molina, J.A. and Wilmarth, M.J. (2022). The social and economic impact of Covid-19 on family functioning and well-being: Where do we go from here? *Journal of Family and Economic Issues*, 43, 205-212.
- Barbieri, F., Anke, L. E., & Camacho-Collados, J. (2021). Xlm-t: Multilingual language models in twitter for sentiment analysis and beyond. *arXiv preprint arXiv:2104.12250*.
- Barrero, J. M., Bloom, N., & Davis, S. J. (2021). Why working from home will stick. *National Bureau of Economic Research*.
- Bloom, N., Liang, J., Roberts, J., & Ying, Z. J. (2015). Does working from home work? Evidence from a chinese experiment. *Quarterly Journal of Economics*, 130(1), 165–218. <https://doi.org/10.1093/qje/qju032>
- Campaña, J.C., Giménez-Nadal, J.I. and Molina, J.A. (2020). Self-employed and employed mothers in Latin American families: are there differences in paid-work, unpaid work and child care? *Journal of Family and Economic Issues*, 41, 52-69.
- Charalampous, M., Grant, C. A., Tramontano, C., & Michailidis, E. (2019). Systematically reviewing remote e-workers' well-being at work: a multidimensional approach. *European Journal of Work and Organizational Psychology*, 28(1), 51–73. <https://doi.org/10.1080/1359432X.2018.1541886>
- Daneshfar, Z., Asokan-Ajitha, A., Sharma, P., & Malik, A. (2022). Work-from-home (WFH) during COVID-19 pandemic – A netnographic investigation using Twitter data. *Information Technology and People*. <https://doi.org/10.1108/ITP-01-2021-0020>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *ArXiv Preprint ArXiv:1810.04805*
- Dubey, A. D., & Tripathi, S. (2020). Analysing the sentiments towards work-from-home experience during COVID-19 pandemic. In *Journal of Innovation Management* (Vol. 8, Issue 1, pp. 13–19). Universidade do Porto - Faculdade de Engenharia. https://doi.org/10.24840/2183-0606_008.001_0003
- Einav, L., & Levin, J. (2014). Economics in the age of big data. *Science*, 346(6210). <https://doi.org/10.1126/science.1243089>
- Giménez, J.I., Molina, J.A. and Ortega, R. (2012). Self-employed mothers and the work-family conflict. *Applied Economics*, 44, 2133-2148.
- Giménez-Nadal, J.I., Molina, J.A. and Velilla, J. (2020). Work time and well-being for workers at home: evidence from the American Time Use Survey. *International Journal of Manpower*, 41(2), 184-206.
- Giménez-Nadal, J.I., Molina, J.A. and Velilla, J. (2022). Intergenerational correlation of self-employment in Western Europe. *Economic Modelling*, 108, 105741.
- Giménez-Nadal, J.I., Molina, J.A. and Velilla, J. (2023). Should we cheer together? Gender differences in instantaneous well-being: An application to COVID-19 lockdowns. *Journal of Happiness Studies*, 24, 529-562.

- Golden, T. D., & Gajendran, R. S. (2019). Unpacking the Role of a Telecommuter's Job in Their Performance: Examining Job Complexity, Problem Solving, Interdependence, and Social Support. *Journal of Business and Psychology*, 34(1), 55–69. <https://doi.org/10.1007/s10869-018-9530-4>
- Golden, T. D., Veiga, J. F., & Dino, R. N. (2008). The Impact of Professional Isolation on Teleworker Job Performance and Turnover Intentions: Does Time Spent Teleworking, Interacting Face-to-Face, or Having Access to Communication-Enhancing Technology Matter? *Journal of Applied Psychology*, 93(6), 1412–1421. <https://doi.org/10.1037/a0012722>
- Jarque, C. M., & Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review/Revue Internationale de Statistique*, 163-172.
- Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why We Twitter: Understanding Microblogging Usage and Communities. <http://www.pownce.com>
- Javadian, R. (2007). Social work responses to earthquake disasters: A social work intervention in Bam, Iran. *International Social Work*, 50(3), 334–346.
- Karl, K. A., Peluchette, J. v., & Aghakhani, N. (2022). Virtual Work Meetings During the COVID-19 Pandemic: The Good, Bad, and Ugly. *Small Group Research*, 53(3), 343– 365. <https://doi.org/10.1177/10464964211015286>
- Kolmogorov, A. N. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giorn Dell'inst Ital Degli Att*, 4, 89-91.
- Lansley, G., & Longley, P. A. (2016). The geography of Twitter topics in London. *Computers, Environment and Urban Systems*, 58, 85–96. <https://doi.org/10.1016/j.compenvurbsys.2016.04.002>
- Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American statistical Association*, 62(318), 399-402.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. <http://arxiv.org/abs/1907.11692>
- Min, H., Peng, Y., Shoss, M., & Yang, B. (2021). Using machine learning to investigate the public's emotional responses to work from home during the COVID-19 pandemic. *Journal of Applied Psychology*, 106(2), 214.
- Mislove, A., Lehmann, S., Ahn, Y. Y., Onnela, J. P., & Rosenquist, J. (2011). Understanding the demographics of Twitter users. In *Proceedings of the international AAAI conference on web and social media* (Vol. 5, No. 1, pp. 554-557).
- Molina, J.A., Velilla, J. and Ortega, R. (2016). The decision to become an entrepreneur in Spain: the role of household finances. *International Journal of Entrepreneurship*, 20(1), 57-73.
- Mouratidis, K., & Papagiannakis, A. (2021). COVID-19, internet, and mobility: The rise of telework, telehealth, e-learning, and e-shopping. *Sustainable Cities and Society*, 74. <https://doi.org/10.1016/j.scs.2021.103182>
- Mouratidis, K., Peters, S., & van Wee, B. (2021). Transportation technologies, sharing economy, and teleactivities: Implications for built environment and travel. *Transportation Research Part D: Transport and Environment*, 92. <https://doi.org/10.1016/j.trd.2021.102716>

Olson, J. S., Covi, L., Rocco, E., Miller, W. J., & Allie, P. (1998). A Room of Your Own: What would it take to help remote groups work as well as collocated groups? CHI 98 Conference Summary on Human Factors in Computing Systems, 279–280.

Osorio-Arjona, J., Horak, J., Svoboda, R., & García-Ruíz, Y. (2021). Social media semantic perceptions on Madrid Metro system: Using Twitter data to link complaints to space. *Sustainable Cities and Society*, 64. <https://doi.org/10.1016/j.scs.2020.102530>

Ozbilen, B., Wang, K., & Akar, G. (2021). Revisiting the impacts of virtual mobility on travel behavior: An exploration of daily travel time expenditures. *Transportation Research Part A: Policy and Practice*, 145, 49–62. <https://doi.org/10.1016/j.tra.2021.01.002>

Saura, J. R., Ribeiro-Soriano, D., & Zegarra Saldaña, P. (2022). Exploring the challenges of remote work on Twitter users' sentiments: From digital technology development to a post-pandemic era. *Journal of Business Research*, 142, 242–254. <https://doi.org/10.1016/j.jbusres.2021.12.052>

Schwartz, H., Eichstaedt, J., Kern, M., Dziurzynski, L., Lucas, R., Agrawal, M., Park, G., Lakshmikanth, S., Jha, S., & Seligman, M. (2013). Characterizing geographic variation in well-being using tweets. *Proceedings of the International AAAI Conference on Web and Social Media*, 7(1), 583–591.

Smirnov, N.V. (1936). Sui la distribution de w_2 (Criterium de M.R.v. Mises), *Comptes Rendus (Paris)*, 202, pp. 449–452 (6.1).

Stieglitz, S., Brachten, F., Ross, B., & Jung, A.-K. (2017). Do social bots dream of electric sheep? A categorisation of social media bot accounts. *ArXiv Preprint ArXiv:1710.04044*.

Student. (1908). The probable error of a mean. *Biometrika*, 6(1), 1-25.

Velilla, J., Molina, J.A. and Ortega, R. (2018). Why older workers become entrepreneurs? International evidence using fuzzy set methods. *The Journal of the Economics of Ageing*, 12, 88-95. <https://doi.org/10.1016/j.jeoa.2018.03.004>.

Welch, B. L. (1951). On the comparison of several mean values: an alternative approach. *Biometrika*, 38(3/4), 330-336.

Wilcoxon, F. (1992). Individual comparisons by ranking methods. In *Breakthroughs in Statistics: Methodology and Distribution* (pp. 196-202). New York, NY: Springer New York.

Tabla 1. Estadística descriptiva para las dos muestras.

VARIABLES	Muestra sin restringir		Muestra Restringida	
	Media	D.T.	Media	D.T.
Seguidores	2,166.590	15,098.586	2,560.184	16,551.090
Seguidos	902.090	1,596.166	1,060.068	1,826.141
Retuits	0.647	16.447	0.637	10.065
Favoritos	5.158	104.004	5.327	47.459
Prob. tuit negativo	0.415	0.344	0.404	0.344
Prob. tuit neutral	0.356	0.243	0.352	0.241
Prob. tuit positivo	0.230	0.274	0.245	0.286
Emprendedor/a	0.012	0.108	0.018	0.134
Hombre	-	-	0.687	0.464
Residente en España	-	-	0.329	0.470

Nota: Estadística descriptiva para la muestra restringida y no restringida. La muestra restringida está compuesta por aquellos tuits que permiten identificar la localización y el género debido a las coincidencias textuales. La muestra sin restringir está compuesta por 10288 tuits y la no restringida por 3478 tuits.

Tabla 2. Correlaciones entre variables en la muestra sin restringir

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) Prob. negativo	1.000							
(2) Prob. neutral	-0.612*	1.000						
(3) Prob. positivo	-0.713*	-0.118*	1.000					
(4) Seguidores	-0.026*	0.021*	0.014	1.000				
(5) Seguidos	-0.049*	0.051*	0.016	0.192*	1.000			
(6) Retuits	0.001	0.007	-0.008	0.142*	0.064*	1.000		
(7) Favoritos	0.003	0.003	-0.006	0.138*	0.048*	0.815*	1.000	
(8) Emprendedor/a	-0.023*	0.015	0.015	0.028*	0.056*	0.081*	0.068*	1.000

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Nota: Correlaciones entre variables para la muestra de 10.288 tuits procesados en castellano sobre teletrabajo.

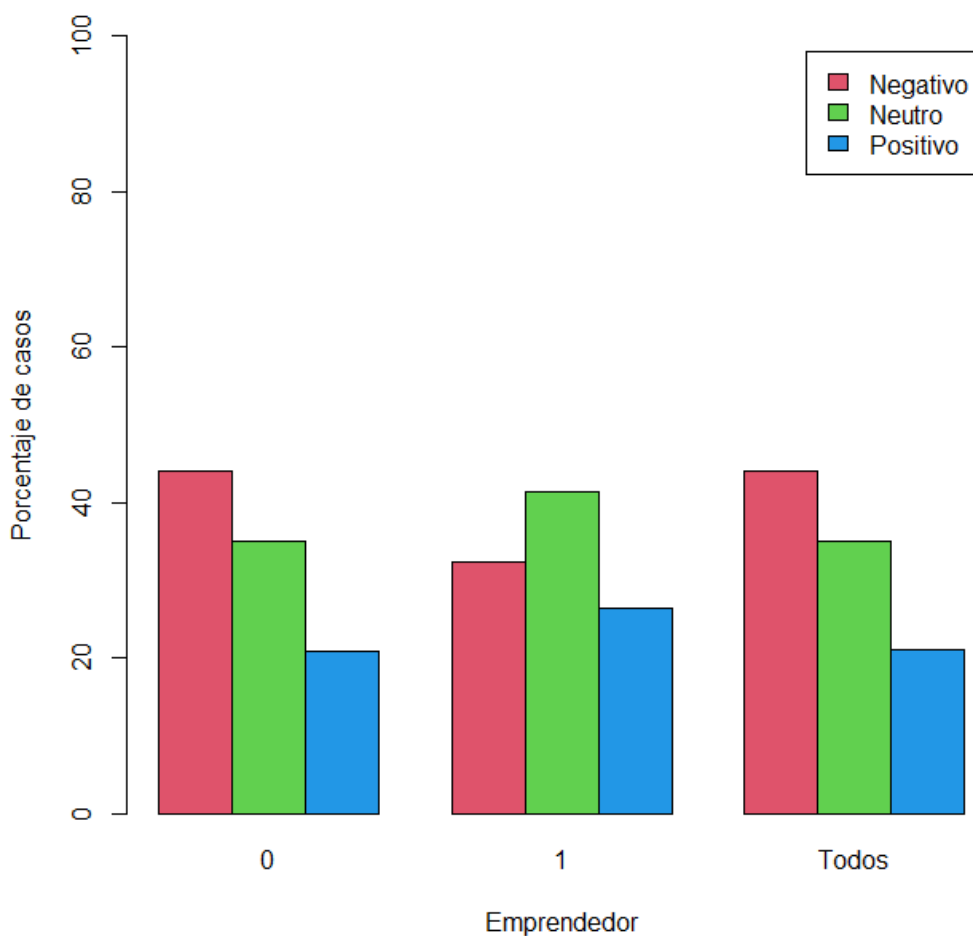
Tabla 3. Correlaciones entre variables en la muestra restringida

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Prob. negativo	1.000									
(2) Prob. neutral	-0.569*	1.000								
(3) Prob. positivo	-0.722*	-0.159*	1.000							
(4) Seguidores	0.004	0.016	-0.018	1.000						
(5) Seguidos	-0.053*	0.076*	0.000	0.261*	1.000					
(6) Retuits	0.026	-0.006	-0.027	0.169*	0.045*	1.000				
(7) Favoritos	0.028	-0.008	-0.028	0.305*	0.041*	0.798*	1.000			
(8) Emprendedor/a	-0.014	0.003	0.014	0.019	0.053*	0.003	0.020	1.000		
(9) Hombre	-0.022	0.055*	-0.020	-0.006	0.021	0.014	0.010	0.028	1.000	
(10) Res. España	-0.016	0.019	0.004	-0.006	0.009	0.006	-0.010	0.013	0.035*	1.000

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Nota: Correlaciones entre variables para la muestra de 3.478 tuits procesados en castellano sobre teletrabajo en los que podemos identificar lugar de residencia y género

Figura 1. Distribución de tuits según su connotación y si es Emprendedor/a



Nota: Distribución porcentual de los tuits por connotación entre emprendedores y no emprendedores sobre la muestra total de 10.288 tuits

Tabla 4. Residuos obtenidos a través de las frecuencias esperadas y observadas

Tuit por connotación	No emprendedor	Emprendedor
Negativo	0,21	-1,95
Neutro	-0,13	1,17
Positivo	-0,14	1,31

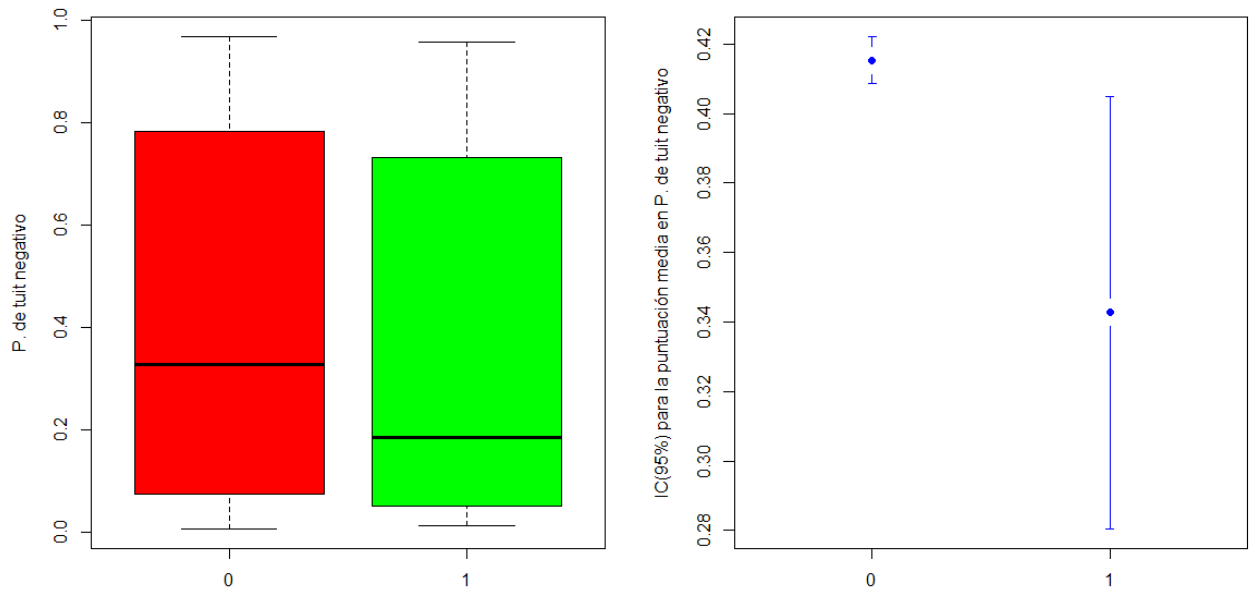
Nota: Los valores pertenecen a los residuos obtenidos mediante la diferencia de las frecuencias esperadas y observadas dividido por la raíz cuadrada de la frecuencia esperada sobre la muestra total de 10288 tuits

Tabla 5. Estadística descriptiva de las probabilidades por connotación en emprendedores

	P. negativa	P. neutra	P. positiva
Media	0,342	0,39	0,267
Mediana	0,186	0,385	0,149
Varianza	0,119	0,064	0,079
Asimetría	0,707	0,2247	0,947
Curtosis	-1,194	-1,221	-0,348

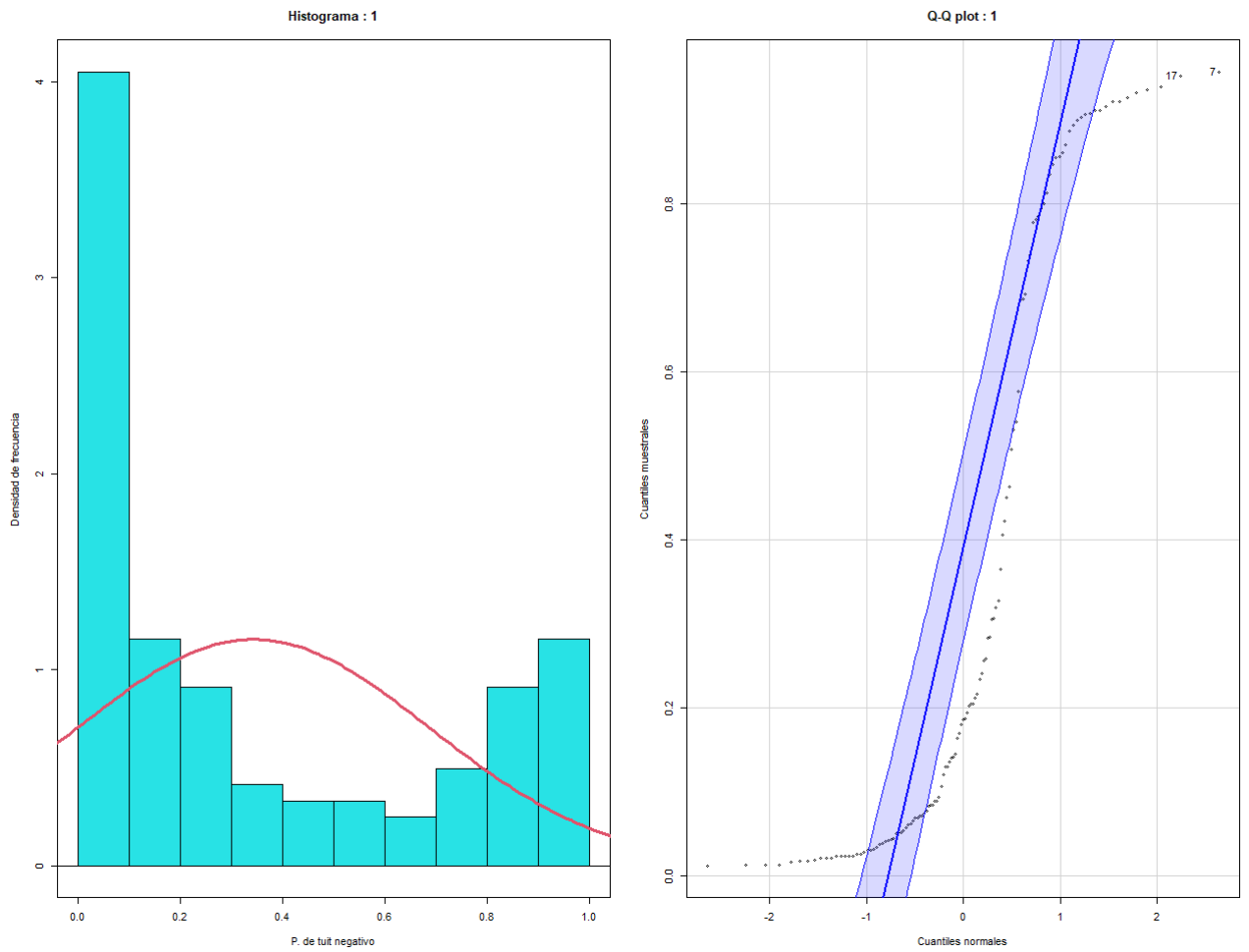
Nota: Estos valores corresponden con los estadísticos descriptivos para la probabilidad de cada connotación para los usuarios que son emprendedores.

Figura 2. Boxplot y Gráfico de medias para la probabilidad de connotación negativa.



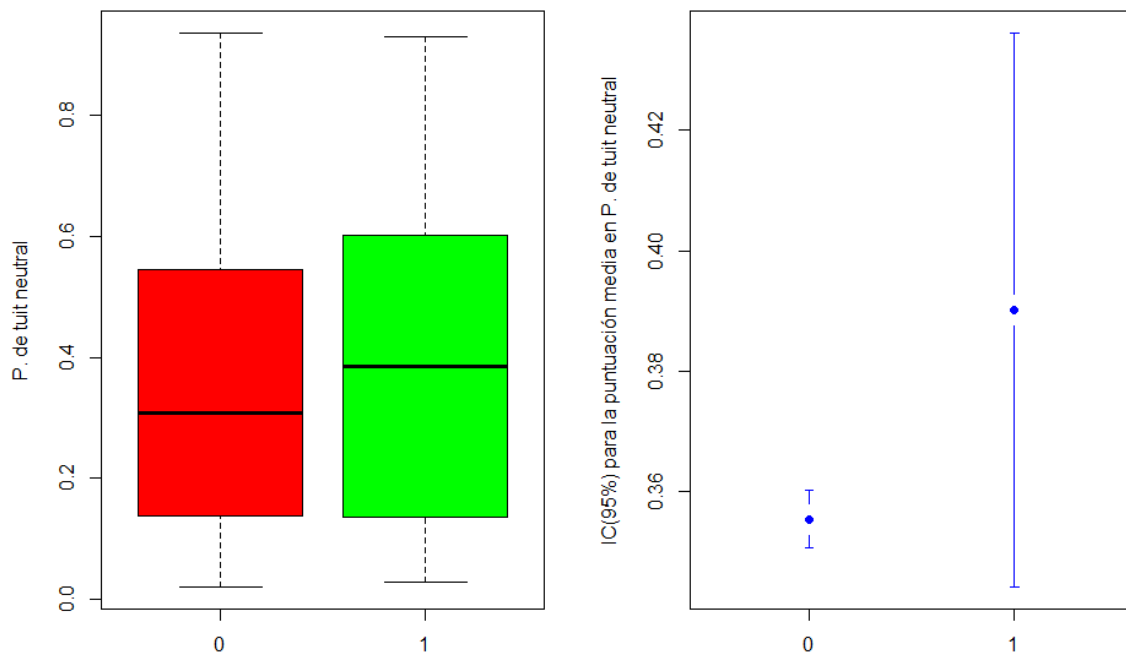
Nota: Diagrama de caja y Gráfica de media para analizar las diferencias entre la probabilidad de que un tuit sea negativo en función si el usuario es emprendedor (valor 1) o en caso contrario, valor 0.

Figura 3. Histograma y Q-Q plot para la probabilidad de connotación negativa de un tuit y emprendedor



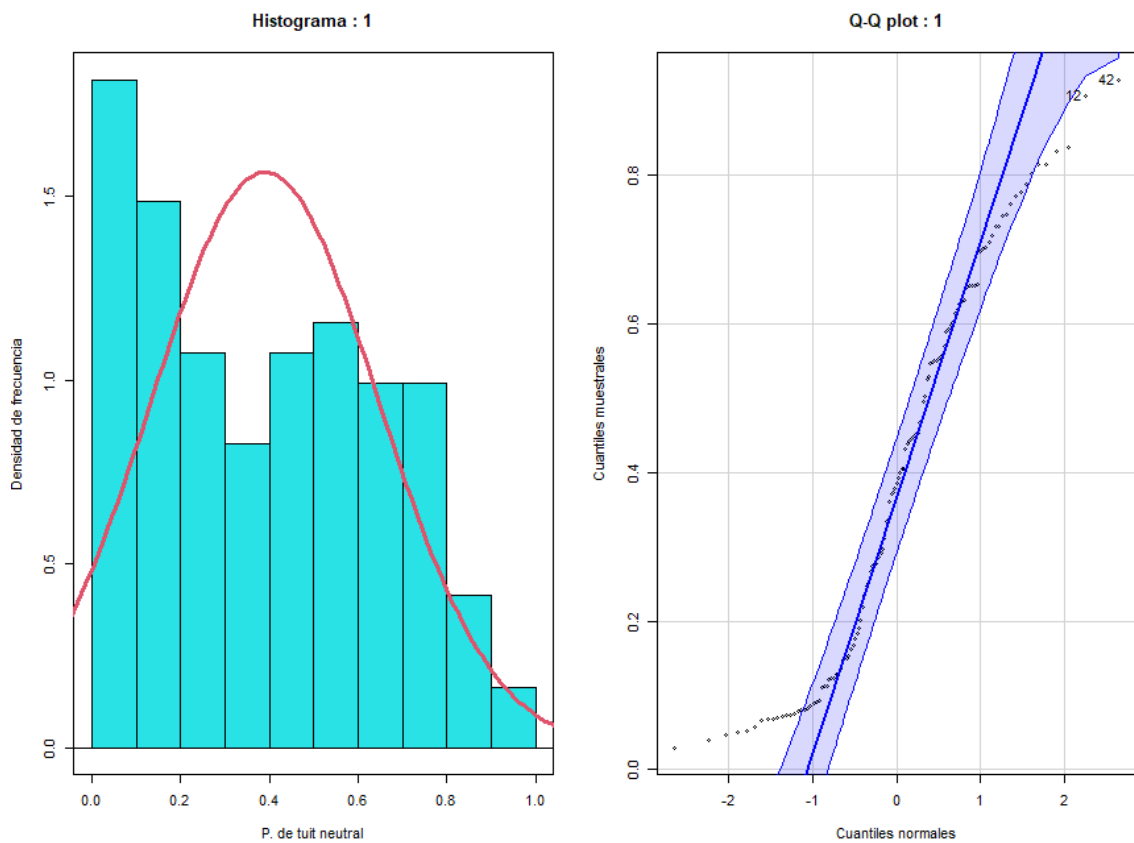
Nota: Histograma y Q-Q plot para el análisis de normalidad de la distribución de la probabilidad de que un tuit sea negativo si el usuario es emprendedor.

Figura 4. Boxplot y Gráfico de medias para la probabilidad de connotación neutra.



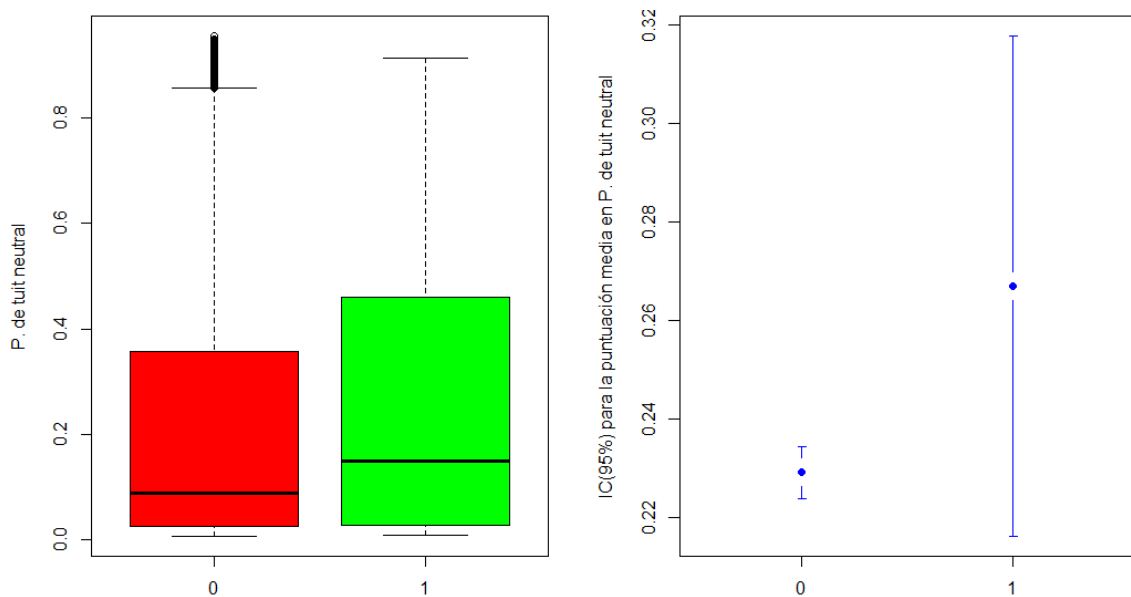
Nota: Diagrama de caja y Gráfica de media para analizar las diferencias entre la probabilidad de que un tuit sea negativo en función si el usuario es emprendedor (valor 1) o en caso contrario, valor 0.

Figura 5. Histograma y Q-Q plot para la probabilidad de connotación neutra de un tuit y emprendedor



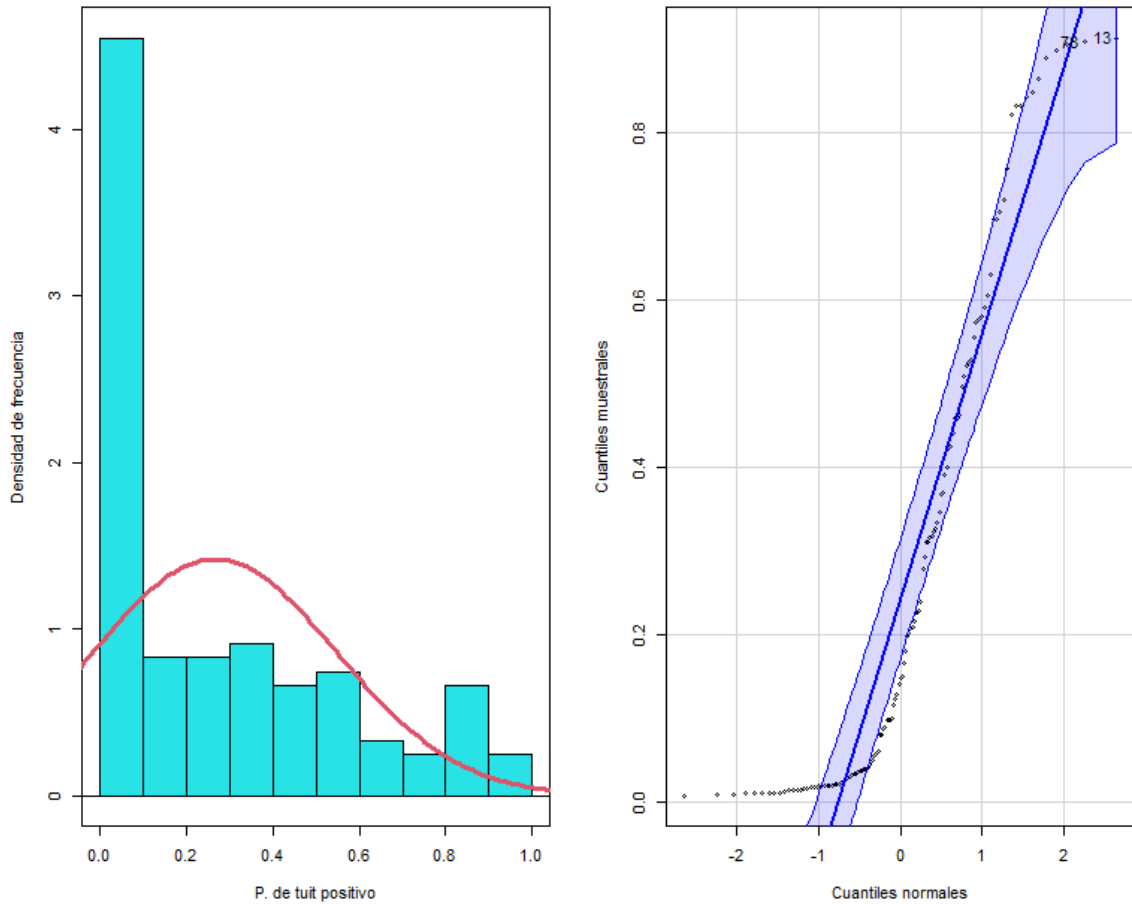
Nota: Histograma y Q-Q plot para el análisis de normalidad de la distribución de la probabilidad de que un tuit sea neutro si el usuario es emprendedor.

Figura 6. Boxplot y Gráfico de medias para la probabilidad de connotación positiva.



Nota: Diagrama de caja y Gráfico de media para analizar las diferencias entre la probabilidad de que un tuit sea positivo en función si el usuario es emprendedor (valor 1) o en caso contrario, valor 0.

Figura 7. Histograma y Q-Q plot para la probabilidad de connotación positiva de un tuit y emprendedor



Nota: Histograma y Q-Q plot para el análisis de normalidad de la distribución de la probabilidad de que un tuit sea neutro si el usuario es emprendedor.

Tabla 6. Estadística descriptiva de las probabilidades por connotación en emprendedores

	P. negativa	P. neutra	P. positiva
Wilcoxon W	689188	570966	567474
Diferencia estimada	0.03**	-0,03	-0,01
IC 95%	(0.004, 0.075)	(-0.071, 0.012)	(-0.031, 0.003)

* $p < 0.1$, ** $p < 0.05$. *** $p < 0.01$.

Nota: Contraste de Mann-Whitney para el análisis en la diferencia de la distribución de la probabilidad de cada connotación entre emprendedores y no emprendedores. Presentamos el calculo del estadístico, la diferencia estimada en las medianas de la distribución y el intervalo de confianza al 95% para dicha diferencia estimada.

Tabla 7. Modelo de regresión lineal

	<i>Variable dependiente:</i>		
	Negativo (1)	Neutro (2)	Positivo (3)
Retuit	0.0005 (0.001)	-0.0001 (0.001)	-0.0004 (0.001)
Favoritos	0.0001 (0.0002)	-0.00004 (0.0001)	-0.0001 (0.0002)
Seguidores	0.00000 (0.00000)	0.000 (0.00000)	-0.00000 (0.00000)
Seguidos	-0.00001*** (0.00000)	0.00001*** (0.00000)	0.00000 (0.00000)
Emprendedor	-0.118 (0.100)	-0.069 (0.070)	0.187** (0.083)
Hombre	-0.016 (0.013)	0.027*** (0.009)	-0.011 (0.011)
Residente en España	-0.011 (0.013)	0.008 (0.009)	0.003 (0.010)
Emprendedor * Hombre	0.101 (0.105)	0.070 (0.073)	-0.171* (0.087)
Emprendedor*España	0.030 (0.090)	0.028 (0.063)	-0.058 (0.075)
Constante	0.428*** (0.012)	0.321*** (0.008)	0.251*** (0.010)
Observaciones	3,478	3,478	3,478
R ²	0.005	0.009	0.003
R ² Ajustado	0.002	0.007	0.0003
Error. Est. Resid. (df = 3468)	0.343	0.240	0.286
Estadístico F (df = 9; 3468)	1.888**	3.623***	1.112

Nota:

* ** *** p < 0.01