

Twitter y Mujeres STEM. Análisis con R.

XX, XX,

Palabras clave: Mujeres STEM, Sentimientos y Emociones en Twitter, NLP con R.

Abstract

Diversos estudios afirman que el problema de la brecha de género sigue vigente en todo el mundo, especialmente en las profesiones STEM. Particularmente en América Latina, este problema se agrava debido a los prejuicios o las normas culturales que condicionan el hacer femenino^{1 2}. Un enfoque desde los casos de éxito recopilados en un corpus creado a partir de los tres libros de la serie “Matilda y las Mujeres en Ingeniería” fue analizado mediante técnicas de Minería de Textos y Procesamiento de Lenguaje Natural (NLP)³, demostrando una casi nula aparición de ciertos términos como “violencia” o “diversidad” y detectando un posible sesgo de representatividad a largo plazo. Surge entonces la necesidad de ampliar el dominio y el corpus de estudio a un campo más igualitario y menos binario como la red social Twitter, incorporando además un análisis de sentimientos y emociones, mediante técnicas NLP codificadas en R. Los resultados son un diagnóstico alentador de la situación a Marzo de 2023. Las bases de datos creadas servirán también para nuevos análisis enfocados a los usuarios que se expresan sobre Mujeres-STEM en esta red social.

Metodología y Resultados

La recopilación de mensajería sobre la red social Twitter que conformó el corpus de estudio se realizó desde el 29 de septiembre de 2022 hasta el 22 de febrero de 2023, en días seleccionados (dado que la nueva administración de la plataforma dispuso medidas restrictivas en cantidad y frecuencia para la minería de datos sobre esta red). La obtención de las claves de acceso sobre la API de Twitter se realizó bajo la categoría “esencial”.

El paquete de R principalmente utilizado entre las fechas mencionadas para la recolección de los datos de Twitter fue “rtweet”. Algunos de los hashtags o expresiones de búsqueda útiles para construir la base de datos sobre la cual se trabajó fueron: “mujereseningenieria”, “mujeresSTEM”, “femSTEM” y “mujeresIT”, entre otros. El dataframe generado por el barrido de la red, contiene: información del texto del mensaje, fecha, hora, autor, y metadata asociada al usuario que postea su mensaje.

El corpus correspondiente a los textos de los tweets fue tokenizado por palabras con la remoción de números, signos y stopwords; realizándose luego un análisis de sentimientos y emociones con el paquete “syuzhet” (diccionario NRC, español). Los resultados recogidos en diferentes fechas demostraron que *los sentimientos positivos cuatuplican los sentimientos negativos expresados en los mensajes*.

Mediante la función “get_nrc_sentiment” las 21875 palabras fueron puntuadas binariamente según 8 emociones (enfado, anticipación, disgusto, miedo, alegría, tristeza, sorpresa, confianza) (Figura 1) y 2 sentimientos: positivo y negativo (Figura 2). Sólo las palabras “relevantes” (puntuadas con 1 por la función) son graficadas.

Por último, se construyó una nube de palabras en cuatro cuadrantes, según las emociones de interés para la temática: tristeza, alegría, enfado y confianza (Figura 3), observándose el impacto de repitencia de

¹ ComunidadMujer (2016). *Informe GET. Género Educación y Trabajo: La brecha persistente*. Ed. ComunidadMujer. Chile.

² INDEC (2023). *Dossier estadístico en conmemoración del 112° Día Internacional de la Mujer*. Argentina, Marzo 2023.

³ Pascal G., et. al. (2021). *Aplicación de Procesamiento de Lenguaje Natural en Ensayos en Primera Persona: Exploración de libros “Matilda y las Mujeres en Ingeniería en América Latina”*. Revista RIIYM – Vol VI – N° 11.

las palabras (por tamaño) y la intencionalidad de su aparición (por su ubicación respecto a las dimensiones definidas en cuadrantes).

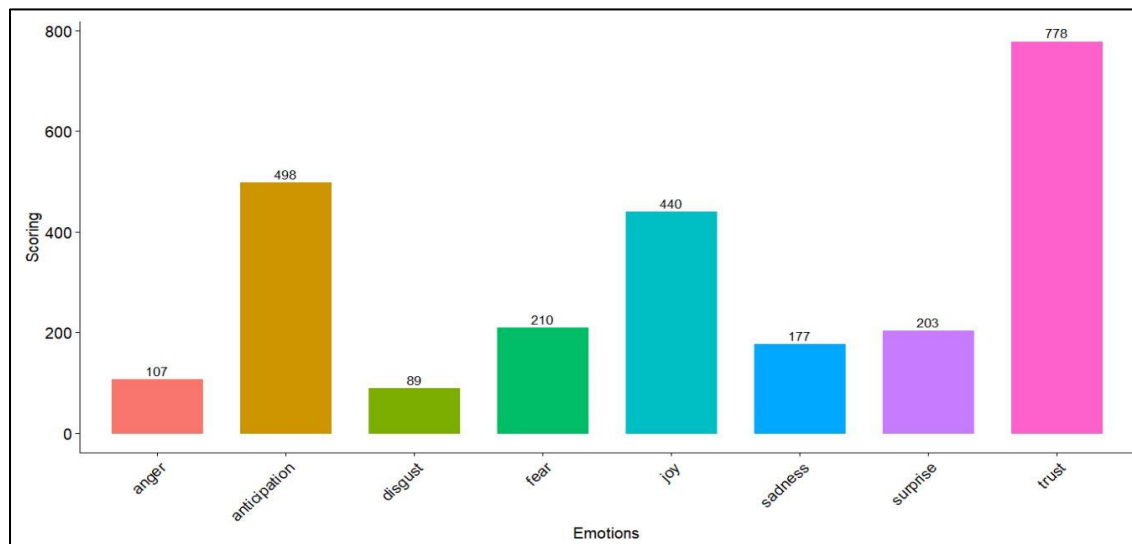


Figura 1 – Cantidad de emociones detectadas en el corpus de los mensajes Mujeres-STEM de Twitter.

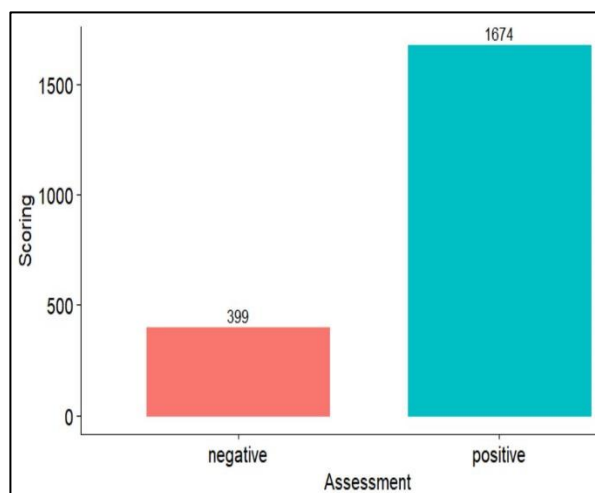


Figura 2 – Sentimientos detectados en los Tweets.



Figura 3 – Palabras por repitencia e intención.

Consideraciones finales

Con la base de datos construida se extenderá el trabajo realizado hacia nuevos niveles de análisis orientados a los usuarios y sus perfiles: vínculos (seguidores y seguidos), ranking de réplica y ubicación geográfica, entre otros.

La actual administración de la red social Twitter a través de la puesta en marcha de su reciente versión 2 (V2) requiere de nuevos perfiles de autenticación, aún difíciles de conseguir a bajo costo. Los paquetes recomendados a futuro para la manipulación de la información de esta red con el lenguaje R *bajo la nueva versión 2* son: “academictwitterR” y “RTwitterV2”⁴.

XX

XX

[XX@XX](#)

⁴ Twitter V2- Tools: <https://developer.twitter.com/en/docs/twitter-api/tools-and-libraries/v2>