

Índice de contenidos:

Introducción.....	1
Desarrollo.....	2
Conclusión.....	7
Referencias bibliográficas:	8

¿Qué es lo que hace sobresalir al comediante Franco Escamilla?

Un informe de proyecto sobre el análisis de sentimiento aplicado a los monólogos de *stand up* mexicano

Introducción

Uno de los componentes universales de la cultura es el humor y, aunque no existe unicidad con respecto al cómo se demuestra, es evidente que juega un papel esencial dentro de las interacciones humanas. Sabemos de la experiencia que el humor se presenta en casi cualquier situación social y, por tanto, tiene también una influencia emocional y cognitiva en nosotros. Así pues, el hecho de crear humor para otros es una característica distintiva del ser humano que responde a nuestras necesidades sociales (Bateson, 2005). De ahí que resulte relevante el estudio de los monólogos que emplean algunos de los comediantes mexicanos con mayor presencia en el medio.

El estilo de comedia en el que se centra este proyecto es el de *stand up*, un anglicismo que describe un formato de comedia en el que el humorista se encuentra *de pie* y se dirige al público mediante un monólogo donde presenta sus reflexiones y valoraciones de manera cómica o irónica. Por consiguiente, el principal reto para realizar un estudio objetivo sobre este tipo de monólogos es el carácter subjetivo del humor, pues queda determinado por diversos elementos culturales y etarios (Bengochea, 2016). Para mantener la objetividad, se utilizó el análisis de sentimiento, ya que es una técnica de procesamiento de lenguaje natural (PLN) que permite analizar las opiniones, valoraciones y actitudes de las personas. De este modo, se puede evaluar el sentimiento general de una oración, un párrafo o del discurso entero (Henríquez et al., 2017; Liu, 2012).

La hipótesis sobre la que se basa el proyecto es que, después de aplicar un análisis de sentimiento a los monólogos, obtendremos como resultado un conjunto de ponderaciones específicas y objetivas para fundamentar nuestra respuesta acerca de lo que hace destacar al comediante Franco Escamilla con respecto de otros similares. Cabe mencionar que el criterio de selección para los comediantes que servirán de comparación fue fundamentalmente arbitrario, aunque guiado por su notoriedad en las plataformas digitales. Por otra parte, los

monólogos fueron escogidos de YouTube en función de la fecha de publicación y extensión total en minutos, de manera que fueran muestras representativas y confiables en la medida que la calidad o cantidad de las transcripciones lo permitiera (véase *Tabla 1*).

Tabla 1: Comediantes y monólogos analizados

Comediante	Nombre del monólogo	Enlace
Álex Fernández	No creo en los noticieros	https://www.youtube.com/watch?v=GqQLuNkuGSI
Alan Saldaña	Una infancia pobre	https://www.youtube.com/watch?v=5rsEBCbIJwg
Carlos Ballarta	Los miedos existenciales	https://www.youtube.com/watch?v=xpGUU0BZhRw
Franco Escamilla	El D.F. y la vida familiar	https://www.youtube.com/watch?v=qo7G516Hjak
Ricardo O’Farrill	Estoy enfermo de gripa	https://www.youtube.com/watch?v=JSdeZMbINpY
Sofía Niño de Rivera	Bebé feo y mamás paranoicas	https://www.youtube.com/watch?v=ZkVj_boxQ8U

Desarrollo

Para el desarrollo del proyecto, se consideraron las siguientes etapas, a saber:

1. Recopilación de los datos.
2. Limpieza de los datos.
3. Análisis exploratorio de datos.
4. Análisis de sentimiento.

El lenguaje de programación que se usó es Python puesto que, hasta el momento, sigue siendo uno de los lenguajes más empleados para trabajar en el campo de la inteligencia artificial. Para abordar el análisis propuesto, Python ofrece librerías que constituyen el estado del arte sobre las técnicas tradicionales de PLN (López-Briega, 2017). Dentro de ellas podemos

destacar NLTK y TextBlob, ambas implementadas en el proyecto para realizar el análisis de sentimiento de los monólogos.

En la primera etapa, se recopilaron los datos a partir de los cuales se desarrolla el análisis que, en este caso, corresponden al conjunto de monólogos disponibles en YouTube enlistados en la *Tabla 1*. Como se trata de un material audiovisual y las librerías de Python trabajan con texto, se tomaron las transcripciones generadas de manera automática. Esto implica una menor calidad de los datos, pues la transcripción no es perfecta y se corre el riesgo de que se modifiquen palabras que, a su vez, alteran el sentido de frases enteras.

Por otra parte, los algoritmos de PLN están mayormente diseñados para el idioma inglés, lo que generó la necesidad de traducir una copia de las transcripciones a este idioma mediante el traductor de Google. Esto, con la intención de mantener los dos conjuntos de datos, tanto la transcripción, como la traducción de los monólogos, puesto que serán usados para el análisis exploratorio y el análisis de sentimiento, respectivamente. Ahora bien, en aras de obtener los resultados más exactos posibles, se hizo una corrección manual para ambos conjuntos. De esta manera, se preservó el sentido de casi todas las frases y se rescataron palabras relevantes que, en otro caso, se hubieran perdido.

En la segunda etapa del proyecto se realizó una limpieza de los datos para poder alimentar los algoritmos de PLN. Al final de esta etapa se obtuvieron como resultado dos matrices documento-término, donde las filas corresponden a cada monólogo del corpus y las columnas corresponden a las palabras. El preprocesamiento del texto para crear dicha matriz consistió en usar expresiones regulares para:

- a) convertir todo el texto en minúsculas,
- b) eliminar puntuación, valores numéricos y palabras vacías (es decir, palabras irrelevantes como artículos y preposiciones),
- c) quitar texto sin sentido (como caracteres especiales), y
- d) *tokenizar* el texto (segmentarlo en palabras).

Posteriormente, como parte de la tercera etapa, se realizó un análisis exploratorio en búsqueda de los patrones más simples. En primera instancia, se generó una nube de palabras con los términos más comunes para cada comediante (véase *Imagen 1*). Al observar la

frecuencia de las palabras, se hacen evidentes los conceptos, personajes y entidades sobre los que se desarrolla el monólogo; como en el caso de Álex Fernández, donde se aprecian claramente los dos temas principales de su rutina: las noticias y la música de banda. Asimismo, se advirtió en este punto el uso de ciertas muletillas, siendo la más notable “damas y caballeros” de Carlos Ballarta. Finalmente, con respecto a Franco Escamilla, no hay mucho que se pueda destacar, pues sencillamente se señalan los elementos sobre los que se forman sus chistes.

Imagen 1: Nubes de palabras



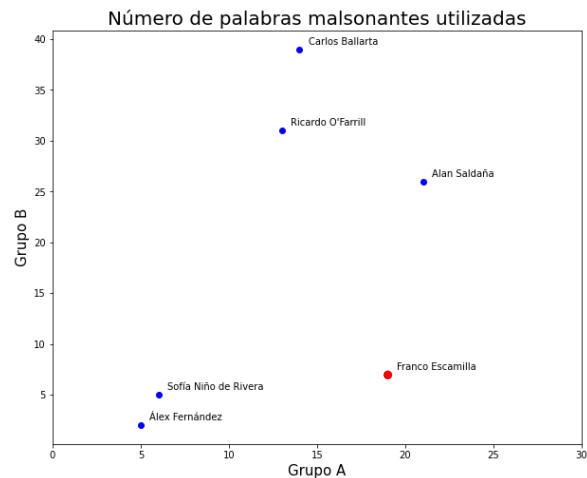
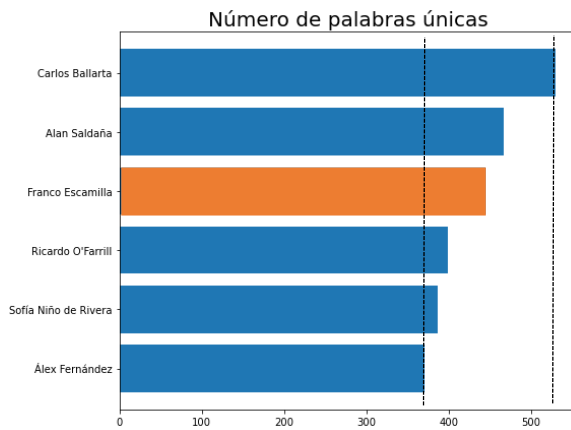
A continuación, se optó por estudiar el tamaño del vocabulario que tiene cada comediante. Para completar este análisis, se calculó el número de palabras únicas para cada uno, es decir, aquellos términos que no se repitieron en la rutina de algún otro comediante (véase *Gráfica 1*). De los resultados obtenidos, es interesante notar que Franco Escamilla se localiza en un punto intermedio, lo que sugiere un

buen balance en la elección del vocabulario o en la secuencia y ritmo de su monólogo. Sin embargo, sería necesaria una revisión más detenida para hacer una generalización al respecto.

El estudio subsiguiente consistió en cuantificar las palabras malsonantes que los comediantes incluyeron en su discurso (véase *Diagrama 1*). Dada la variedad de insultos y eufemismos que existen, se agruparon las palabras soeces en dos grupos: las que comienzan con la letra ‘p’, ‘c’ y ‘m’, así como sus derivadas, se asignaron al *Grupo A*; mientras que el resto se contempló en el *Grupo B*. Con este último análisis exploratorio, se pudo llegar a la conclusión de que Franco no abusa de las palabras malsonantes, como podría decirse de Carlos Ballarta o de Alan Saldaña. Pero, por otra parte, tampoco tiene un humor tan limpio, como lo es el de Sofía Niño de Rivera o Álex Fernández.

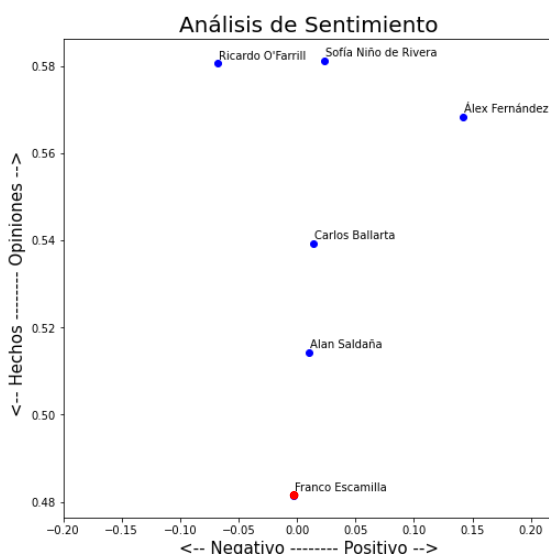
Gráfica 1:

Diagrama 1:



Hasta ahora, el análisis efectuado ha sido bastante genérico. Estos recuentos y diagramas de dispersión también podrían aplicarse a datos numéricos. De ahí que surja, entonces, la necesidad de poder realizar análisis automáticos de criterios subjetivos presentes en un texto. Para ello, el análisis de sentimiento es la técnica más adecuada, pues nos permite descubrir los patrones ocultos en un corpus gracias a sus bases en el aprendizaje automático (Hernández & Gómez, 2014). Con esta intención, la cuarta y última etapa del proyecto implementó el análisis de sentimiento a los monólogos traducidos, que se encontraban ya a modo de matriz documento-término, para evaluar dos características principales (véase *Diagrama 2*):

Diagrama 2:



a) Polaridad: qué tan positivas o negativas eran las palabras, siendo -1 muy negativa y +1 muy positiva.

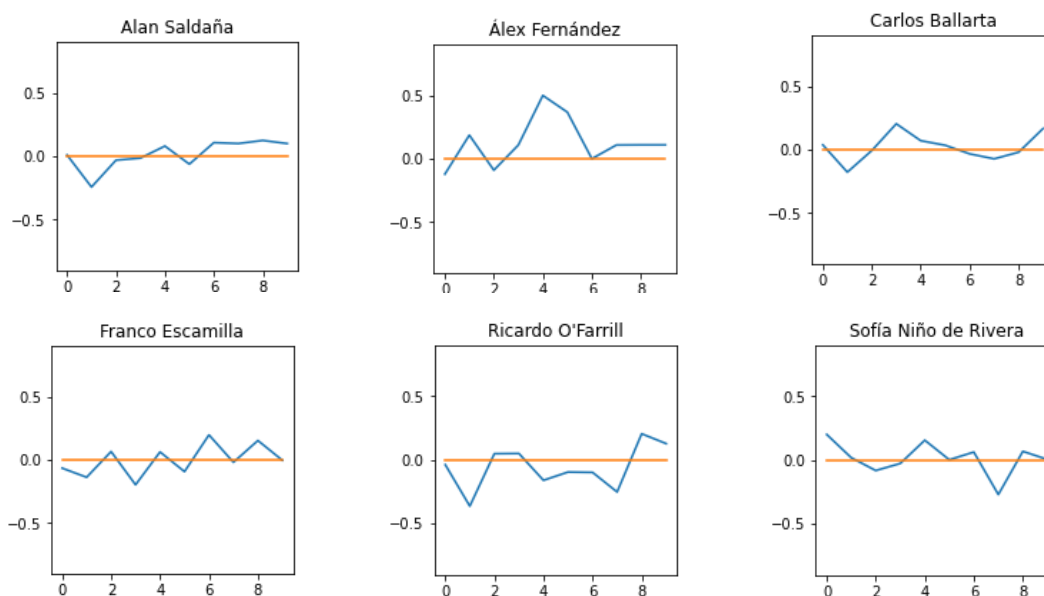
b) Subjetividad: qué tan subjetivas u objetivas eran las palabras, siendo 0 muy objetivo y +1 muy subjetivo.

Para este fin, se empleó el módulo TextBlob, una librería para Python desarrollada a partir del trabajo de investigadores lingüísticos que se han dado a la tarea de etiquetar el sentimiento de las palabras en función de su significado y contexto (Loria, 2020). Dentro de los hallazgos

referentes a la subjetividad de los monólogos, se encontró que la narrativa empleada por Franco Escamilla es significativamente más factual y anecdótica que el resto de los comediantes estudiados. De igual manera, la polaridad de su discurso fue la más neutra, seguida de la de Alan Saldaña. Esto deja entrever un humor basado, mayormente, en hechos que son compartidos con la audiencia mediante un vocabulario que, en promedio, resulta en una connotación neutra.

Por último, para concluir con el análisis, se examinó el comportamiento del sentimiento general de los monólogos en el tiempo (véase *Gráfica 2*). Para ello, se dividió la rutina de cada comediante en subintervalos de, aproximadamente, un minuto cada uno. De los resultados obtenidos se puede destacar que el monólogo de Franco Escamilla tiene la particularidad de asemejarse a una función cosenoidal, con crestas y valles que tienen una ligera tendencia positiva, cuando se evalúa su sentimiento con respecto del tiempo. Es un descubrimiento notable que apunta hacia una estructura lingüística tal que el interlocutor perciba mensajes con significados ligeramente negativos, intercalados con otros ligeramente positivos que terminan por neutralizarse mutuamente, a medida que avanza la narración.

Gráfica 2: Sentimiento general de cada monólogo en el tiempo



Conclusión

Como respuesta a nuestra interrogante sobre lo que hace sobresalir al comediante Franco Escamilla, se concluye del análisis que su humor se caracteriza por ser objetivo y de polaridad neutra. Es decir, sus monólogos tienden a construirse sobre bases anecdóticas, de modo que su narrativa resulta más factual que subjetiva al considerársele como un todo. Esto concuerda con el estudio realizado por Soto-Coelho (2020), pues refiere, sobre las características del humor de Franco, una narración que atrapar la atención del público y hace que nos pongamos en su lugar mientras narra sus experiencias (p. 48). Asimismo, la evolución de sus rutinas tiene la singularidad de mantener una polaridad neutra que se alcanza gracias a un balance excepcional entre las frases de connotaciones más negativa con aquellas más positivas, que se traduce, a su vez, en un equilibrio del sentimiento general del monólogo con respecto al tiempo.

De igual manera, como parte de los hallazgos, aunque de menor valor y significación que los anteriores, se encontró que Franco hace un uso mesurado de las palabras soeces. Similarmente, su elección de vocabulario sugiere una secuencia y ritmo adecuados en comparación con los monólogos de los otros comediantes evaluados. Ahora bien, en el caso específico de Franco Escamilla, la entonación y los gestos del comediante juegan un papel muy importante al momento de estar ejecutando el monólogo frente al público (Soto-Coelho, *op cit*, p. 49). Por tanto, ha de aclararse expresamente que el presente proyecto tiene un alcance limitado, pues incluye únicamente algunos de los elementos lingüísticos que se pueden estudiar a través del análisis de sentimiento. En este sentido, los resultados obtenidos son de carácter preliminar y menesterosos de revisiones más integrales, que incluyan una mayor cantidad de datos y técnicas similares basadas en aprendizaje automático¹.

¹ En atención a esta necesidad, se adjunta el enlace al repositorio de GitHub donde se puede consultar libremente el código: <https://github.com/HarimHidal/nlp-mxstandup>

Referencias bibliográficas:

1. Bateson, P. (2005). The nature of play: Great apes and humans. En A. D. Smith, *The role of play in the evolution of great apes and humans* (pp. 13-24). Nueva York: Guilford Press.
2. Bengochea, M. M. (2016). *El stand up a través de la comunicación. Análisis y caracterización de un fenómeno en crecimiento*. Tesis de grado, Universidad Nacional de Rosario. Recuperado de <https://rephip.unr.edu.ar/bitstream/handle/2133/8267/Tesina%20final%20-%20Bengochea%20Digital.pdf?sequence=3&isAllowed=y>
3. Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. En Synthesis Lectures on Human Language Technologies (Vol. 5, pp.1-167). Morgan & Claypool Publishers.
4. Henríquez, C. et al. (2017). *Análisis de sentimientos a nivel de aspecto usando ontologías y aprendizaje automático*. Procesamiento del Lenguaje Natural, (59),49-56. Recuperado de <https://www.redalyc.org/articulo.oa?id=515754427005>
5. López-Briega, R. P. (2017). *Matemáticas, análisis de datos y python*. Procesamiento del Lenguaje Natural con Python. Recuperado de <https://relopezbriega.github.io/blog/2017/09/23/procesamiento-del-lenguaje-natural-con-python/>
6. Hernández, M. & Gómez, J. (2014). *Análisis de Sentimientos Aplicado a Referencias Bibliográficas*. Revista EPN, 33(3), 64-71. Recuperado de https://revistapolitecnica.epn.edu.ec/ojs2/index.php/revista_politecnica2/article/download/126/pdf/1697
7. Loria, S. (2020). *TextBlob*. TextBlob: Simplified Text Processing. Recuperado de <https://textblob.readthedocs.io/en/dev/>
8. Soto-Coelho, N. (2020). *Estrategias lingüísticas utilizadas para la creación de humor en los monólogos de Franco Escamilla*. Tesis de grado, Pontificia Universidad Católica del Ecuador. Recuperado de <http://repositorio.puce.edu.ec/bitstream/handle/22000/18559/Estrategias%20ling%C3%BC%C3%ADsticas%20utilizadas%20para%20la%20creaci%C3%B3n.pdf?sequence=1&isAllowed=y>