

Estudio predictivo del uso colectivo de hashtags en museos de la red REMEDI

Collective predictive analysis in REMEDI museums hashtag's usage

Víctor Yeste^a, Ángeles Calduch-Losa^b y Jorge Serrano-Cobos^c

^aUniversitat Politècnica de València, València, España, vicesmo@inf.upv.es, ^bUniversitat Politècnica de València, València, España, mcalduch@cio.upv.es y ^cUniversitat Politècnica de València, València, España, jorserc2@har.upv.es

Resumen

En la literatura científica se suele estudiar el marketing cultural de museos desde la perspectiva del caso, analizando proyectos específicos; o, por el contrario, de forma genérica, sin basar el estudio en datos, en ocasiones difíciles de capturar. En este caso, se analiza la actividad conjunta de una muestra de los museos de la red REMEDI en Twitter, extrayendo datos cibernéticos de hashtags comunes y específicos utilizados por los miembros de la red. Sobre esos datos, se estudiará y aplicará la capacidad predictiva (para ciertos parámetros de éxito en términos de marketing cultural digital) de la regresión binomial negativa o la regresión de Poisson, unos tipos de modelo lineal generalizados que se usan para trabajar con datos de conteo (número de veces que ocurre cierto fenómeno). Como resultado obtendremos un mapa de ciertas características de comunicación empleadas por este subconjunto de cuentas sociales, a través de esta red social, cuyo rendimiento futuro podrá ser aproximado en base a la capacidad de predicción de distintas variables, que permitirían a esas cuentas conseguir mejor rendimiento, en términos de éxito digital.

Palabras clave: *analítica de redes sociales, analítica de Twitter, análisis de hashtags, análisis de tendencias, predicción de tendencias*

Abstract

In scientific literature, museum cultural marketing is usually studied from the perspective of the case, analyzing specific projects; or, on the contrary, in a generic way, without basing the study on data, sometimes difficult to capture. In this case, the joint activity of a sample of the museums of the REMEDI network on Twitter is analyzed, extracting cybermetric data from common and specific hashtags used by the members of the network. On these data, the predictive capacity (for certain parameters of success in terms of digital

cultural marketing) of the negative binomial regression or Poisson regression, generalized linear model types used to work with count data (number of times that certain phenomenon occurs). As a result, we will obtain a map of certain communication characteristics used by this subset of social accounts, through this social network, whose future performance may be approximated based on the prediction capacity of different variables, which would allow those accounts to achieve better performance, in terms of digital success.

Keywords: *social media analytics, Twitter analytics, hashtag analysis, trend analysis, trend prediction*

1. Introducción y Estado de la Cuestión

Los museos son una de las tipologías de entidades culturales que más trabajan el marketing cultural para atraer visitantes, según Monistrol (2009) Pero ¿qué es el marketing cultural? La expresión atiende a dos acepciones. Según IGI-Global (n.d.), es “un tipo específico de marketing cuyo objetivo es promover un mensaje, un servicio o un producto a un grupo de compradores potenciales de una cultura o demografía en particular”. En español, haríamos más bien referencia a que el marketing cultural “es el proceso que se desarrolla en las organizaciones culturales y en la sociedad para facilitar el intercambio a través de relaciones colaborativas que crean un valor recíproco mediante el uso de recursos complementarios” (Leal y Quero, 2011). La diferencia viene de un problema de traducción, porque el concepto español se debería traducir en realidad por “Marketing for Cultural Organizations”.

Por tanto, el marketing cultural tal como lo entendemos en España, se desarrolló a finales de los años 60 (Monistrol, 2009) en organizaciones culturales como lo son los museos, y se produjo a través de relaciones colaborativas, antes sólo offline y hoy día también en buena parte, a través del mundo online, donde el valor recíproco deviene mejora de marca para la organización cultural, y, por ende, un deseable aumento de visitantes.

En el caso que nos ocupa, los museos, entre las técnicas de marketing digital más usadas, destaca el uso de las redes sociales. Así, Claes y Deltell (2019) analizaron la comunicación online de los diecinueve museos que pertenecen a la Dirección General de Bellas Artes de España. En la metodología se contabilizó el número de seguidores, el uso de idiomas, el volumen de publicaciones, el estudio de las webs, los backlinks de las mismas, así como el seguimiento de las actividades. El estudio reflejó la especial incidencia de la red social Facebook en la comunicación digital de los mismos, y la carencia en el uso de otras lenguas.

Mas Iglesias (2018) analizó el contenido de mensajes publicados en Facebook durante 6 meses de 2018 por los 10 museos españoles con más seguidores en este medio social, descubriendo que los mensajes promocionales prevalecían sobre los de creación de comunidad. Y Villaespesa (2018) se centró en la conceptualización de motivaciones de usuarios para interactuar con los sitios web de museos, lo que ayuda a establecer una base para la segmentación de los mismos.

Sólo en el caso de Abad y Caerols se estudian en 2016 los hashtags, en este caso el hashtag “#5Museos”, impulsado conjuntamente por los museos Lázaro Galdiano, Cerralbo, Romanticismo, Sorolla y Artes Decorativas. Por otro lado, el intento más serio para profundizar en los hábitos de marketing cultural en redes sociales, y en particular en Twitter, es el del estudio de Susana Llerena (2015), que analizó múltiples variables de análisis cualitativo y cuantitativo de 10 museos españoles, en particular referentes a la detección de buenas prácticas. Pero hasta la fecha, no hay estudios de mayor calado, que cuenten con Big Data en marketing cultural digital para el análisis de la actividad de un número exhaustivo de museos en redes sociales.

Este trabajo pretende ser el primero de dicha tipología de análisis, donde se intenta no sólo descubrir patrones de comportamiento a nivel grupal, sino predecir mediante alguna variable, el posible rendimiento del contenido de estas cuentas sociales, de forma que sirva para detectar mejores prácticas para el colectivo de museos españoles, en concreto aquellos pertenecientes a REMED (Red de Museos y Estrategias Digitales).

2. Objetivos

Este estudio consiste en una investigación cuantitativa, explicativa y no experimental, en la cual se hace uso de una inferencia inductiva de una manera longitudinal. Así, se estudia el uso de hashtags por parte de las cuentas de Twitter del conjunto de museos que forman parte de REMED, y se realiza el análisis de tendencias de los hashtags por parte de los usuarios de Twitter en español.

La variable principal es el conteo de favoritos, y se toma como hipótesis de este estudio que es posible realizar la predicción de la variable principal cinco semanas después. El ámbito de estudio está formado las 104 cuentas de Twitter de los museos que forman parte de REMED.

Se cuenta con siete variables de análisis que explican la información relacionada con el uso de los hashtags, tanto en la dimensión de las cuentas de Twitter de museos de la muestra escogida (prefijo “m_” en las variables) como de los usuarios de Twitter en español en general (prefijo “tw_” en las variables). Todas las variables representan los datos en modo conteo, lo cual significa que suman el total de los datos captados para cada tweet de cada hashtag procesado:

- 1) Número de tweets (nombre de variable “num_tweets”)
- 2) Número de retweets (nombre de variable “retweet_count”)
- 3) Número de favoritos (nombre de variable “favorite_count”)
- 4) Número de seguidores de los autores de los tweets (nombre de variable “user_num_followers”)
- 5) Número de tweets publicados por los autores de los tweets (nombre de variable “user_num_tweets”)
- 6) Edad en días de las cuentas de Twitter de los autores de los tweets (nombre de variable “user_age”)
- 7) Número de tweets que incluyen una URL (nombre de variable “url_inclusion”)

Con las variables anteriormente mencionadas, se ha realizado una investigación comprobando las correlaciones entre las variables y realizando un análisis de regresión. Así, se averigua las relaciones entre las variables y se analiza si es posible realizar la predicción de la cantidad de favoritos de los hashtags utilizados por los museos. La primera toma inicial se ha presentado con el sufijo “_t0” en las variables y la toma realizada 5 semanas después se ha mostrado con el sufijo “_t1”.

3. Desarrollo de la innovación

El desarrollo del estudio ha consistido en la extracción de la información, el análisis de las variables, correlaciones y regresiones estadísticas y el cálculo de la precisión con el que se han realizado las regresiones.

3.1 Obtención de la información

Se han realizado dos tomas de datos: la toma inicial se ha llevado a cabo entre el 3 y el 9 de enero de 2021, y la segunda toma ha sido cinco semanas después, entre el 7 y el 14 de febrero del mismo año. La duración ha sido la necesaria para extraer los datos según las limitaciones metodológicas de la versión estándar de la API de Twitter (versión 1.1), puesto que sólo se pueden realizar 180 peticiones en franjas de 15 minutos, y cada petición consta de un conjunto de 100 tweets. El resultado total es de 18.000 posibles tweets en cada intervalo de 15 minutos (Standard search API, s.f.). La funcionalidad de obtener datos de los perfiles de Twitter también imponía un máximo de los últimos 3.200 tweets publicados por cada cuenta de Twitter (GET statuses/user_timeline, s.f.). La funcionalidad de búsqueda, por otra parte, permite buscar cada hashtag en Twitter en general, pero sólo devuelve una muestra de los tweets publicados en los siete días anteriores a cuando se realiza la petición (Search Tweets: standard v1.1, s.f.).

3.2 Análisis de la información

La información extraída en forma de las variables del análisis descritas en el apartado 2 permite que se realice el análisis conjunto de las variables de la toma inicial y las variables de favoritos de la toma final. Así, se pueden comprobar las correlaciones Pearson entre las variables que se van a utilizar en el estudio, de manera que puedan detectarse correlaciones muy fuertes (cercanas a 1 en valor absoluto) entre alguna de las variables independientes (toma inicial) y las variables dependientes (favoritos en la toma final). Si fuera así, se escogerá aquella que tenga un mayor coeficiente de correlación o, en caso de ser el mismo entre varias, la que tenga un sesgo y curtosis estandarizados menores. La variable escogida servirá para llevar a cabo la regresión lineal simple de la variable objetivo, siempre y cuando ambas sigan una distribución normal, condición para que se pueda realizar una regresión lineal (Barón López & Téllez Montiel, s.f.).

Si no hay ninguna correlación muy fuerte (cercana a 1), entonces será adecuado realizar un análisis de regresión siguiendo el modelo de Poisson o el de regresión binomial negativa, puesto que ambos son los más utilizados para datos de tipo conteo. El modelo de Poisson es más recomendable si no existe sobre dispersión en los datos, pero si la hubiera se apostaría por la regresión binomial negativa (Lawless, 1987), ya que la regresión binomial negativa capta parte de la varianza que no es capaz de detectar el modelo de Poisson. Si hubiera sobre dispersión, se produciría una subestimación en los errores estándares de los coeficientes. La sobre dispersión, sin embargo, es muy frecuente en fenómenos que se producen con

recurrencia, como es el caso del contenido que se publica en formato digital (Navarro, Utzet, Puig, Caminal, & Martín, 2001).

Para detectar la sobre dispersión, se ha realizado la prueba de dispersión estadística chi-cuadrado de Pearson, por lo que si devuelve un valor cercano a uno la varianza será igual a la media y el modelo de Poisson será el escogido. En el caso contrario, se optará por la regresión binomial negativa.

Para todos los análisis de este estudio se ha utilizado el software STATGRAPHICS 18.

3.3 Cálculo de la precisión

Las ecuaciones de regresión presentan una variabilidad que está incluida en sus modelos (R-cuadrado), que representa hasta qué punto el modelo explica la información en la toma final. Además, se calcula el error medio absoluto (MAE) y la desviación estándar de los errores de predicción en forma de la raíz del error medio cuadrático (RMSE), de manera que se pueda interpretar en base a ellas la precisión con la que se ha realizado la predicción de los datos de la toma realizada cinco semanas después en base a la toma inicial.

Ambos valores de error son muy usados a la hora de evaluar modelos. No obstante, RMSE es una medida muy sensible a valores atípicos, ya que les da un mayor peso en su cálculo, mientras que MAE cancela valores con signos contrarios al no realizar el cuadrado y la raíz posterior. En este estudio se ha comprobado que existe en la información una gran sobre dispersión, por lo que es conveniente añadir el cálculo del MAE para incluir otra medición que no se vea tan afectada por esta situación. (Chai & Draxler, 2014)

El error medio absoluto (MAE) se calcula de la siguiente manera: (Chai & Draxler, 2014)

$$MAE = \sum(V_o - V_p)/N \quad (1)$$

La raíz del error medio cuadrático (RMSE) se calcula así: (Sridhar & Khan, 2014)

$$RMSE = \sqrt{\sum(V_o - V_p)^2/N} \quad (2)$$

Donde

V_o : Valor obtenido en la toma 14 días después

V_p : Valor predicho por la ecuación de regresión

N : Número de puntos de datos medidos

4. Resultados

4.1 Resumen general

Se han almacenado 950.259 datos en total, provenientes del análisis del uso de 17.627 hashtags en dos tomas de datos por parte de la muestra de 104 cuentas de Twitter de museos.

Se han procesado 7.737.155 tweets en total, de los cuales 209.727 tweets han sido publicados por los museos de la muestra y los 7.527.428 restantes por usuarios de Twitter en general.

Se ha detectado una diferencia significativa entre la inclusión de una URL en los tweets publicados por los dos segmentos: el 32,91% de los tweets publicados por usuarios de Twitter en general incluyen una URL, mientras que en los tweets publicados por los museos el ratio ha sido del 81,7%. También se ha comprobado una diferencia grande en la edad en días de las cuentas que han publicado tweets en ambos segmentos. Extrayendo la edad en días en modo conteo de cada tweet procesado, el promedio en días de las cuentas de los usuarios de Twitter ha sido de 2.175,71 días, mientras que las cuentas de los museos han presentado un promedio de 3.208,21 días.

4.2 Hashtags más usados y con más favoritos

Se han extraído asimismo los hashtags que han mostrado un mayor uso por parte de cada segmento en la toma inicial, así como aquellos que más favoritos han recibido.

Tabla 1. Diez hashtags que han presentado un mayor uso en forma de tweets

Museos		Twitter	
Hashtag	Tweets	Hashtag	Tweets
mequinenza	2186	quédateencasa	26627
museumweek	1270	felizmiércoles	24956
museo	1237	ecuador	24495
picasso	911	panamá	21117
museos	888	niunamenos	17722
museopusol	867	toquedequeda	17722
despiertaonarte	687	opinión	17722
fycma	648	envivo	17722
villajoyosa	642	empleo	17719
málaga	637	amlo	17716

Se puede comprobar en la tabla anterior que no coincide ninguno de los tweets más utilizados por los museos con los de Twitter en sí, lo cual sin duda responde a una intención de uso y una especificación diferente. Los hashtags más usados por los museos están, en su mayoría, relacionados con su sector, mientras que Twitter atiende a muchísimas otras temáticas y tendencias más generalistas.

Tabla 2. Diez hashtags que más favoritos han recibido

Museos		Twitter	
Hashtag	Tweets	Hashtag	Tweets
metgala	183597	nuevafotodeperfil	103219
metcamp	177766	libra	89226

despiertaconarte	36907	felizmiércoles	76079
ezramiller	22293	realmadrid	75036
katyperry	15301	mamá	72131
metkidsfunfact	13826	buenasnoches	63503
allbyyourshelf	12079	undíacomohoy	52895
themet	12001	cuartomilenio	52061
picasso	11902	buenosdiasatodos	50216
taldíacomohoy	10374	barcelona	49231

En la tabla anterior, de nuevo, ambos listados no coinciden en ninguno de sus hashtags, aunque hay dos muy similares como son “taldíacomohoy” para los museos y “undíacomohoy” para Twitter. Esta vez, el listado proveniente de los museos incorpora hashtags más generalistas, mientras que el de Twitter incluye tendencias todavía más generales (felicitaciones de día de la semana, de buenas noches, de buenos días...).

4.3 Análisis estadístico de los favoritos

El análisis del uso de los hashtags se ha realizado en base a las siete variables tanto de los museos como de Twitter en sí, mencionadas en el apartado 2, por lo que son un total de catorce variables estadísticas de tipo conteo. Esta información servirá para tratar de predecir los favoritos obtenidos por parte de los hashtags cinco semanas después, siendo por tanto dos objetivos y por ello dos variables dependientes: los favoritos de las cuentas de los museos y los favoritos en Twitter.

La variable de los favoritos en los tweets provenientes de las cuentas de los museos en la toma final ha presentado un promedio de 76,18, un sesgo estandarizado de 4.693,27 y una curtosis estandarizada de 212.774. Para comprobar su relación con las variables estadísticas iniciales, se ha realizado un análisis multivariado mediante el cual se han extraído las correlaciones Pearson entre todas ellas y la variable dependiente:



Fuente propia desde STATGRAPHICS 18

Fig. 1 Correlaciones Pearson entre las variables de la toma inicial y los favoritos en las cuentas de los museos en la toma final

La variable de los favoritos en las cuentas de los museos (m_favorite_count_t1) presenta una correlación muy fuerte (muy cercana a 1) con dos variables: con los retweets en las cuentas de los museos (m_retweet_count_t0) tiene 0,9977 y con los favoritos en las cuentas de los museos (m_favorite_count_t0) tiene 0,9998, por lo que se elige esta última para evaluar la regresión.

Si la correlación hubiera sido cercana a 1 en valor absoluto pero menor, se justificaría el uso de una regresión lineal entre esta última y la variable objetivo. Como no siguen una distribución normal debido a que tienen unos valores muy grandes de sesgo y curtosis estandarizados, incluso tras tratar de normalizarlas con una transformación logarítmica o logarítmica doble, esta posibilidad se descartaría. Sin embargo, al tratarse de una correlación prácticamente igual a 1, se interpreta que el número de favoritos en las cuentas de museos varía de manera despreciable en un intervalo de cinco semanas, y por tanto la regresión no tiene sentido a nivel estadístico.

Si se desea comprobar de todas maneras la información que aportan el resto de variables independientes a los favoritos en las cuentas de los museos, es necesario realizar una regresión de Poisson o binomial negativa. Para ello, en primer lugar, se hace la prueba chi-cuadrado en la variable objetivo con el fin de calcular su desviación estándar. Esta da un valor de 6,85925E10 con p-valor cercano a cero, un valor muy lejano del 1, por lo que se escoge la regresión binomial negativa con un 95% de confianza.

Como una de las condiciones estadísticas para realizar una regresión binomial negativa es que las variables de predicción sean independientes, es decir, que no sean combinaciones lineales de las otras variables presentes en el modelo, se aplica un filtro de alta correlación (0,7 o más de manera absoluta) dando preferencia a aquellas que tengan una correlación Pearson mayor con la variable objetivo (tw_favorite_count_t1) y, en caso de un valor igual, escogiendo aquella que presente un sesgo y curtosis estandarizados menores. Tras aplicar el filtro, se obtiene la siguiente tabla de correlaciones Pearson:

	m_favorite_count_t0	m_user_num_followers_t0	m_user_num_tweets_t0	tw_retweet_count_t0	tw_favorite_count_t0	tw_user_num_followers_t0	m_favorite_count_t1
m_favorite_count_t0	1,00	0,44	0,18	0,00	0,02	0,01	0,44
m_user_num_followers_t0	0,44	1,00	0,39	0,00	0,02	0,00	0,44
m_user_num_tweets_t0	0,18	0,39	1,00	0,02	0,10	0,04	0,18
tw_retweet_count_t0	0,00	0,00	0,02	1,00	0,26	0,25	0,00
tw_favorite_count_t0	0,02	0,02	0,10	0,26	1,00	0,42	0,02
tw_user_num_followers_t0	0,01	0,00	0,04	0,25	0,42	1,00	0,01
m_favorite_count_t1	0,44	0,44	0,18	0,00	0,02	0,01	1,00

Fuente propia desde STATGRAPHICS 18

Fig. 1 Correlaciones Pearson entre las variables de la toma inicial tras aplicar el filtro de alta correlación y los favoritos en el conjunto de los museos en la toma final

Se realiza, por tanto, la regresión binomial negativa mediante una selección de factores de Pasos hacia Atrás, que comienza con todas las variables disponibles salvo la de favoritos de las cuentas de los museos en el instante inicial ($m_favorite_count_t0$) y elimina aquellas que tengan un p-valor mayor a 0,05 en el modelo. El resultado es un modelo ajustado con p-valor cercano a cero que incluye el 11,41% de la variabilidad del número de favoritos. Presenta un error absoluto medio (MAE) de 2,75056E13 y una raíz del error medio cuadrático (RMSE) de 3,44959E25. La ecuación de regresión es la siguiente:

$$m_favorite_count_t1 = \exp(24,3763 + 1,83154E - 8 * m_user_num_followers_t0 - 5,37998E - 7 * m_user_num_tweets_t0)$$

Donde:

$m_favorite_count_t1$ = Favoritos para un hashtag en el conjunto de los museos en la toma de cinco semanas después

$m_user_num_followers_t0$ = Número de seguidores de los autores de los tweets de las cuentas de los museos en la toma inicial

$m_user_num_tweets_t0$ = Número de tweets publicados por los autores de los tweets de las cuentas de los museos en la toma inicial

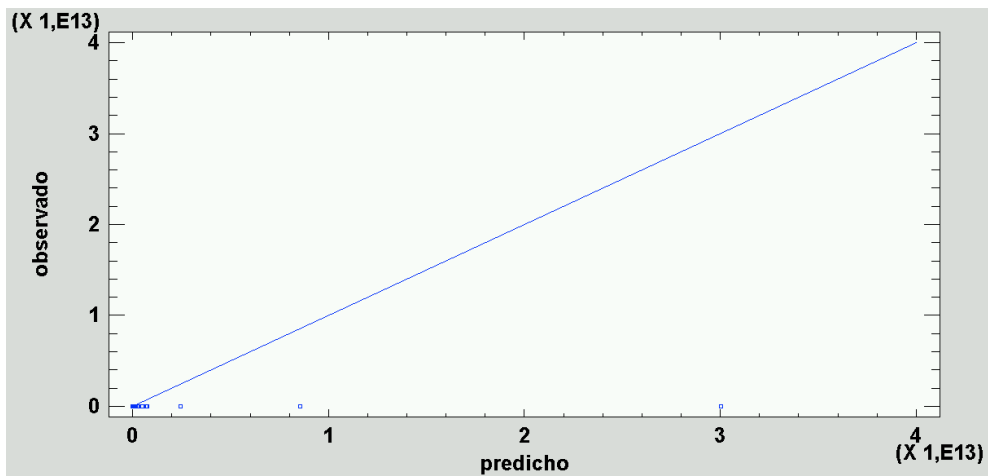


Fig. 2 Gráfico Observado vs Predicho según la ecuación de regresión de los favoritos en el conjunto de los museos

En el gráfico anterior se puede ver que los puntos fruto de la relación entre los valores observados y los predichos se encuentran cercanos al principio y luego se van alejando de la recta de regresión conforme aumenta la cifra del predicho, lo cual demuestra visualmente una variabilidad baja del modelo.

Por otro lado, la variable de los favoritos en los tweets provenientes de Twitter en la toma final ha presentado un promedio de 427,67, un sesgo estandarizado de 772,1 y una curtosis estandarizada de 7.666,83. A la hora de comprobar su relación con las variables estadísticas

iniciales, se ha realizado un análisis multivariado mediante el cual se han extraído las correlaciones Pearson entre todas ellas y esta nueva variable dependiente:



Fig. 1 Correlaciones Pearson entre las variables de la toma inicial y los favoritos en el conjunto de usuarios de Twitter en la toma final

La variable de los favoritos en las cuentas de Twitter en la toma final (tw_favorite_count_t1) no presenta ninguna correlación muy fuerte (cercana a 1 absoluto) con ninguna de las variables de la toma inicial, por lo que se decide realizar una regresión de Poisson o binomial negativa. Para ello, en primer lugar, se hace la prueba chi-cuadrado en la variable objetivo con el objetivo de calcular su desviación estándar. Esta da un valor de 1,44599E11 con p-valor cercano a cero, un valor muy lejano del 1, por lo que se escoge la regresión binomial negativa con un 95% de confianza.

Como una de las condiciones estadísticas para realizar una regresión binomial negativa es que las variables de predicción sean independientes, es decir, que no sean combinaciones lineales de las otras variables presentes en el modelo, se aplica un filtro de alta correlación (0,7 o más de manera absoluta) dando preferencia a aquellas que tengan una correlación Pearson mayor con la variable objetivo (tw_favorite_count_t1) y, en caso de un valor igual, escogiendo aquella que presente un sesgo y curtosis estandarizados menores. Tras aplicar el filtro, se obtiene la siguiente tabla de correlaciones Pearson:



Fuente propia desde STATGRAPHICS 18

Fig. 1 Correlaciones Pearson entre las variables de la toma inicial tras aplicar el filtro de alta correlación y los favoritos en el conjunto de usuarios de Twitter en la toma final

Se realiza, por tanto, la regresión binomial negativa mediante una selección de factores de Pasos hacia Atrás, que comienza con todas las variables disponibles y elimina aquellas que tengan un p-valor mayor a 0,05 en el modelo. El resultado es un modelo ajustado con p-valor cercano a cero que incluye el 2,83% de la variabilidad del número de favoritos. Presenta un error absoluto medio (MAE) de 9,7718E12 y una raíz del error medio cuadrático (RMSE) de 1,09343E25. La ecuación de regresión es la siguiente:

$$tw_favorite_count_t1 = \exp(24,3283 + 0,000113038 * m_retweet_count_t0 + 0,0000293429 * tw_favorite_count_t0)$$

Donde:

tw_favorite_count_t1 = Favoritos para un hashtag en el conjunto de los usuarios de Twitter en la toma de cinco semanas después

m_retweet_count_t0 = Retweets para un hashtag en el conjunto de los museos en la toma inicial

tw_favorite_count_t0 = Favoritos para un hashtag en el conjunto de los usuarios de Twitter en la toma inicial

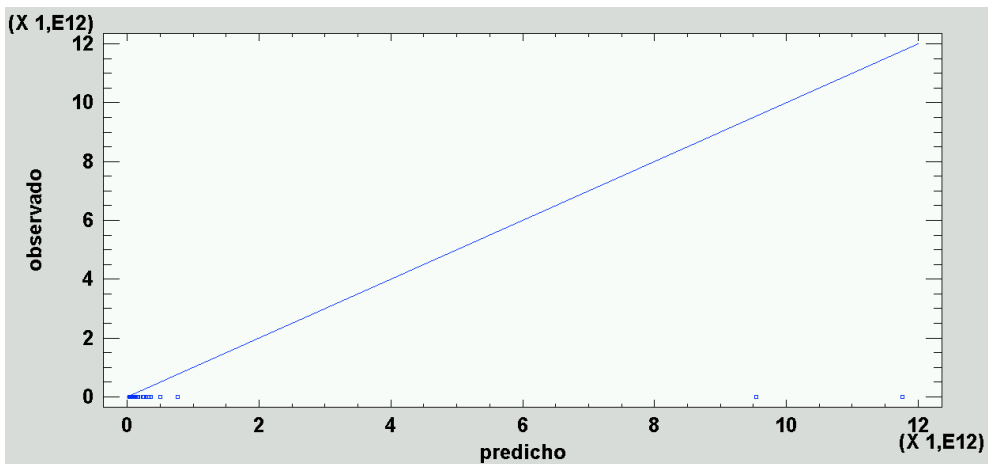


Fig. 2 Gráfico Observado vs Predicho según la ecuación de regresión de los favoritos en el conjunto de los usuarios de Twitter

En el gráfico anterior se puede ver que los puntos que relacionan los valores observados y los predichos se alejan de la recta de regresión, lo cual demuestra visualmente una variabilidad muy baja del modelo.

5. Conclusiones

El análisis de regresión de los favoritos se ha realizado a cinco semanas vista, y desde la perspectiva de las cuentas de Twitter de los museos se ha obtenido una correlación prácticamente absoluta, que ha permitido confirmar la hipótesis de que en ese conjunto se puede predecir los favoritos de un hashtag en concreto, ya que la variación es despreciable y no justifica una regresión estadística. Sin embargo, no ha sido así en el caso de los usuarios de Twitter, que responden al análisis de la tendencia del hashtag en general. No ha sido posible realizar una regresión lineal simple, por lo que se ha escogido la regresión binomial negativa y ha devuelto una variabilidad muy pequeña. Por tanto, en el caso del conjunto de los usuarios de Twitter, la hipótesis de poder predecir los favoritos se rechaza.

También se ha podido comprobar que los hashtags vienen acompañados de un uso más intensivo de la inclusión de URLs en los tweets por parte de los museos que en Twitter en general, lo cual responde probablemente a una intención diferente y más orientada a la de la emisión de información publicada en las webs de los museos y otros medios.

Se podría estudiar en el futuro la posibilidad de predecir los favoritos con nuevas variables estadísticas que aporten más información sobre el uso de los hashtags por parte de las cuentas de Twitter. También sería interesante realizar más tomas de datos y comprobar la precisión de las regresiones estadísticas mediante datos de test, es decir, realizando una tercera toma y comprobando si el error aumenta o disminuye con un conjunto totalmente diferente de datos del que ha servido para las regresiones.

Otro detalle que podría analizarse sería la implicación de las fechas de festivos en España, así como los fines de semana con respecto a los días de entre semana. Además, sería conveniente analizar los valores anómalos de las variables y las características que los acompañan ya que su presencia, aunque perjudicial para la regresión estadística al favorecer la sobre dispersión, puede ser muy beneficiosa para la estrategia de marketing de los museos: buscar, de esta manera, el mayor número de favoritos posible.

Referencias

- BARÓN LÓPEZ, F. J. & TÉLLEZ MONTIEL, F. (s.f.). “Regresión múltiple” en *Apuntes de Bioestadística*. <<https://www.bioestadistica.uma.es/baron/apuntes/ficheros/cap06.pdf>> [Consulta: 15 de febrero de 2021]
- CAEROLS-MATEO, R., VIÑARÁS-ABAD, M., & GONZÁLVEZ-VALLES, J. E. (2017). “Social networking sites and museums: Analysis of the Twitter campaigns for International Museum Day and Night of Museums” en *Revista Latina de Comunicación Social*, 72, 220–234. <<https://doi.org/10.4185/RLCS-2017-1162>>
- CAPRIOTTI, P., & LOSADA-DÍAZ, J. C. (2018). “Facebook as a dialogic communication tool at the most visited museums of the world” en *Profesional de la Información*, 27(3), 642–650. <<https://doi.org/10.3145/epi.2018.may.17>>

- CHAI, T. y DRAXLER, R. R. (2014). "Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature" en *Geoscientific Model Development*, vol. 7, p. 1247-1250.
- CLAES, F., & DELTELL, L. (2014). "Social museums: Social media profiles in Twitter and Facebook 2012-2013" en *Profesional de La Informacion*, 23(6), 594-602. <<https://doi.org/10.3145/epi.2014.nov.06>>
- CLAES, F., & DELTELL, L. (2019). "Social museum in Spain: Social networks and websites of state museums" en *Profesional de La Informacion*, 28(3). <<https://doi.org/10.3145/epi.2019.may.04>>
- IGI GLOBAL. (n.d.). "What is Cultural Marketing" en *IGI Global*. <<https://www.igi-global.com/dictionary/the-assessment-of-cultural-experience-through-the-measurement-of-cross-cutting-skills/39240>> [Consulta: 10 de enero de 2021]
- LAWLESS, J. F. (1987). "Negative binomial and mixed Poisson regression" en *The Canadian Journal of Statistics*, vol. 15, issue 3, p. 209-225.
- LEAL, ANTONIO; QUERO, M. J. (2011). *Manual de marketing y comunicación cultural. Cádiz. I. y C. de la J. de A. Dirección General de Universidades de la Consejería de Economía (ed.). Universidad de Cádiz. Servicio de Publicaciones.* <https://www.bizkaia.es/home2/archivos/DPTO4/Temas/producto44manual-de-marketing-y-comunicacion-cultural_web.pdf>
- LLERENA IGLESIAS, S. (2016). "La comunicación de los museos españoles en Twitter: análisis de la situación y establecimiento de buenas prácticas". <<https://e-archivo.uc3m.es/handle/10016/22291Z>> [Consulta: 10 de enero de 2021]
- LOSADA-DÍAZ, J. C., CAPRIOTTI, P. (2015). "The communication of art museums in facebook: Comparison of key international and Spanish institutions" en *Palabra Clave*, 18(3), 889-904. <<https://doi.org/10.5294/pacla.2015.18.3.11>>
- MAS IGLESIAS, J. M. (2018). "Spanish museums on Facebook: Analysis of their communication as social museums" en *Revista de Comunicación*, 17(2), 185-207. <<https://doi.org/10.26441/RC17.2-2018-A8>>
- MONISTROL, R. (2009). "Evolución y aplicación del marketing cultural en los museos" en *BiD: textos universitaris de biblioteconomia i documentació*. Facultad de Biblioteconomía y Documentación, Universidad de Barcelona. <[Http://bid.ub.edu/23/monistrol2.htm](http://bid.ub.edu/23/monistrol2.htm)>
- NAVARRO, A ; UTZET, F. ; PUIG, P. ; CAMINAL, J. y MARTÍN, M. (2001). "Negative binomial distribution versus Poisson in the analysis of recurrent phenomena" en *Gaceta sanitaria / S.E.S.P.A.S.*, vol. 15, issue 5, p. 447-452.
- SRIDHAR, B. y KHAN, M. Z. A. (2014) "RMSE comparison of Path Loss Models for UHF/VHF bands in India". López, D. (coord.) En: *IEEE Region 10 Symposium (14-16 abril 2014, Kuala Lumpur, Malasia)*. 330-335
- TWITTER DEVELOPERS (s.f.). *Standard search API*. <<https://developer.twitter.com/en/docs/twitter-api/v1/tweets/search/api-reference/get-search-tweets>> [Consulta: 15 de febrero de 2021] [sin autoría reconocida]
- TWITTER DEVELOPERS (s.f.). *GET statuses/user_timeline*. <https://developer.twitter.com/en/docs/twitter-api/v1/tweets/timelines/api-reference/get-statuses-user_timeline> [Consulta: 15 de febrero de 2021] [sin autoría reconocida]

- TWITTER DEVELOPERS (s.f.). *Search Tweets: standard v1.1*. <<https://developer.twitter.com/en/docs/twitter-api/v1/tweets/search/overview>> [Consulta: 15 de febrero de 2021] [sin autoría reconocida]
- VILLAESPESA, E. (2018). “Evaluación de la estrategia digital en los museos: métodos y herramientas para web, redes sociales y móviles” en *Periférica Internacional. Revista Para El Análisis de La Cultura y El Territorio*, 19, 38–53. <<https://revistas.uca.es/index.php/periferica/article/view/4902>>
- VIÑARÁS ABAD, M., CAEROLS MATEO, E. (2016). “#5Museos: un caso de éxito sobre la oportunidad de las redes sociales para generar engagement” en *Revista internacional de relaciones públicas*, 12(12), 169–190. <<https://doi.org/10.5783/RIRP-12-2016-10-169-190>>