

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Facultad de Ciencias Sociales y Humanísticas

Eficiencia del gasto en publicidad analizado a través de datos en Twitter para empresas del sector comercio en el periodo 2018.

PROYECTO INTEGRADOR

Previo la obtención del Título de:

Economía con Mención en Gestión Empresarial

Presentado por:

ALEX ENRIQUE AMAGUAYA PACALLA

JOHNNY ENRIQUE CAJAPE TUBAY

GUAYAQUIL - ECUADOR

Año: 2019

DEDICATORIA

El presente proyecto se lo dedicamos a Dios, a nuestras familias y amigos que han sido pilar fundamental durante estos años de estudio.

AGRADECIMIENTOS

Al Econ. Juan Carlos Campuzano por impartir su conocimiento y acompañarnos en el proceso de titulación

A nuestra Alma mater la Escuela Superior Politécnica del Litoral por el conocimiento que nos ha brindado durante los años de estudio.

A Dios, que me ha brindado las fuerzas necesarias para poder lograr mis objetivos.

A mis padres, quienes me han guiado y formado durante toda la vida.

A la Dra. Carmen Vaca, quien ha sido una gran consejera tanto en lo profesional como en lo personal.


- Alex A.

A mi familia y amigos...

- Johnny C.

DECLARACIÓN EXPRESA

"Los derechos de titularidad y explotación, nos corresponde conforme al reglamento de propiedad intelectual de la institución; Amaguaya Pacalla Alex Enrique y Cajape Tubay Johnny Enrique damos nuestro consentimiento para que la ESPOL realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual"

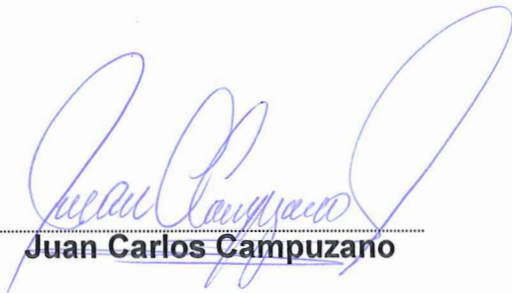


Alex Amaguaya



Johnny Cajape

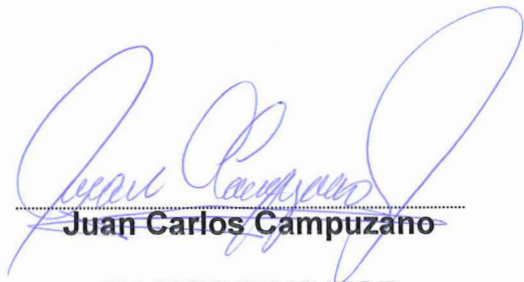
EVALUADORES



~~Juan Carlos Campuzano~~

Juan Carlos Campuzano

PROFESOR DE LA MATERIA



~~Juan Carlos Campuzano~~

Juan Carlos Campuzano

PROFESOR TUTOR

RESUMEN

Posterior a la llegada de las redes sociales y la interconexión que provocó entre sus usuarios, la generación de negocios dentro de estas plataformas resultó inevitable, el principal de ellos la venta de publicidad, en Ecuador la venta de publicidad a través de redes sociales asciende a un monto superior a los treinta millones de dólares solo para empresas del sector comercio por tanto, el presente trabajo analiza la eficiencia del gasto en publicidad mediante la condición de Dorfman-Steiner para ochenta y ocho empresas del sector comercio durante el año 2018, generando indicadores de gestión para el marketing digital (KPI's) empleando técnicas de Machine Learning para el procesamiento de datos provenientes de la red social twitter y relacionándolos con los resultados financieros de estas empresas durante el mismo periodo, a través de una regresión lineal múltiple. En el análisis realizado se encontró un efecto significativo por parte de los indicadores hacia la condición de Dorfman-Steiner para empresas con poca cantidad de tweets, los mayores efectos encontrados se dieron a través de la interacción entre los indicadores concluyendo que, para reducir el nivel de gasto en publicidad se debe apuntar a la propagación y popularidad del contenido que se publica, teniendo en cuenta la calidad del contenido que se divulga.

Palabras Clave: Dorfman-Steiner, KPI's, métricas, machine learning, twitter.

ABSTRACT

After the arrival of social networks and the interconnection that it caused among its users, the generation of businesses within these platforms was inevitable, the main one of them selling advertising, in Ecuador the sale of advertising through social networks amounts to a amount exceeding thirty million dollars only for companies in the commerce sector, therefore, this paper analyzes the efficiency of advertising spending through the condition of Dorfman-Steiner for eighty-eight companies in the commerce sector during 2018, generating indicators of Management for digital marketing (KPI's) using Machine Learning techniques for the processing of data from the social network twitter and relating them to the financial results of these companies during the same period, through a multiple linear regression. In the analysis performed, a significant effect was found by the indicators towards the Dorfman-Steiner condition for companies with a small number of tweets, the greatest effects found were given through the interaction between the indicators concluding that, to reduce the level Advertising spending should be aimed at the propagation and popularity of the content that is published, taking into account the quality of the content that is disclosed.

Keywords: Dorfman-Steiner, KPI's, metrics, machine learning, twitter.

ÍNDICE GENERAL

EVALUADORES.....	5
RESUMEN.....	I
<i>ABSTRACT</i>	II
ÍNDICE GENERAL	III
ABREVIATURAS.....	V
ÍNDICE DE FIGURAS	VI
ÍNDICE DE TABLAS.....	VII
CAPÍTULO 1.....	1
1. Introducción	1
1.1 Antecedentes	2
1.2 Descripción del problema	3
1.3 Preguntas de investigación	3
1.4 Justificación del problema	4
1.5 Objetivos	4
1.5.1 Objetivo General	4

1.5.2	Objetivos Específicos.....	4
1.6	Marco teórico.....	5
1.6.1	Key Performance Indicators (KPI's).....	6
1.6.2	Condición de Dorfman-Steiner.....	19
CAPÍTULO 2.....		10
2.	Metodología.....	10
2.1	Dataset.....	10
2.2	Variables.....	11
2.2.1	Pre-procesamiento de datos.....	11
2.2.2	Descripción de etapas de pre-procesamiento de texto de tweets.....	12
2.3	Descripción de variables.....	13
2.4	Modelos.....	17
2.4.1	Modelo de Clasificación SVMs.....	17
2.4.2	Modelo de Regresión.....	18
CAPÍTULO 3.....		21
3.	Resultados Y ANÁLISIS.....	21
3.1	Análisis de correlación.....	21
3.2	Análisis de regresión.....	22
CAPÍTULO 4.....		26
4.	Conclusiones Y Recomendaciones.....	26
4.1	Conclusiones.....	26
4.2	Recomendaciones.....	27
Bibliografía.....		28
APÉNDICES.....		¡Error! Marcador no definido.

ABREVIATURAS

ML	Machine Learning
SVMs	Support Vector Machines
SUPERCIAS	Superintendencia de Compañías, Valores y Seguros
API	Application Programming Interface
URL	Uniform Resource Locator
KPI's	Key Performance Indicators

ÍNDICE DE FIGURAS

Figura 2.1: Frecuencia acumulada del número de tweets durante el 2018	14
Figura 2.2: Cantidad de retweets y favoritos durante el 2018	15
Figura 2.3: Cantidad de tweets durante el 2018	15
Figura 2.4: Ranking de empresas acorde al total de tweets durante el 2018	16
Figura 3.1: Correlación entre las variables (gráfico de calor)	22
Figura 3.2: Histograma de la cantidad de Tweets	23

ÍNDICE DE TABLAS

Tabla 1.1. KPI de publicidad.....	7
Tabla 2.1: Estadísticas descriptivas de las variables.....	17
Tabla 3.1 Resultado de la regresión.....	24

CAPÍTULO 1

1. INTRODUCCIÓN

Existe una relación positiva prácticamente indiscutible entre la publicidad y las ventas, fenómeno que ha sido estudiado en incontables ocasiones, sin embargo, en el actual contexto tecnológico con el auge y apogeo de medios de comunicación masivos entre estas las redes sociales, el abanico de sitios web donde es posible publicitar a bajos costos se incrementa considerablemente, llevando a cuestionamiento la relación positiva antes mencionada y si los resultados de las ventas realmente justifican su gestión.

Según datos publicados de la Superintendencia de Compañías Valores y Seguros (Supercias) los ingresos reportados para las 100 empresas más grandes del sector comercio para el año 2018 ascendió a veintitrés mil millones de dólares, representando el 48.5% del monto acumulado por el sector, mientras que el gasto de publicidad por parte de estas 100 empresas superaba los trescientos millones de dólares (Supercias, 2019); de este monto cerca del 10% corresponde a gasto de publicidad realizado en redes sociales.

Resulta complejo atribuir una venta específica a la gestión realizada en publicidad y se transforma en una tarea aún más compleja cuando se trata de publicidad en redes sociales, por tanto las empresas utilizan instrumentos de control diseñados exclusivamente para el marketing digital conocidos como Key Performance Indicators (KPI's) que brindan una perspectiva del desarrollo de sus campañas publicitarias o Social Media Profile, estos indicadores nos permiten observar una relación con las ventas, no obstante mediante el uso de herramientas de procesamiento de datos como Machine Learning (ML) es posible definir una relación basado en la información producida por la interacción de los usuarios (C. Valencia, 2019).

Sin duda el desarrollo de un estudio que demuestre la existencia de una relación entre la publicidad en redes sociales y las ventas, de tal manera que brinde

información para realizar un mejor direccionamiento del gasto de publicidad para una empresa y que esta pueda generar mayores beneficios (Del Alcázar Ponce, 2019).

1.1 Antecedentes

Con la llegada de las redes sociales, compartir diferentes aspectos de la vida así como gustos y preferencias se ha vuelto una práctica común, comprendiendo por red social a aquellas plataformas digitales que permiten la obtención de un perfil propio y a la vez modificable con acceso y conexión a otros usuarios, además con la opción de compartir contenido (Boyd, 2008).

Para el año 2018, el número de ecuatorianos con acceso a internet y presencia en alguna red social ascendía a 13,8 millones de usuarios activos, por tanto, resulta debatible no considerar la información proveniente de este tipo de plataformas para observar desde otra perspectiva aspectos claves que resultan determinantes en el desarrollo de un negocio o empresa (Morán Q. & Cañarte R., 2017).

Malte y otros (2015) en su trabajo “¿Qué impulsa el éxito publicitario en Facebook?” realizaron un análisis de los diferentes tipos de interacciones con el volumen de ventas, considerando para estudio la página oficial de Facebook de una de las principales empresas alemanas dedicadas a la venta al por menor. Posteriormente hicieron una recolección mensual de datos durante 12 periodos, llegando a la conclusión de que muchos de los usuarios están dispuestos a realizar sus propias búsquedas de información sin necesidad de que este sea recomendado por un tercero, y en el peor de los casos que este llegue de forma invasiva a su buzón de mensajes en la misma aplicación, además concluyeron que la interacción de comentarios o reacciones están ligadas de forma positiva a las ventas de la empresa.

Por otra parte, Ramón S. y otros (2017) en su investigación “Comprendiendo el entorno del marketing digital con KPI's y análisis web”, analizaron diversos indicadores de gestión del marketing digital que las empresas deben comprender, y a la vez cuales de estos deben ser administrados de mejor forma para lograr un

aumento de efectividad en las campañas de publicidad a realizarse en redes sociales, y finalizaron con la propuesta de formulación de nuevos indicadores.

1.2 Descripción del problema

Según, Power Data en el año 2017 (empresa dedicada al análisis de datos) menciona que del total de volumen de datos que se genera en la actualidad, solo un 20% son datos numéricos y el porcentaje restante no es numérico que corresponde a blogs, redes sociales, videos o imágenes, entre otras. Hoy en día el uso y análisis de datos recolectados en internet principalmente de redes sociales, permite obtener mayor información de los consumidores y de las empresas.

Esta información es generada mediante la interacción de los usuarios con las páginas o post publicitarios de una empresa y que es sugerida por la misma red social en base a la actividad o por patrocinio de la cuenta, dicho patrocinio o publicidad tiene un costo relativamente bajo en comparación con medios tradicionales, además que la forma en que este llega al interesado se basará en la exposición que sea deseada, empezando con una tarifa promedio en las diferentes redes sociales de 0,55 ctvs de dólar para que esta llegue al perfil de 1000 usuarios.

Sin embargo esta inversión no garantiza que usuario interactúe o muestre un interés en el contenido del mismo; como lo indica Malte en su investigación, los usuarios reaccionan de forma negativa cuando un tipo de publicidad aparece de forma masiva, ya sea en su perfil o buzón de correos mientras este se encuentra en su momento de ocio, convirtiendo esta publicación en un gasto innecesario incurrido por parte de las compañías, repitiendo un escenario similar con otros usuarios y volviéndolo un gasto poco eficiente.

1.3 Preguntas de investigación

¿El gasto de publicidad en redes sociales se encuentra en un nivel óptimo?

¿Los datos de redes sociales constituyen una fuente de información para la toma de decisiones en el área de publicidad?

1.4 Justificación del problema

La presente investigación busca generar una metodología alterna en comparación a la que actualmente utilizan las empresas del sector comercial para la gestión del gasto de publicidad en redes sociales, metodología que implementará herramientas de machine learning (ML) para la generación de un indicador capaz de medir el nivel de aceptación o afinidad de los consumidores con la publicidad generada, y que posteriormente se relacionará con la proporción publicidad - ingresos de las empresas.

Dicha proporción que luego será puesta bajo el análisis de la condición Dorfman-Steiner, con el propósito de determinar el nivel óptimo de publicidad a invertir, esto sin duda servirá como un instrumento de decisión o control para evitar un gasto excesivo en publicidad que no genere un retorno significativo para las empresas. Con esto será posible relacionar distintas fuentes de información y formular estrategias que contemplen un panorama donde se tome en cuenta la opinión de los clientes.

1.5 Objetivos

1.5.1 Objetivo General

Analizar los KPI's del marketing digital utilizando datos de la red social Twitter de las empresas del sector comercio para el periodo 2018 y su relación con la proporción de gasto de publicidad - ingresos (condición de Dorfman - Steiner), con el propósito de tomar mejores decisiones financieras en el área de publicidad.

1.5.2 Objetivos Específicos

- Plantear una métrica de sentimiento basado en datos recolectados de Twitter para las empresas del sector comercio del Ecuador.

- Estimar el efecto entre los KPI's del marketing digital, la métrica de sentimiento y el ratio gasto de publicidad-ingresos (condición de Dorfman – Steiner).
- Analizar e interpretar los resultados obtenidos al relacionar los KPI's del marketing digital, la métrica de sentimiento con el ratio gasto de publicidad-ingresos (condición de Dorfman – Steiner).

1.6 Marco teórico

El análisis de datos provenientes de redes sociales como una fuente de información que proporcione variables predictoras, es un tema que se ha desarrollado con mayor ímpetu durante los últimos años, considerando que la tasa de usuarios se ha incrementado de forma considerable llegando al cierre de 2018 un total de tres mil cuatrocientos ochenta y cuatro millones de usuarios activos, dicha cifra corresponde al 45% de la población para el mismo año (Hootsuite, 2019). Por tanto, el volumen de datos sin duda se ha incrementado dando paso a un mayor soporte al análisis realizado.

Considerando la diversidad de información que es compartida a través de los usuarios, por lo cual su aplicabilidad o empleo en diversos estudios es muy amplio, una muestra de los estudios realizados es el de Zhang, Fuehlers, & Gloor (2011) donde detallan uno de los primeros intentos de utilizar datos proveniente de la red social Twitter como un predictor de los indicadores del mercado de valores, para ello los autores recolectaron las publicaciones de los perfiles de empresas en twitter en un periodo de seis meses, posterior a ello de forma aleatoria escogieron como muestra el 1% del total de tweets recolectados que luego serían tratados y analizados.

Utilizando métricas de sentimientos se calculó la esperanza y el miedo colectivo de cada empresa para posteriormente realizar la correlación con los índices bursátiles, en el cual encontraron una correlación negativa y significativa entre los tweets tomados como muestra y cada uno de los índices. Además a manera de comprobación realizaron el análisis con el índice VIX (índice de volatilidad de

opciones), en el que encontraron una correlación positiva y significativa (este índice es considerado una medida de las expectativas del mercado bursátil que coincide fuertemente con las alzas y bajas del mercado corriente), lo que indicaba que las métricas de sentimientos proveniente de twitter sirven como predictores en el mercado de valores y a la vez brindan información de como se desempeñaría este mercado en el futuro.

Chanchal Bahadur (2016) en su trabajo “Rendimiento de las finanzas de las empresas según el impacto de redes sociales” muestra la relación creciente entre el uso de las redes sociales y la apreciación de las marcas, y también describe como los cambios en la apreciación de los consumidores deberían verse traducido como una mejora en el desempeño financiero de las empresas. Para lo cual realizó un estudio para ciento ochenta empresas correspondientes a 10 industrias diferentes, partiendo con una hipótesis de que al clasificar a las empresas con mayor número de publicaciones en redes sociales estas tendrían retornos más altos en comparación a las empresas con una baja participación en redes sociales. En este trabajo utilizó el modelo de cuatro factores de Fama French, en el que no se encontró diferencias significativas entre en los retornos reportados por estas empresas. Como segunda opción planteó la posibilidad de predecir el flujo de caja trimestral para los grupos de empresas, empleando un modelo de regresión multivariante de series de tiempo Lorek y Willinger, para lo cual implementó como variable adicional una métrica construida con los comentarios realizados por los usuarios para el grupo de empresas, y además utilizó como método de contraste una predicción empleando datos de panel, sin embargo no se encontró una mejora en su predicción para ninguno de los dos casos planteados.

1.6.1 Key Performance Indicators (KPI's)

Según Neiger las redes sociales dentro del análisis de marketing digital tienen cuatro propósitos que se encuentran ligados a generar información sobre el mercado, dichos propósitos son: establecer una marca, difundir información crítica, generar un mayor alcance entre la audiencia y crear un compromiso con el público

u otras empresas (Neiger, y otros, 2012), y cada uno de estos propósitos a su vez mantiene una relación con los distintos indicadores de gestión.

Los Key Performance Indicators (KPI's) de publicidad permiten registrar o medir la actividad generada a partir de las acciones realizadas dentro de las redes sociales, entre los KPI's de publicidad más utilizados se encuentran los siguientes:

Tabla 1.1. KPI de publicidad

KPI's	Definición	Medida
Insights	Comentarios de los consumidores en redes sociales	Número y tipos de sugerencias.
Exposición	Número de veces que el contenido ha sido visto	Vistas, vistas de videos. Número de clics, comentarios, valoraciones, compartidos.
Alcance	Número de personas que tienen contacto con la publicación.	Número de seguidores de la página. Número de personas que se suscriben y dejan de suscribirse. Demografía de los seguidores/suscriptores. Ratio de crecimiento de seguidores
Compromiso (Bajo)	Número de personas que reconocen o prefieren el contenido	Calificaciones. Número de Likes o dislikes. Frecuencia de favoritos.

Compromiso (Medio)	Número de personas que participan activamente, compartiendo y usando el contenido para influir en otros	Número de publicaciones por usuario. Contenido generado por el usuario. Comentarios en publicaciones. Frecuencia de discusiones y nuevos temas. Descargas. Menciones.
Compromiso (Alto)	Número de personas que participan en eventos fuera de línea (que pueden ser además de una actividad en línea continua) como consumidor o como socio del programa, voluntario o patrocinador	Número de personas que se registran para recibir información al respecto. Participación en eventos como voluntarios o patrocinadores. Número de personas que asisten a eventos off-line como participantes. Número de participantes satisfechos.

Adaptado de: Neiger y otros (2012)

En un escenario en donde las empresas participan en un mercado muy competitivo y donde los clientes se encuentran adecuadamente informados al momento de elegir entre distintas compañías, se impulsa a que las empresas posean una mayor cantidad de información relevante con el objetivo de poder ejecutar estrategias comerciales efectivas y así poder fidelizar a los clientes. Actualmente, existen técnicas que permiten analizar patrones de conducta, nichos de mercado y muchos otros tipos de información no trivial, para lo cual se utilizan de sofisticados modelos que combinan métodos estadísticos y machine learning (ML), entre las herramientas mayormente utilizadas en el campo del aprendizaje automático se encuentra Support

vector machines (SVMs) que tienen como objetivo separar un conjunto de datos de forma lineal o no lineal con características similares de información, la misma que se analiza mediante la optimización de una función de costo.

Así también, autores como Cecchini y Aytug (2010) utilizaron un clasificador bayesiano empleando la metodología de Support vector machines (SVMs) para identificar factores que afecten en la percepción de la marca de 15 empresas provenientes de diferentes industrias, y por consiguiente analizaron si estas mejoraban su desempeño financiero. Los autores realizaron monitoreos constantes a sus páginas de redes sociales, encontrando finalmente una baja relación entre los factores y el desempeño financiero para empresas pertenecientes a la industria telefónica, ordenadores y hardware, mientras que las empresas de la industria de juguetería y ropa deportiva tenían una relación moderada y alta.

Los resultados obtenidos por parte de Cecchini y Aytug concuerdan fuertemente con los casos de estudios realizados por Mikolaj y Ryan (2014) de la Harvard Business School, donde detalla como Nike® había cambiado su estrategia de marketing y publicidad pasando de medios tradicionales a las redes sociales, en el que se realizó desembolsos de dinero que superaban los ochocientos millones USD, dicha estrategia tuvo como objetivo reducir cada vez el gasto en medios de prensa visual o impresa. Sin embargo, cabe resaltar que la estrategia de Nike en cuanto al desarrollo de marketing en los medios sociales se encuentra prácticamente a la par del surgimiento de las redes sociales como las conocemos hoy en día, y que el nivel de sus ventas a nivel global exige una fuerte inversión en publicidad para mantener un crecimiento y fidelización por parte de los consumidores.

CAPÍTULO 2

2. METODOLOGÍA

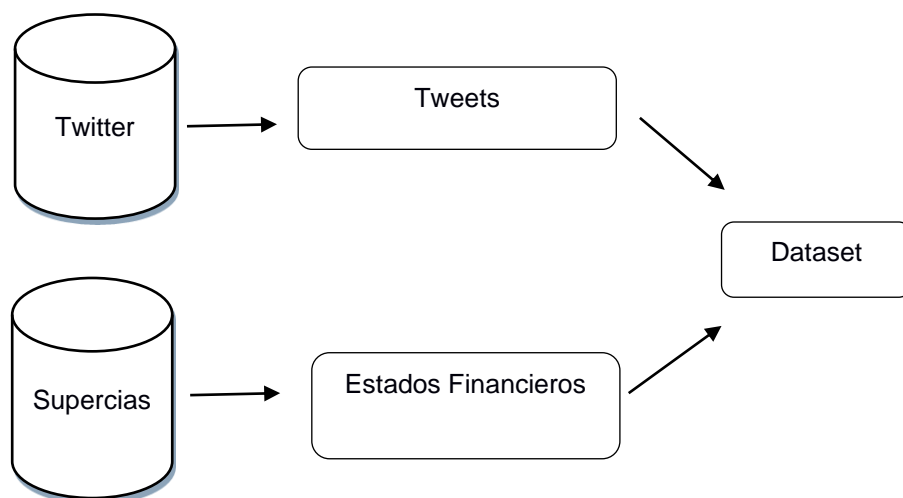
2.1 Dataset

El dataset construido tuvo dos fuentes de información, una de ellas fue la página web de la Superintendencia de Compañías Valores y Seguros (Supercias) y la red social Twitter. Para la recopilación de datos de las distintas fuentes se tomó en consideración ochenta y ocho empresas, las cuales fueron seleccionadas de acuerdo al ranking según el nivel de ingreso para las empresas del sector comercio durante el periodo 2018.

Para las ochenta y ocho empresas se extrajo los estados financieros, de los cuales se utilizó la cantidad de dinero destinado a publicidad y su nivel de ingreso por venta para el año 2018.

Mientras que, para la recolección de datos en twitter se hizo uso de la Application Programming Interface (API) de twitter que permitió acceder a la función de recolección de tweets con su respectiva información. Para este proceso se tomó en consideración únicamente los tweets que fueron publicados durante período 2018.

Estructura de recolección de datos



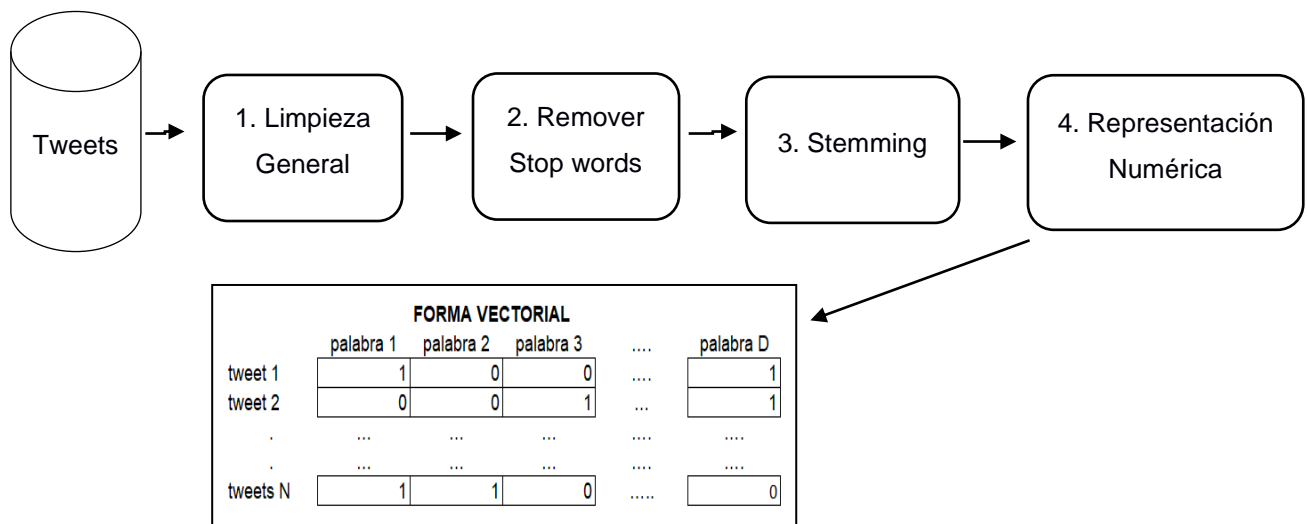
Elaborado: Autores

2.2 Variables

2.2.1 Pre-procesamiento de datos.

Para el uso de los datos de twitter se realizó un pre-procesamiento al texto de los tweets, con el objeto de convertir el texto a una representación numérica para su posterior uso en el cálculo de la métrica de sentimiento. A continuación se presenta la estructura usada para el pre-procesamiento de texto (Vaibhavi Patodkar, Sheikh I.R):

Estructura de pre-procesamiento de tweets



Elaborado: Autores

2.2.2 Descripción de etapas de pre-procesamiento de texto de tweets

Limpieza General.

En esta etapa se removi6 del texto:

- Signos de puntuaci6n
- Hashtags
- Direcciones URL
- Usuarios
- N6meros
- Emoticones

Remove Stop Words.

Se define como stop words o palabras vac6as a aquellas palabras que carecen de un significado, debido a que no aportan un mensaje a la estructura o contenido del texto, como por ejemplo:

- Art6culos
- Conjunciones
- Preposiciones
- Pronombres
- Adverbios
- Abreviaturas

Stemming.

Esta etapa consiste en reducir las palabras a su ra6z, con el objetivo de eliminar palabras repetitivas una vez empleado Stemming.

Representaci6n num6rica.

Para la representaci6n num6rica de texto se us6 t6cnicas de Word Embedding basado en frecuencia (TF-IDF), el cual consiste en asignar una palabra a un vector usando un diccionario.

2.3 Descripción de variables.

Con el dataset construido se construyeron 6 variables, de las cuales 3 son variables en base a los datos de Twitter tales como número de retweets escalados, favoritos escalados y la métrica de sentimiento. Y por el lado financiero se obtuvo el monto de dinero destinado a publicidad, el nivel de ingreso por venta y la condición de Dorfman-Steiner, a continuación se muestra una descripción de las variables utilizadas y su respectiva formulación:

- **Retweets escalados:** Cantidad promedio de retweets por tweet que tuvo firma durante el periodo 2018 (Malte Brettel, Jens Reich, José Gavilanes y Tessa Flatten, 2015).

$$Retweet_escalado = \frac{\text{cantidad total de retweets}}{\text{cantidad total de tweets}}$$

- **Favoritos escalados:** Cantidad promedio de favoritos por tweet que tuvo firma durante el periodo 2018 (Malte Brettel, Jens Reich, José Gavilanes y Tessa Flatten, 2015).

$$Favorito_escalado = \frac{\text{cantidad total de favoritos}}{\text{cantidad total de tweets}}$$

- **Métrica de sentimiento:** Ratio entre la diferencia de la cantidad total de tweets positivos menos los tweets negativos, dividido para la cantidad de tweets (Stefan Feuerriegel, Julius Gordon, 2017).

$$Métrica_Sent = \frac{\text{Total de tweets positivos} - \text{Total de tweets negativos}}{\text{cantidad total de tweets}}$$

- **Gasto de Publicidad:** Cantidad total de dinero destinado a publicidad por una firma en el año 2018.
- **Ingreso por venta:** Cantidad total de dinero obtenido producto de las ventas por una firma en el año 2018.

- **Condición de Dorfman-Steiner:** Ratio entre el gasto en publicidad y el ingreso por venta.

$$\text{Cond. Dorfman - Steiner} = \frac{\text{Gasto de publicidad}}{\text{Ingreso por venta}}$$

2.1.1. Estadística descriptiva

A continuación se muestra un gráfico relacionado a la cantidad total de tweets obtenidos de twitter, en el cual la Figura 2.1 se observa la frecuencia acumulada de los tweets para cada mes, logrando obtener al final de los 12 meses un total de cuarenta y nueve mil seiscientos ochenta y cinco tweets para las ochenta y ocho empresas seleccionadas en la presente estudio.

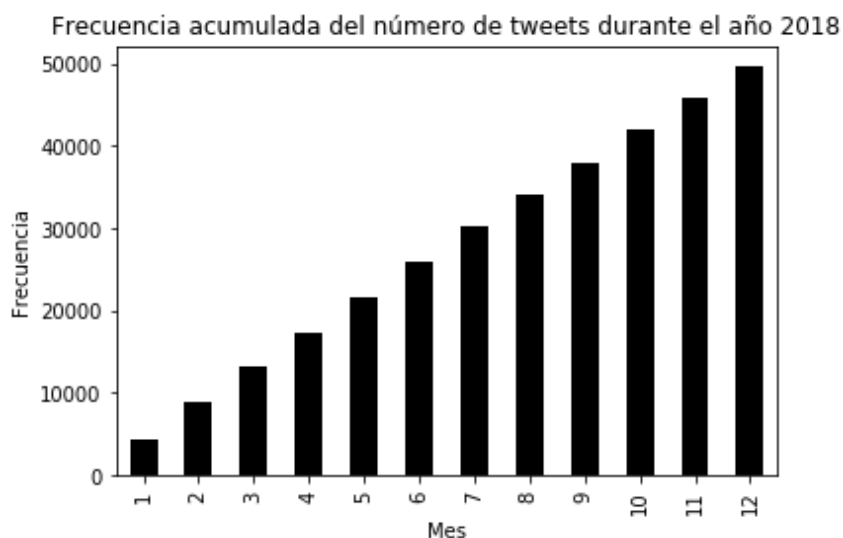


Figura 2.1: Frecuencia acumulada del número de tweets durante el 2018

Elaborado por: Autores

Además, en la figura 2.2 se expone la cantidad total de favoritos y retweets obtenidos por semana para el total de las ochenta y ocho empresas durante el año 2018, en el cual se observa que la cantidad de favoritos es superior a los retweets.

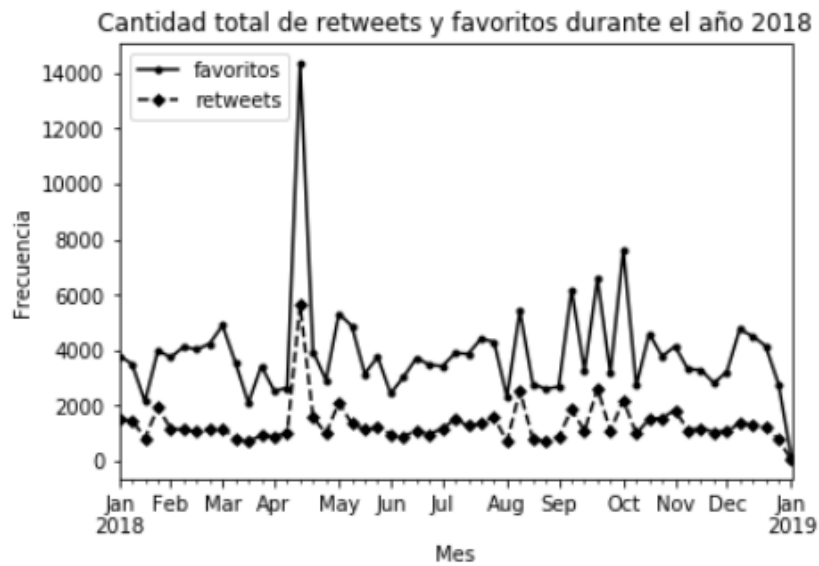


Figura 2.2: Cantidad de retweets y favoritos durante el 2018
Elaborado por: Autores

También se muestra en la figura 2.3 la cantidad total de tweets obtenidos por semana para las ochenta y ocho empresas durante el año 2018, en el cual se visualiza que el primer y último trimestre fueron los periodos de tiempo con mayor frecuencia de tweets.

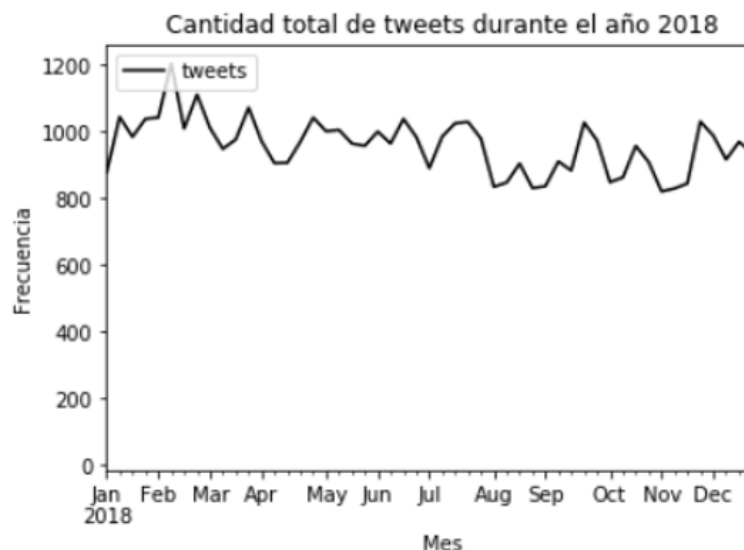


Figura 2.3: Cantidad de tweets durante el 2018
Elaborado por: Autores

Asimismo en la figura 2.4 se muestra un gráfico en el cual se detalla cuáles fueron las empresas que tuvieron mayor cantidad de tweets para el periodo 2018. Este gráfico muestra el ranking de las 15 empresas que tuvieron mayor actividad en twitter del total de ochenta y ocho empresas seleccionadas para nuestro análisis.

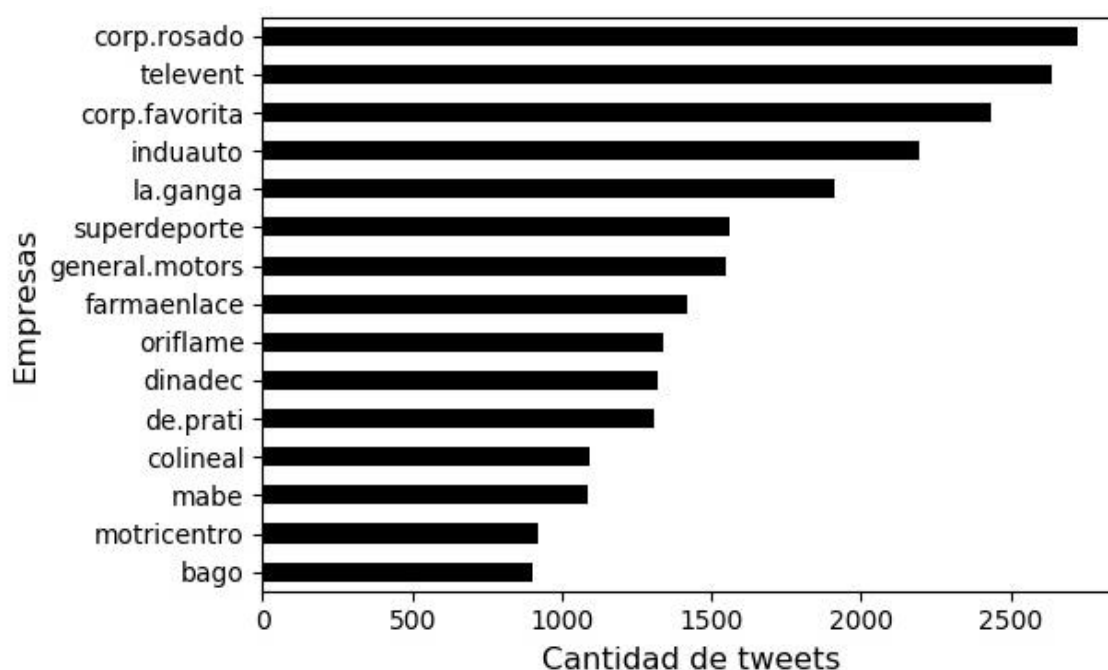


Figura 2.4: Ranking de empresas acorde al total de tweets durante el 2018
Elaborado por: Autores

Una vez visualizado los datos totales de twitter, se procedió a agruparlos por empresa y realizar el respectivo cálculo de las métricas para su posterior análisis, a continuación se muestra una tabla descriptiva sobre las variables empleadas:

Tabla 2.1 Estadísticas descriptivas de las variables

	Cantidad	Media	Desv. Est.	Min.	Máx.
Favoritos_escalado	88	3.52	6.05	0.08	32.12
Retweets_escalado	88	1.35	2.40	0.02	14.90
Cantidad de tweets	88	564.60	574.52	27	2726
Gasto en publicidad(miles USD)	88	2314.59	4033.26	1	25643
Mét. De Sentimiento	88	0.64	0.23	-0.74	0.97
Ingreso por venta (miles USD)	88	181047.98	305348.71	13387	2089984
Condición de Dorfman-Steiner	88	0.02	0.02	2.195 e-05	0.09

Elaborado por: Autores

2.4 Modelos.

2.4.1 Modelo de Clasificación SVMs.

El modelo escogido para realizar la tarea de clasificación de la polaridad de un tweet fue Support Vector Machines (SVMs) debido a su gran aplicación para problemas de clasificación como de regresión. El objetivo de este modelo es encontrar un hiperplano en un espacio dimensional N que clasifique de forma clara los datos, el cual presenta la siguiente formulación:

Función de Costo a Minimizar.

$$c(x, y, f(x)) = \min_w \lambda \|w\|^2 + \sum_{i=1}^n (1 - y * f(x))$$

$$f(x) = w^T x + b$$

Donde:

- X: son las variables de entrada
- Y: categorías asignadas.
- λ : hiperparámetro de regularización
- W: pesos/parámetros del modelo
- f(x): predicción del modelo
- b: sesgo

2.4.2 Modelo de Regresión.

Debido a la periodicidad anual de los datos de gasto en publicidad y del nivel de ingreso por venta, se procedió a calcular para el resto de variables un valor en el mismo periodo de tiempo.

Para evaluar efecto de los indicadores en base a datos de twitter frente a la condición de Dorfman-Steiner, se planteó un modelo de regresión lineal múltiple estimado mediante mínimos cuadrados generalizados, el cual se lo realizó de esta manera debido a que se tomó en consideración el problema de heterocedasticidad (Herrington y Dempsey, 2005).

En la especificación del modelo de regresión planteado no solo se buscó evaluar el efecto que tiene cada métrica por separado, también se evaluó la interacción entre las métricas frente a la condición de Dorfman-Steiner (Malte Brettel, Jens Reich, José Gavilanes y Tessa Flatten, 2015). A continuación se presenta la especificación del modelo planteado:

Cond. de Dorfman – Steiner

$$\begin{aligned} &= \beta_0 + \beta_1 \text{Retweets_escalado} + \beta_2 \text{Favoritos_escalado} \\ &+ \beta_3 \text{Mét. de Sentimiento} + \beta_4 (\text{Retweets_escalado} * \text{Favoritos_escalado}) \\ &+ \beta_5 (\text{Retweets_escalado} * \text{Mét. de Sentimiento}) \\ &+ \beta_6 (\text{Mét. de Sentimiento_escalado} * \text{Favoritos}) + \varepsilon \end{aligned}$$

Donde:

- Retweets_escalado: es la cantidad de retweets promedio para una empresa.
- Favoritos_escalado: es la cantidad de favoritos promedio para una empresa.
- Métrica De Sentimiento: es la métrica de sentimiento para una empresa.
- Cond. de Dorfman-Steiner: Ratio entre el gasto en publicidad y el ingreso por venta.
- ε : es el término de error del modelo.

Además para evaluar la relación entre los indicadores frente a la condición de Dorfman-Steiner se realizó un análisis de correlación entre las variables antes mencionadas, medidas a través del coeficiente de correlación de Spearman (Malte Brettel, Jens Reich, José Gavilanes y Tessa Flatten, 2015), presentado la siguiente formulación para el cálculo:

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

Donde:

- d: diferencia entre rangos.
- n: número de datos.
- r: Coeficiente de correlación de Spearman.

2.4.3 Condición de Dorfman-Steiner

La condición de Dorfman-Steiner calcula el gasto óptimo de publicidad para el precio de un producto mediante una regla que relaciona los ingresos marginales de publicidad y la elasticidad precio de la demanda del producto. Para esto el supuesto a considerar es que la firma opera como un monopolio.

A continuación se muestra el procedimiento del cálculo partiendo desde la función de beneficio:

$$\Pi(P, A) = P(Q(P, A)) - TC_M(Q(P, A)) - A$$

Donde:

- $Q = Q(P, A)$ es la cantidad demandada del producto por periodo de tiempo a la firma.
- P = Precio por unidad demandada
- A = Gasto de publicidad total (en dólares)
- $TC_M = TC_M(Q)$ Costo total de manufactura (en dólares)

Derivando la función de beneficio en función del Precio (P) y gasto de publicidad (A) se obtienen las siguientes condiciones de primer orden:

$$\frac{d\pi}{dP} = Q(P, A) + P \frac{dQ}{dP} - \frac{dTC_M}{dQ} * \frac{dQ}{dP} = 0$$

$$\frac{d\pi}{dA} = P \frac{dQ}{dA} - \frac{dTC_M}{dQ} * \frac{dQ}{dA} - 1 = 0$$

Empleando la ecuación $\frac{d\pi}{dA}$, la definición de costo marginal de manufactura $\frac{dTC_M}{dQ} = MC_M$ y la elasticidad precio de la demanda (en valor absoluto) $\eta = -\frac{dQ}{dP} * \frac{P}{Q}$ y

realizando una operación entre variables se obtiene $P * \frac{\partial Q}{\partial A} = \frac{P}{P-MC_M} = \eta$

Con este último factor se obtiene la condición de Dorfman Steiner que se expresa como:

$$\frac{A}{p * q} = \frac{\theta}{\eta}$$

Donde:

Θ : representa la elasticidad de la publicidad frente a la producción

η : representa la elasticidad precio de la demanda

Según la teoría a medida que los precios de los productos se conviertan en inelásticos el nivel de gasto de publicidad aumenta considerablemente. Haciendo que la condición DS crezca y por consiguiente el gasto no sea eficiente.

CAPÍTULO 3

3. RESULTADOS Y ANÁLISIS

3.1 Análisis de correlación.

A través del análisis de correlación en la figura 3.1 se encontró que existe una relación inversa entre la condición de Dorfman-Steiner y la métrica de sentimiento el cual tuvo un valor de -0.26 , el resultado de esta relación nos indicó que el sentimiento que se expresa en los tweets es importante para los consumidores debido a que permite reducir la condición de Dorfman-Steiner, lo cual implica efectos en la reducción del nivel de gasto. A la vez se encontró que existe una relación negativa entre la métrica de sentimiento y el gasto en publicidad lo cual nos indica que mientras los tweets expresen un sentimiento positivo hacia las personas el gasto en publicidad se reducirá, por lo cual esto se reflejará en la condición de Dorfman Steiner.

Además se encontró una relación positiva tanto para la cantidad promedio de retweets como la cantidad promedio favorita con la condición de Dorfman-Steiner, obteniendo valores de 0.21 y 0.26 respectivamente. Esta relación indicó que mientras los tweets generen más interacción con el consumidor a través de los retweets y favoritos, esto provoca que exista un mayor nivel de ingreso y a su vez más gasto en publicidad debido a que se destinan más recursos con el fin de obtener más interacción con los tweets publicados.

También se halló una fuerte relación positiva de 0.87 entre la cantidad de retweets y favoritos, lo cual indica que si los tweets generan mayor popularidad a través de la cantidad promedio de favoritos, las personas tenderán a compartir mayormente los tweets logrando así una mejor propagación de la publicidad a través de ellos.

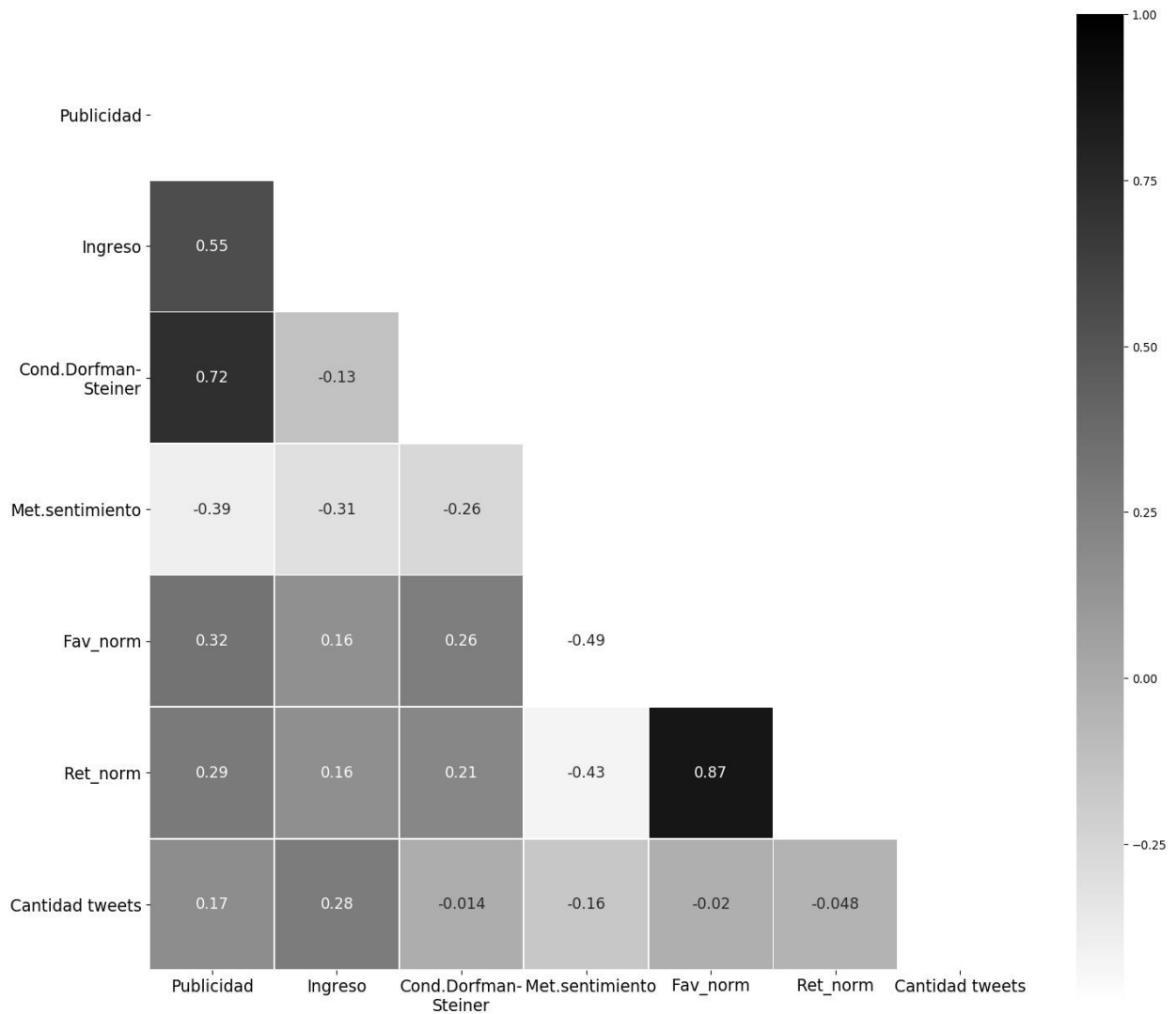


Figura 3.1: Correlación entre las variables (gráfico de calor)
 Elaborado por: Autores

3.2 Análisis de regresión.

Se realizó una separación en cuartiles a las 88 empresas en base a la cantidad de tweets publicados, esta separación se la realizó debido a que se observó una distribución sesgada positivamente para la cantidad de tweets como se puede visualizar en la figura 3.2. Para el presente análisis se planteó la misma especificación del modelo presentado en el capítulo 2 para todos los grupos de empresas, a continuación en la tabla // se muestran los principales resultados obtenidos:

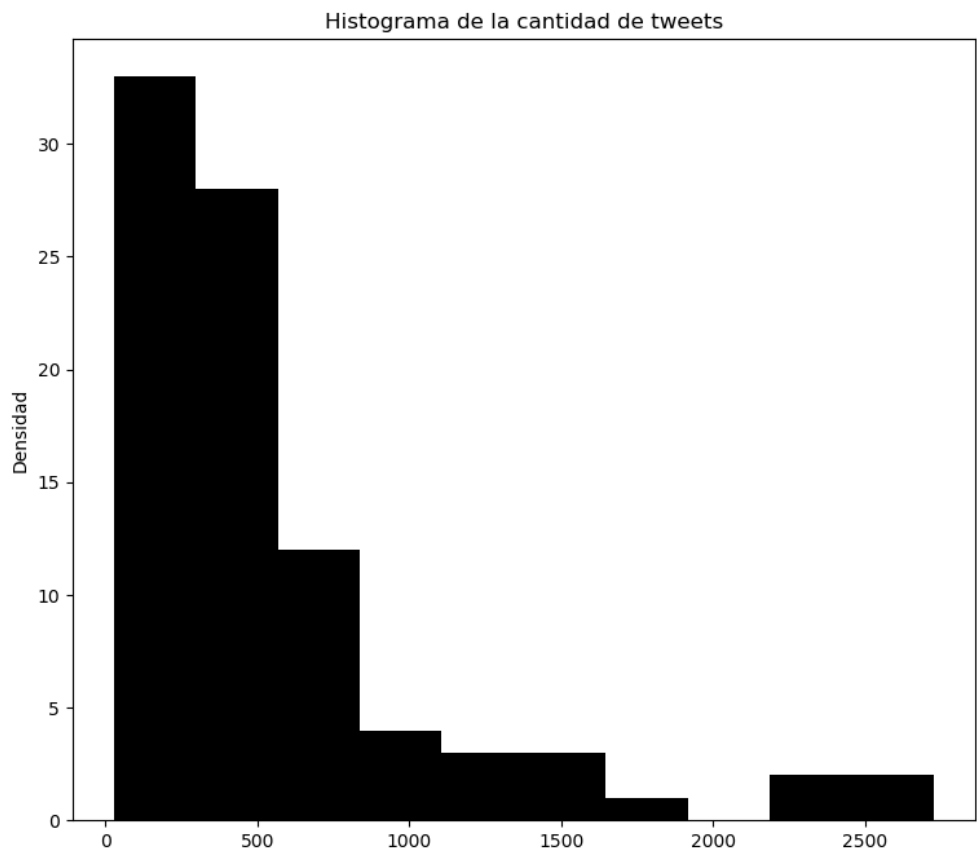


Figura 3.2: Histograma de la cantidad de Tweets
Elaborado por: Autores

Tabla 3.1 Resultado de la regresión

VARIABLES	(Modelo 1)	(Modelo 2)
	Cond_Dorfman_Steiner	Cond_Dorfman_Steiner
favorito_promedio	-0.0011 (0.0032)	-0.0487*** (0.0098)
retweet_promedio	0.0073 (0.0081)	0.1859*** (0.0330)
Sentimiento	-0.0032 (0.0089)	-0.0346 (0.0335)
favorito_Prom * retweet_Prom	-0.0002** (0.0001)	-0.0007 (0.0007)
favorito_Prom * Sentimiento	0.0024 (0.0050)	0.0785*** (0.0166)
retweet_Prom * Sentimiento	-0.0023 (0.0118)	-0.2975*** (0.0511)
Constante	0.0136** (0.0065)	0.0453* (0.0245)
Observaciones	88	22
R-squared	0.1133	0.6946

Errores estándar en paréntesis

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

Elaborado por: Autores

En el modelo de la tabla 3.1 correspondiente al total de empresas, solo se encontró un efecto significativo entre la interacción de las variables cantidad promedio de favoritos con la cantidad promedio de retweets cuyo valor fue de -0.0002.

En cambio para el modelo 2, correspondiente a las empresas ubicadas en el segundo cuartil (empresas con una cantidad de tweets entre ciento ochenta y dos y cuatrocientos treinta y dos) se halló más de una variable con efectos significativos para este grupo de empresas. Además se encontró que al aumentar en una unidad la cantidad promedio de retweets se esperaba que el valor de la condición de Dorfman-Steiner aumente en 0.19, lo cual da lugar a un efecto positivo entre estas variables. En cambio existió un efecto negativo por parte de la métrica de sentimiento de -0.03 pero este efecto no resultó significativo para este grupo de empresas.

Además se encontró efectos significativos cuando se realizó la interacción entre las variables, en el cual al aumentar en una unidad la interacción entre la cantidad

promedio de favoritos y la métrica de sentimiento producirá un aumento de 0.08 en la condición de Dorfman-Steiner. También resultó significativo la interacción entre la cantidad promedio de retweets y la métrica de sentimiento, pero esta interacción producirá una reducción de 0.30 en la en la condición de Dorfman-Steiner.

Para el presente trabajo de investigación y con los datos obtenidos el gasto de publicidad en redes sociales eficiente debe encontrarse en el punto 0,1576 resultado que procede de una elasticidad de la publicidad frente a la producción de 0.55 es decir que cambios porcentuales (1%) en gasto de publicidad genera un impacto en el 0.55 en los productos producidos; dichos en termino monetarios inversiones de 1 ctv en publicidad tiene una repercusión en 55 dólares de productos, mientras que la elasticidad precio de la demanda muestra que los precios en el mercado ecuatoriano son elásticos con un valor de 3.47 lo que explicaría porque la publicidad ha tenido un lento desarrollo en el país.

La forma en que la condición medida a través de los diferentes KPI's propuestos para el estudio alcancen el mismo nivel que el cálculo encontrado (0,1576) es mejorando la calidad del contenido sin incrementar el nivel de gasto mantenido actualmente.

CAPÍTULO 4

4. CONCLUSIONES Y RECOMENDACIONES

4.1 Conclusiones.

- Como se pudo observar la métrica de sentimientos propuesta analizó la interacción de los usuarios con las diferentes publicaciones realizadas por las empresas en twitter, clasificando esta interacción como positiva, negativa o neutra (utilizando la metodología de SVMs) y creando un indicador que se puede interpretar el grado de aceptación que tienen los usuarios con la publicidad de las empresas.
- La métrica de sentimiento resulta tener un efecto significativo cuando se realizó una interacción junto con otras variables, logrando concluir que dicho indicador propuesto tiene un efecto conjunto frente a la condición de Dorfman- Steiner.
- A través del modelo de regresión múltiple se buscó obtener el retorno y nivel de significancia que tiene cada una de los indicadores antes expuestos con la variable dependiente que en este caso corresponde a la condición de Dorfman-Steiner, obteniendo resultados poco significativos cuando se analizaron a todas las empresas en su conjunto. Sin embargo al desagregar en grupos por cuartiles de acuerdo a la cantidad de publicaciones, se observó que para empresas con una menor cantidad de contenido publicitario subido a sus cuentas se obtuvo resultados significativos.
- La cantidad promedio de retweets presentó uno de los mayores efectos equivalente a un aumento de 0.19 condición de Dorfman-Steiner, lo que indica

que a medida que una publicación consigue llegar a un mayor número de personas el gasto en publicidad se reduce y en cambio el nivel de ingreso por venta aumenta en mayor proporción.

- La relación Dorfman-Steiner busca la relación de gasto/ingreso óptimo que permita a las empresas una asignación correcta de recursos sin embargo, para este caso los resultados obtenidos permiten observar que la condición de gastos no se encuentra en equilibrio, ya que para los datos recolectados la condición se colocó en un ratio de 0.1576, mientras que ninguno de los intercepto de los diferentes escenarios no se encuentra en dicho nivel.

4.2 Recomendaciones

- Las redes sociales representan una fuente de información de la cual se puede aprovechar el volumen de datos que se producen en estas, el cual debe ser aprovechado con el objetivo de generar valor agregado y tomar mejores estrategias comerciales.
- Analizar más de una red social sería una muestra más representativa sobre la variedad de consumidores que existen en la actualidad, proveyendo con ello un mayor conocimiento sobre las características de estos.

BIBLIOGRAFÍA

- Boyd, D. &. (2008). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 210-230. doi:<http://dx.doi.org/10.1111/j.1083-6101.2007.00393.x>
- Brettel, M., Reich, J.-C., Gavilanes, J., & Flatten, T. (2015). What Drives Advertising Success on Facebook? An Advertising-Effectiveness Model Measuring the Effects on Sales Of “Likes” and Other Social-Network Stimuli. *Journal of advertising and research*, 162-175. doi:10.2501/JAR-55-2-162-175
- C. Valencia. (2019, Marzo 1). *Cámara Valencia*. Retrieved from <https://www.mastermarketing-valencia.com/comunicacion-publicidad/kpi-de-la-comunicacion-y-publicidad/>
- Chanchal, B. (2016, Agosto). *University of Iowa*. doi:<https://doi.org/10.17077/etd.wuio6qmv>.
- Del Alcázar Ponce, J. (2019). *Ecuador Estado Digital Ene/19*. Quito: Mentinno.
- Hootsuite. (2019). *Digital 2019*. Vancouver: Hootsuite & We are Social.
- Malte, B., Jens, R., Jose, G., & Tessa, F. (2015). What drives advertising success on facebook? *Journal of Advertising Research*, 162-175.
- Mikolaj, J. P., & Ryan, J. (2014). Estrategia social en Nike. *Harvard Business School*, 1-24.
- Morán Q., C., & Cañarte R., T. (2017). Redes sociales factor predominante en la fidelización del cliente. *Dominio de la Ciencias*, 519-532.
- Neiger, B., Thackeray, R., Van Wagenen, S., Hanson, C., West, J., Barnes, M., & Fagen, M. (2012). Use of Social Media in Health Promotion: Purposes, Key Performance Indicators, and Evaluation Metrics. *Health Promotion Practice*, 159-164.

- Ramón, J., Palos, P., & Cerdá, L. (2017). Understanding the Digital Marketing Environment with KPIs and Web Analytics. *Future Internet*, 2-13. doi:10.3390/fi9040076
- Supercias. (2019, Abril). *Superintendencia de Compañías Valores y Seguros*. Retrieved Junio 2019, from <https://appscvs.supercias.gob.ec/rankingCias/rankingCias.zul?id=G&tipo=5>
- UJAT. (2012). *Universidad Juárez Autónoma de Tabasco*. Retrieved from <https://sites.google.com/site/estadisticadm/b-organizacion-y-presentacon-de-datos/b-2-organizacion-y-presentacion-de-datos/b-2-3-datos-no-numericos>
- Zhang, X., Fuehlers, H., & Gloor, P. A. (2011). Predicting Stock Market Indicator Through Twitter. *Procedia - Social and Behavioral Sciences*, 55-62.