



EL MISIONERO DEL AGRO

INFLUENCIA DE LOS BOTS SOCIALES EN LA PERCEPCIÓN DE
MARCAS DE EMPRESAS ECUATORIANAS

INFLUENCE OF SOCIAL BOTS ON ECUADORIAN COMPANIES
BRANDS' PERCEPTION

Filiación:

Universidad Agraria del Ecuador – Facultad de Ciencias Agrarias –
Escuela de Computación e Informática

Autora:

Ing. Elke Jacqueline Yerovi Ricaurte, MSc.

Teléfono: 0986215720

eyerovi@uagraria.edu.ec; eyerovi@espol.edu.ec; elke.yerovi@hotmail.com

Coautores:

Ing. Pablo Andrés Calderón Castro

Teléfono: 0991740714

pacalder@espol.edu.ec; pablo.calderonc@ug.edu.ec

Ing. Mariuxi Ileana Tejada Castro

Teléfono: 0994439044

mtejada@uagraria.edu.ec; mtejada@espol.edu.ec; mariuxi_ileana@hotmail.com

Guayaquil - Ecuador

Fecha de presentación: 30/noviembre/2015

Fecha de aceptación: 18/diciembre/2015

INFLUENCIA DE LOS BOTS SOCIALES EN LA PERCEPCIÓN DE MARCAS DE EMPRESAS ECUATORIANAS

INFLUENCE OF SOCIAL BOTS ON ECUADORIAN COMPANIES BRANDS' PERCEPTION

Universidad Agraria del Ecuador - Facultad de Ciencias Agrarias - Escuela de Computación e Informática

Autora: Ing. Elke Jacqueline Yerovi Ricaurte, MSc. / eyerovi@uagraria.edu.ec; Teléfono: 0986215720

Coautores: Ing. Pablo Andrés Calderón Castro / pacalder@espol.edu.ec; Teléfono: 0991740714

Ing. Mariuxi Ileana Tejada Castro / mtejada@uagraria.edu.ec; Teléfono: 0994439044

RESUMEN

En esta investigación se examinó algoritmos de toma de decisiones para determinar quién es bot o humano en un ambiente de red social como Twitter. Se realizó métodos de gráficas estadísticas que mostraron las cuentas de Twitter y luego los árboles de decisión con la ponderación de variables que revelan rasgos humanos como son la amistad, usuarios, contenido, tiempo de respuesta (temporalidad), calidad de respuestas (sentimentalidad) y composición de redes, esta es una clasificación de características empleadas por sistemas basados en rasgos de comportamientos para detección de los bots sociales siendo la base de la investigación la recuperación de la data que muestra resultados de influencia considerables de Bot Or Not perteneciente a 50 marcas de empresas ecuatorianas que tienen mayor presencia en internet y específicamente en sus cuentas de Twitter. Esto parte de la premisa que brinda el Test de Turing, el cual determina si un computador, que tiene un grado relativamente elevado de inteligencia artificial pre-programado, es capaz de pensar como un humano. En el dendrograma 41 empresas de las 50 (82%) mostraron mayor concentración de influencia de bots sociales y luego se incluyó el análisis resultante del árbol de decisiones que nos mostró como conclusión que el contenido de los "Tuits" es la evidencia de la presencia de bots. Hoy en día, los ecosistemas de los medios sociales están poblados de cientos de millones de individuos que presentan reales incentivos -incluyendo económicos y políticos- para diseñar algoritmos que exhiban comportamientos cercanos a los humanos.

Palabras clave: Árbol de Decisión, Bots, Data Mining, Twitter, Bot Or Not.

ABSTRACT

In this research, decision-making algorithms to determine who is bot or human was examined in an environment of social networks like Twitter. Methods of statistical graphs showing Twitter accounts and then the decision trees with the weighting of variables that reveal human features such as friendship, users, content, response time (temporary), quality of answers (sentimentality) was performed and composition of networks, this is a classification of characteristics used by behavioral traits based detection of bots social systems being based on research data recovery results showing considerable influence Bot Or Not belonging to 50 brands Ecuadorian companies that have a greater presence on the Internet and specifically in their Twitter accounts. This starts from the premise that gives the Turing Test, which determines if a computer, which has a relatively high degree of pre-programmed artificial intelligence is able to think like a human. In the dendrogram 41 companies of the 50 (82%) showed the highest concentration of influence of social bots and then the resulting decision tree analysis showed us that conclusion included the contents of the "Tweets" is evidence of the presence bot. Today, the social media ecosystems are populated by hundreds of millions of individuals who have real incentives, including economic and political to design algorithms to exhibit near human behavior.

Keywords: Decision Tree, Bots, Data Mining, Twitter, Bot Or Not.

INTRODUCCIÓN

A lo largo del nuevo milenio se han venido desarrollando en el internet diferentes tipos de algoritmos versátiles y adaptables caracterizados como bots o “programas robot” que al ser sistemas de ingeniería social se alimentan inicialmente de información digital personal, estos procuran capturar datos de la mayor cantidad de usuarios que en general sus antecedentes son parte del “Big Data” que según Edd Dumbill (O’Reilly Media, 2012) sucede cuando cantidades masivas de datos son inmanejables por las bases de datos que los mismos salen de las restricciones de sus arquitecturas. Además, estudios en recientes años subrayan (Wagner, Mitter, Körner, & Strohmaier, 2012) que al entrar en el entorno social objetivo, los bots requieren pasar desapercibidos (Huber, Kowalski, Nohlberg, & Tjoa, 2009) al entremezclarse entre el público digital de una red social “aparentando” y tomando el lugar de un ser humano usando la abstracción de nuestro comportamiento e inteligencia por medio de datos pre-procesados por algoritmos sofisticados que involucran inteligencia artificial, todo ejecutado con la conexión a una cuenta o cuentas de usuarios aparentemente regulares de la red social monitoreada por bots sociales.

En el Ecuador todavía no se ha planteado una revisión de la influencia de los bots sociales en ninguna de las redes sociales usadas ampliamente en el país o al menos determinar su presencia, la problemática que esto conlleva es que en general no se está consciente de su presencia cada vez más pronunciada en el

ámbito local y las repercusiones o alcance que puedan tener las acciones de estos autómatas digitales, sean estas repercusiones positivas o negativas en distintas situaciones de decisión, tampoco no se ha estudiado el incremento en incidencias por la manipulación de los medios de comunicación masiva en los servicios de redes sociales (SRS) en el ambiente digital ecuatoriano tal como se ha tratado al respecto en distintas investigaciones científicas en otros países, para llegar a entender el impacto de su ataque (Mitter, Wagner, & Strohmaier, 2014) al manipular la conciencia pública, el tener pautas para entender su impacto y poder defendernos de estos (Boshmaf, Muslukhov, Beznosov, & Ripeanu, 2012) o conocer más a fondo del alcance de la manipulación de información que son capaces (Ferrara, 2015) en situaciones sociales y políticas.

Lo que se quiere destacar en el estudio presente es el aplicar conceptos dados y mejorar metodologías específicamente para el mercado ecuatoriano, inspirados en estas y otras investigaciones mencionadas anteriormente, mostrando entre las distintas soluciones una metodología reproducible que permita discernir entre la máquina y el hombre en los medios sociales digitales locales, específicamente en cuentas de Twitter ecuatorianas, pudiendo también contribuir a investigaciones locales previas que tratan de eventos electorales recientes (Camada Fiallos, 2012) pero que no mencionan la influencia de los bots sociales en sus resultados y conclusiones, tomando en cuenta los datos descritos a continuación.

Objetivo de la investigación

Contribuir con procedimiento reproducible y aplicable basado en algoritmos de árboles de decisión, más el uso de dendrogramas para revelar la presencia de bots los cuales pueden ejercer influencia en medios digitales específicamente en la red social Twitter.

MATERIALES Y MÉTODOS

Esta investigación brinda un planteamiento metodológico y técnico que busca reproducir un procedimiento para descubrir cuentas de Twitter que pertenecen a bots sociales en el mercado ecuatoriano, distinguiéndose de las que son cuentas de humanos, sin hacer distinción entre cuentas falsas y reales pero diferenciando su comportamiento por medio de un análisis de los resultados estadísticos obtenidos posteriormente. Para la gestión en la recopilación de datos iniciales se tomó en cuenta la herramienta de análisis web Socialbakers, esta recopila datos generales de distintos servicios redes sociales (SRS) de forma directa e indirecta, con la posibilidad de seleccionar solamente la red social que es de interés al estudio presente, en este caso Twitter y el país específico, en este caso Ecuador, se dispuso de cincuenta (50) cuentas más destacadas en internet de marcas de empresas respectivamente ecuatorianas .

El lista resultante mostró todas las empresas con presencia en Twitter que al momento de la investigación (Julio de 2015) mostraron considerable base de clientes y actividad (Socialbakers, 2015) en el mercado ecuatoriano, para este análisis el orden o aparición de las empresas es irrelevante ya que en un principio sólo se requiere una base de marcas de empresas locales, en este caso cincuenta (50) de estas, aparte las lista de empresas varía su ubicación cada minuto al ser datos generados en vivo. A través de la lista mostrada por la mencionada aplicación web (socialbakers) que proporciona rankings de cuentas de Twitter por diferentes variables de actividad social, los datos reunidos se muestran a Julio de 2015.

Luego se introduce, las cuentas de Twitter seleccionadas, una a una por el sistema Truthy: “Bot or Not Bot?” para recolectar los porcentajes de BotOrNot, los cuales involucran los factores de Contenido (Content), Temporal (Temporalidad), Sentiment (Sentimiento), Network (Redes), Friend (Amigos), User (Usuario).

Tabulando los datos de Socialbakers y Truthy en una hoja de cálculo de Microsoft Excel que permite finalmente obtener una base de datos para luego llevarla a su análisis posterior con las debidas herramientas de cálculos estadísticos que permitirán hacer un análisis empírico a fondo,

Se procede a convertir esta hoja de cálculo en un documento de valores separados por comas (comma separated values) grabando el documento como CSV, luego este formato se lo carga en RStudio, el cual contiene al lenguaje R que es orientado a realizar y ejecutar algoritmos para resultados en datos y gráficas estadísticas, este es un lenguaje de código y distribución libre de licencias de pago, el RStudio con R permiten convertir cualquier documento CSV en una tabla de datos (dataset) o grupos de datos “crudos”, a lo cual Myatt (2014) afirma que “el punto de partida del análisis de datos es la tabla de datos (conocida como data set) que contiene las mediciones o valores cotejados de datos representados como números o textos” (p. 17), que al ser empleado con algoritmos estadísticos se pueden asignar valores adecuados para dar las respuestas más exactas posibles al momento del análisis.

Métodos gráficos en Dendrogramas y Árboles de Decisión.

Un ejemplo de clasificación jerárquica o dendrograma es la de las razas de la humanidad o datos genealógicos de los mismos, para elaborar dichos trazados existen variados algoritmos para construir árboles jerárquicos

o dendrogramas. Un algoritmo aglomerativo comienza con n sub-clústeres, cada uno conteniendo un sólo punto de información y en cada etapa se fusionan en dos grupos similares más para formar un nuevo clúster,

así reduciendo el número de clústeres a uno. Las sucesivas subdivisiones de un árbol de decisión da criterios de agrupación de los mismos, según la relación de estos se acercarán más unos datos que otros. Al existir un orden aleatorio de generación de las ramificaciones el algoritmo asigna un valor numérico a cada una de las posiciones que se asocian con una medida de distancia o desigualdad según lo afirman distintas investigaciones anteriores (Cuadras, 2014; Peña M., 2001; Webb & Copesey, 2011) que cubren el desarrollo y análisis de dendrogramas.

Ahora Lacourty (2011) describe un árbol de decisión como “el resultado gráfico de un método que permite tomar “buenas” decisiones involucrando “riesgos” y “costos”. Utiliza un enfoque visual de agrupamientos de datos mediante reglas fáciles de entender” (p. 7). Un árbol de decisión se compone de entradas, las mismas que pueden ser un

objeto descrito mediante un conjunto de atributos que presentan resultados y a su vez estos facilitan una decisión que será tomada según las entradas. El análisis de los datos e interpretación de estos, después de mostrarse el árbol con sus hojas, se hace en general de forma personalizada a mano, en el cual el encargado del análisis (especialista en estadística) desarrolla un escrito basado en sus conocimientos y experiencia que reflejen las tendencia, proyecciones y puntos destacados de estos resultados, al mostrar sus conclusión esta sirve para tomar decisiones de peso en una investigación. Dicho proceso es aletargado, oneroso y bastante subjetivo, más cuando el volumen de datos es extremadamente grande o en crecimiento exponencial porque sobrepasa nuestra capacidad de entender su potencial a menos que se utilice una herramienta digital especializada para el efecto, en este caso una aplicación para crear y ejecutar algoritmos estadísticos.

Resumen de aplicaciones web y escritorio utilizadas.

R y RStudio: El lenguaje para estadísticas R ofrece gran variedad de técnicas gráficas estadísticas como análisis de series de tiempo lineal y no lineal, modelado, evaluaciones estadísticas clásicas, clasificación, agrupación, siendo además altamente extensible, R proporciona una ruta de código abierto a la participación en el campo de la estadística. Adicionalmente RStudio es mucho más que una interfaz bonita, en realidad es un entorno de desarrollo integrado (IDE), el cual ha sido cuidadosamente diseñado en base a las necesidades de los usuarios del lenguaje R, los investigadores pueden usarla como una herramienta principal para sus estudios científicos, es fácil de usar y los resultados podrán visualizarse mediante representaciones gráficas que permitirán tomar decisiones fácilmente.

Text Mining usando R: Análisis de datos de Twitter de acuerdo a Zhao (2014): Este es otro de los métodos con los que nos permite trabajar RStudio, por lo tanto se debe

considerar cada uno de los siguientes puntos a seguir para aplicar. (p. 34)

- Datos de texto no estructurados
- Categorización de textos
- Agrupación de textos
- Extracción de entidades
- Análisis de sentimientos
- Documento resumen

Dentro del RStudio con R se puede ingresar directamente datos de Twitter pero su API fue limitada recientemente a Octubre de 2015, el procedimiento usual era el configurar el API de Twitter y autenticarlo en RStudio para recopilar información variada de conteos, retweets, etc. El estudio realizado en el artículo científico no se utilizan los datos de Twitter directamente sino de Socialbakers y Truthy respectivamente porque se requerían ciertos datos que den una idea de abstracción humana como sentimentalidad, calidad de respuestas humanas, atención en las respuestas, etc., que lo brinda el proyecto Truthy de la Universidad de Indiana.

Cuenta Twitter.

Twitter está basado en un formato muy simple, tanto en su diseño como en su utilización, y quizás en eso radique su éxito mundial. Los usuarios que lo utilizan son de diversa índole y los usos que se le dan son múltiples. Van desde un uso meramente personal, hasta un uso enfocado en el ámbito empresarial. Muchas veces, es tanta la cantidad de información publicada por los que uno sigue (followings), que es difícil que un usuario sea capaz de mantenerse al día y leer todo lo que éstos publiquen. Puede que algunos mensajes de interés pasen desapercibidos debido a que no hay filtros que separan la información que resulta relevante de la que no lo es. Para evitar que esto suceda, se ha creado un servicio de filtrado y agregación de cuentas de Twitter

a partir de temáticas definidas a través de hashtag.

Finalmente se describe un procedimiento resumido de lo que usualmente se hace para activar bots para Twitter: Un usuario de una cuenta de Twitter puede tomar el control como administrador con una interfaz web, este tendrá a disposición una comunidad conformada por varios usuarios reales y una cuenta bot contando que el usuario administrador tiene control total sobre la distribución del contenido de la cuenta bot, de acuerdo a González (2012) es sólo “necesario seguir únicamente a la cuenta robot, puesto que ésta publicará los contenidos ya filtrados de todos los usuarios de Twitter pertenecientes a la comunidad” (p. 9)

Truthy – Bot or Not?.

Proyecto realizado por la universidad de Indiana (Indiana University, 2015) con los investigadores del Center for Complex Networks and Systems Research en conjunto con el School of Informatics and Computing ha estado estudiando las formas en que la información se propaga a través de las redes sociales tal como Twitter. Este proyecto es fundado por el gobierno federal de los estados unidos.

El proyecto informalmente se llama “Truthy” (CNetS, 2015) y hace uso de modelos

complejos para analizar la distribución de información en redes sociales para determinar qué sentimiento es popular, la influencia de usuarios, la atención a mensajes y contenido, la estructura de la red social de un usuario y otros factores que afectan la manera en que la información es diseminada. Una meta importante del proyecto es entender cómo los medios sociales digitales pueden ser abusados. El proyecto no ha estado exento de crítica pero de miembros políticos que difundieron noticias mal infundadas del servicio (Uberty, 2014).

Truthy – Bot or Not?.

La implementación que se utilizó en este estudio fue analizar los datos con el modelo Árbol de decisiones y el modelo de Dendrograma, por lo que se empleó el software de estadísticas RSTUDIO con el lenguaje R (x64 3.2.1). Además, para determinar si una cuenta es un Bot, se ejecutaron las siguientes variables: Friends, Content, Temporal, User, Sentiment y Network, del proyecto

Truthy de las lista de 50 marcas de empresas ecuatorianas con mayor presencia en Twitter desde la página web de Socialbakers (Socialbakers, 2015), además se empleó el uso de Dendrogramas y Árboles de Decisión que pueden demostrar gráficamente los resultados obtenidos para finalmente analizar los resultados obtenidos.

A continuación se detallan los pasos que se realizaron para presentar el árbol jerárquico de la Figura 1:

Paso 1.

Se importó el dataset <marcas> con el archivo <marcas.csv>

```
> marcas2_kmean = kmeans(marcas2, 4)
> marcas2_kmean$cluster
[1] 1 2 4 1 2 3 4 1 3 1 1 4 4 3 1 1 3 3 4 1 3 1 1 1
2 1 4 1 3 2 3 4 4 3 1 3 3 2 3 2 1 3 2 1 4 2 1 2 4 1
```

Paso 2.

Copiar el dataset <marcas> en <marcas2> y quitamos la columna <Marca>.

```
> marcas2 = marcas
> marcas2$Marca = NULL
```

Paso 4.

Unir el resultado obtenido con el dataset para hacer el dendrograma.

```
> marca_resultado = data.frame(marcas2,
marcas2_kmean$cluster)
> hc = hclust(dist(marcas2), method = "ave")
> plot(hc, hang=-1, labels=marcas$Marca,
xlab = "Marcas Twitter")
> rect.hclust(hc, k=4)
> groups <- cutree(hc, k=4)
```

Paso 3.

Aplicar el comando kmeans para calcular cuales filas van en el cluster, por lo tanto usamos 4 para el número de clúster ya que vimos anteriormente que era el número óptimo.

RESULTADOS

De tal manera se obtuvo como resultado el siguiente gráfico que representa el dendrograma obtenido con un muestreo y agrupación en 4 clusters. Como se muestra en la Figura 1.

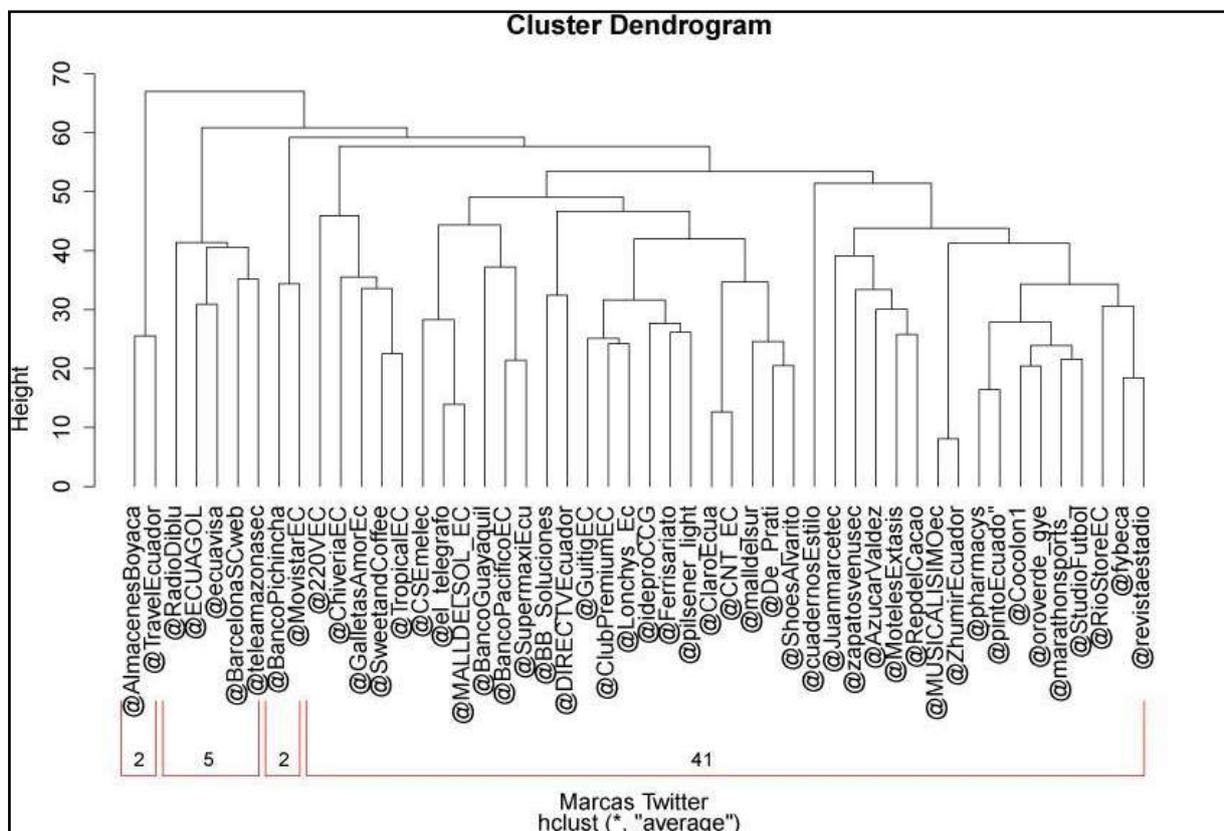


Figura 1. Dendrograma obtenido con las cuentas de las 50 marcas ecuatorianas con mayor presencia en el Twitter mediante el programa RStudio que utiliza el lenguaje de estadísticas R

Se ejecutaron las siguientes variables para determinar si una cuenta es un Bot o no Bot: Friends, Content, Temporal, User, Sentiment y Network y se realizó un Árbol de Decisión, mostrando gráficamente los datos obtenidos. A continuación se detallan los pasos que se realizaron para presentar el árbol de decisiones como se aprecia en la Figura 2.

Paso 1.

Ejecutar la librería <party>
> library(party)

Paso 2.

Luego ejecutar el comando ctree buscando las condiciones que definen si una cuenta es Bot or Not.

> Marcas_ctree = ctree(BotOrNot ~ Friend + User + Content + Temporal + Sentiment + Network, data = marcas)

Paso 3.

Imprimir el resultado.

> print(Marcas_ctree)

Interfaz gráfica que muestra un árbol con 3

nodos terminal

Response: BotOrNot

Inputs: Friend, User, Content, Temporal, Sentiment, Network

1) Content <= 47; criterion = 1, statistic = 16.168

2) User <= 59; criterion = 0.996, statistic = 11.761

3) * weights = 22

2) User > 59

4)* weights = 18

1) Content > 47

5)* weights = 10

x

Paso 4.

Imprimir el gráfico del árbol de decisión.

> plot(Marcas_ctree)

El gráfico finalmente se muestra como en la Figura 2.

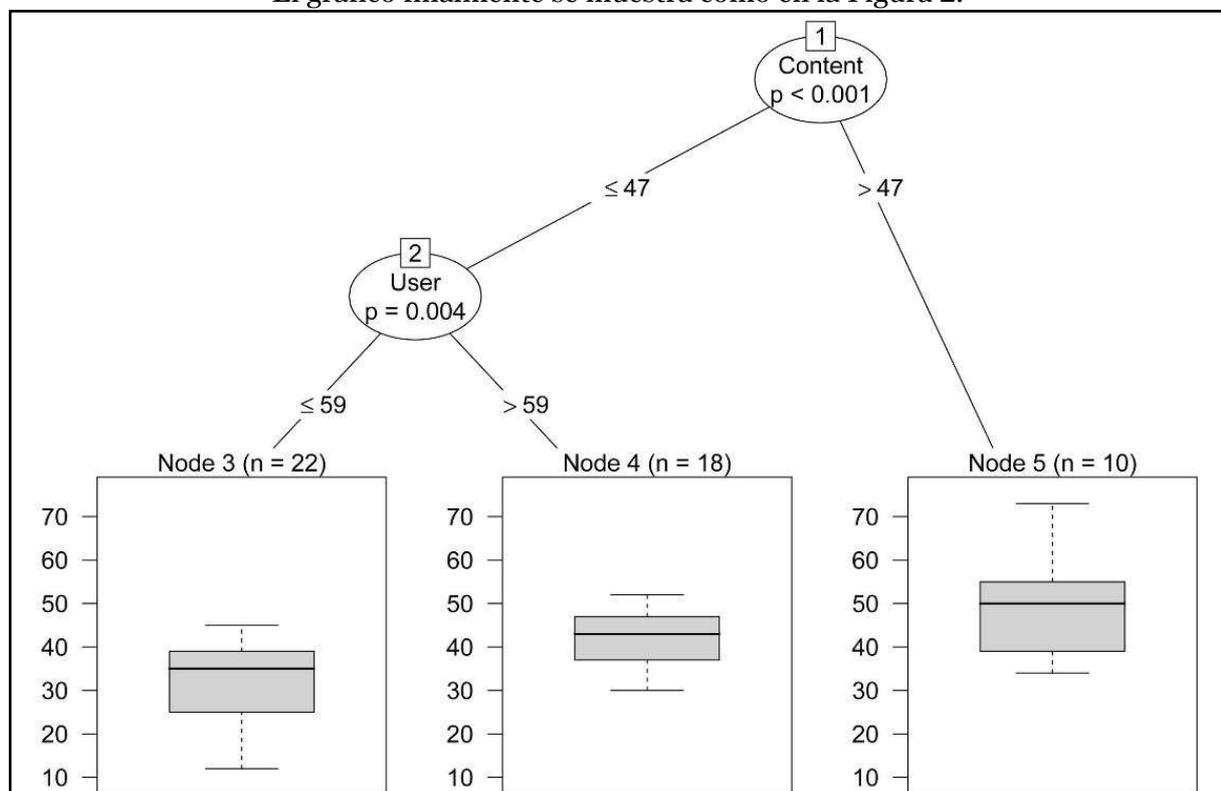


Figura 2. Representación gráfica del Árbol de decisiones que muestra el resultado obtenido del análisis de las 50 marcas ecuatorianas con mayor presencia en Twitter que determinan si sus cuentas son BotOrNot.

DISCUSIÓN

Así como aparece en la Figura 2¹, en el árbol de decisión se denota que:

- Si una cuenta de Twitter obtiene un valor de Content mayor a 47% sus probabilidades de tratarse de una cuenta Bot van desde 34% aproximadamente hasta más de 70%. Lo que representa 10 de los 50 casos analizados, ósea que la calidad del contenido revelaría que es bot.
- Si el valor de Content es inferior o igual a 47% y el valor de User es inferior a 59%, las probabilidades de que se trate de una cuenta Bot se ubica en 30% hasta el 54% aproximadamente. Lo que representa 18 de los 50 casos analizados.
- Si el valor de Content es menor a 47% y el valor de User es menor o igual que 59% las probabilidades de que la cuenta sea Bot se ubica en 10% hasta el 45%. Esto ocurre en 22 de los 50 casos analizados.

¹ Según comenta Zhao (2013) en la documentación acerca del lenguaje R dice que para evitar ciertos inconvenientes con las recientes versiones del algoritmo “ctree()” se tiene que repetir la construcción del árbol de decisión por segunda ocasión (Zhao, 2013, p. 31)

Con respecto al árbol de decisiones se puede notar que si una cuenta de Twitter indica si es bot ó not bot (robot social o no robot social) su influencia en general es negativa, refiriéndose a que los tweets que son escritos por un humano son diferentes a un retuit o un contenido de un bot que retuitea o genera un contenido repetido en varias instancias.

Las cuentas de Bots tienden a ser más recientes, además los retuitean más que los humanos y tienen nombres de usuario más amplios, mientras que los mismos retuitean menos, tienen menos respuestas o menciones especiales y son mucho menos retuiteados que los humanos. El contenido revela en general

cual es cual, pero con minucioso escrutinio.

En investigaciones previas, Ferrara et al. (2014) recalca que cuando el contenido apropiado es identificado por los bots, estos adaptan sus respuestas a través de algoritmos sofisticados, creando situaciones familiares en el usuario víctima, mientras otros toman vías de suplantación de identidad a tal punto que se comportan igual al usuario de la cuenta original. Esto brinda a colación la cantidad real de bots que el público ecuatoriano no identifica como tal, nos encontramos con millones de bots a identificar para intentar concienciar de su prominente presencia e influencia en el mercado ecuatoriano.

CONCLUSIONES

En efecto después de haber ponderado las variables que revelan rasgos humanos como son la amistad (Friends), usuarios (User), contenido (Content), temporalidad (Temporal), sentimentalidad (Sentiment) y redes (Network), con el servicio web (Bot or Not Bot) que analiza el comportamiento de un usuario de Twitter y a su vez lo compara con el comportamiento de los robots sociales (bots sociales) y junto al programa

de estadísticas con los procedimientos de dendograma y árbol de decisión, esto puede ser metodológicamente utilizado, al ser fácilmente reproducible por el público que requiera dicha solución y confirmar que el contexto del contenido (Content) es dónde se ve de primero si la cuenta es Bot o no.

En el árbol de decisiones se obtienen dos casos: el primer caso que si una cuenta de

Twitter se obtuvo un valor de Content es inferior o igual a 47% y el valor de User es inferior a 59%, las probabilidades de que se trate de una cuenta Bot van desde el 30% hasta el 54% aproximadamente, esto representa 18 de los 50 marcas analizadas, en cambio en el segundo caso si el valor de Content es menor a 47% y el valor de User es menor o igual que 59% las probabilidades de que la cuenta sea Bot bajan desde un 10% al 45% esto ocurre en 22 de los 50 marcas analizadas, esto implica que ambos casos en un futuro los robots sociales (bots sociales) van a reemplazar a los humanos o al menos presentar una gran

presencia virtual a tal punto que no se pueda distinguir que es humano y que es robot del todo.

En el árbol de decisiones al verificar si una cuenta de Twitter se obtuvo un valor similar al dendrograma en donde el Content es mayor a 47% sus probabilidades de tratarse de una cuenta Bot van desde 34% aproximadamente hasta más de 70%, esto representa a 10 de las 50 marcas analizadas, implica que en el futuro los bots sociales van a reemplazar a los humanos al no discernir que es humano y lo que es robot.

BIBLIOGRAFÍA

- Anderson, M. (2015). Twitter's withdrawal of reliable share count API is a bold monetising move. Retrieved November 29, 2015, from <https://thestack.com/cloud/2015/10/05/twitters-withdrawal-of-reliable-share-count-api-is-a-bold-monetising-move/>
- Boshmaf, Y., Muslukhov, I., Beznosov, K., & Ripeanu, M. (2012). Key challenges in defending against malicious socialbots. In Proceedings of the 5th USENIX conference on Large-Scale Exploits and Emergent Threats (p. 12). USENIX Association.
- Camada Fiallos, R. (2012). Aplicación de Técnicas de Minería de Datos para la Indagación y Estudio de Resultados Electorales. CIENCIAMÉRICA, 1, 67–76.
- CNetS. (2015). The Truth about Truthy. Retrieved November 29, 2015, from <http://cnets.indiana.edu/blog/2014/08/27/the-truth-about-truthy/>
- Cuadras, C. M. (2014). Nuevos métodos de análisis multivariante. Barcelona, Spain: CMC Editions.
- Ferrara, E. (2015). Manipulation and abuse on social media. arXiv Preprint arXiv:1503.03752.
- Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2014). The Rise of Social Bots. arXiv Preprint arXiv:1407.5225.
- Forelle, M., Howard, P., Monroy-Hernandez, A., & Savage, S. (2015). Political Bots and the Manipulation of Public Opinion in Venezuela. arXiv Preprint arXiv:1507.07109.
- García González, M. (2012). Implementación de un servicio de filtrado y agregación para comunidades en Twitter. Universidad Carlos III de Madrid.
- Huber, M., Kowalski, S., Nohlberg, M., & Tjoa, S. (2009). Towards automating social engineering using social networking sites. In Computational Science and Engineering, 2009. CSE'09. International Conference on (Vol. 3, pp. 117–124). IEEE.
- Hwang, T., Pearce, I., & Nanis, M. (2012). Socialbots: Voices from the fronts. Interactions, 19(2), 38–45.
- Indiana University. (2015). Truthy

- BotOrNot Tool. Retrieved from <http://truthy.indiana.edu/botornot/>.
- Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, 60(11), 2169.
 - Mitter, S., Wagner, C., & Strohmaier, M. (2014). Understanding the impact of socialbot attacks in online social networks. arXiv Preprint arXiv:1402.6289.
 - Mowbray, M. (2014). Automated Twitter Accounts. *TWITTER AND SOCIETY*, 183.
 - Myatt, G. J., & Johnson, W. P. (2014). *Making Sense of Data I: A Practical Guide to Exploratory Data Analysis and Data Mining*. (J. Wiley, Ed.) (2nd ed.). Wiley.
 - Nilsson, N. J. (2009). *The Quest for Artificial Intelligence*. Cambridge University Press.
 - O'Reilly Media, I. (2012). *Big Data Now: 2012 Edition*. O'Reilly Media.
 - Peña M., J. (2001). *Arquitectura distribuida de control para sistemas con capacidades de data mining*.
 - Romportl, J., Ircing, P., Zackova, E., Polak, M., & Schuster, R. (2012). *Beyond AI: Artificial Dreams*.
 - Romportl, J., Zackova, E., & Kelemen, J. (2014). *Beyond Artificial Intelligence: The Disappearing Human-Machine Divide*. Springer International Publishing.
 - Socialbakers. (2015). *Socialbakers Brand Statistics from Twitter, Ecuador 2015*. Retrieved July 25, 2015, from <http://www.socialbakers.com/statistics/twitter/profiles/ecuador/brands/page-1-5/>
 - Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 433–460.
 - Uberty, D. (2014). How misinformation goes viral: a Truthy story - *Columbia Journalism Review*. Retrieved November 30, 2015, from http://www.cjr.org/behind_the_news/how_misinformation_goes_viral.php
 - Wagner, C., Mitter, S., Körner, C., & Strohmaier, M. (2012). When social bots attack: Modeling susceptibility of users in online social networks. *Making Sense of Microposts (# MSM2012)*, 2.
 - Webb, A. R., & Copsey, K. D. (2011). *Statistical Pattern Recognition*, 3rd ed. (J. Wiley, Ed.) (3rd ed.). Marlvern, United Kingdom: Wiley & Sons, Ltd.
 - Weller, K., Bruns, A., Burgess, J., Mahrt, M., & Puschmann, C. (2014). *Twitter and Society*. (K. Weller, A. Bruns, J. Burgess, M. Mahrt, & C. Puschmann, Eds.). New York: Peter Lang.
 - Zhao, Y. (2014). *R and Data Mining: Examples and Case Studies*, 1–160.