

Uso de minería de textos para comparar los contenidos relacionados a calidad y acreditación generados en redes sociales por universidades de Perú y Chile

Nicolás A. Núñez¹, Rodrigo A. Crisóstomo² y Sandro A. Sánchez³

(1) Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Santiago, Chile.

(2) Facultad de Ingeniería, Universidad Andrés Bello, Santiago, Chile.

(3) Centrum PUCP Business School, Pontificia Universidad Católica del Perú, Lima, Perú.

(Correo-e: nnunez@ing.uchile.cl; r.crisostomozuiga@uandresbello.edu; sasanchez@pucp.edu.pe)

Recibido May. 26, 2020; Aceptado Jul. 20, 2020; Versión final Oct. 5, 2020, Publicado Feb. 2021

Resumen

Este estudio analiza el contenido generado en redes sociales online por universidades peruanas y chilenas en temas de calidad y acreditación. Se comparan los niveles de uso de palabras clave entre dos categorías de universidades (acreditadas y no acreditadas), en el marco de las estrategias de comunicación externa de las instituciones. La recolección de datos se realizó en Twitter y Facebook, aplicando minería de textos para hacer un conteo de palabras más frecuentes. Se aplicó el algoritmo bosque aleatorio (random forest) para efectuar una clasificación binaria. Los resultados muestran que los términos más utilizados por las universidades fueron “calidad”, “servicio” y “gestión”. Los resultados del clasificador refrendan lo obtenido en la minería de textos, donde la cantidad de contenidos publicados sobre calidad y acreditación no tendrían relación con las posiciones obtenidas en los rankings universitarios. Se concluye que las instituciones universitarias deben revisar sus estrategias de contenidos en redes sociales a fin de lograr una mayor diferenciación para asegurar su posicionamiento.

Palabras clave: acreditación; educación superior; bosque aleatorio; minería de textos

Use of text mining to compare quality and accreditation content generated on social media by Peruvian and Chilean universities

Abstract

This study analyzes quality and accreditation content generated on online social media by Chilean and Peruvian universities. Keywords regarding university external communication strategies are compared between two types of universities (accredited and unaccredited). Data are collected from Twitter and Facebook by applying text mining techniques to count the most frequently used keywords. The random forest algorithm is applied to perform a binary classification. The results show that the terms most used by universities were “quality,” “service,” and “management.” The results obtained from the classifier are in agreement with the results obtained by text mining, where the amount of publications related to quality and accreditation do not correlate with university rankings. It is concluded that universities should revise their content strategies on social media to achieve greater differentiation to secure their classification in university rankings.

Keywords: accreditation; higher education; random forest; text mining

INTRODUCCIÓN

En los últimos años los avances tecnológicos han permitido la generación de grandes volúmenes de datos en distintos entornos online, siendo uno de los más relevantes las redes sociales (Stieglitz et al., 2018), entre las que destacan Facebook, Twitter o WhatsApp. Las redes sociales han ido evolucionando desde la generación de contenidos personales hasta llegar a ser una herramienta poderosa para que las organizaciones puedan transmitir sus mensajes a sus distintos grupos de interés (Leonardi et al., 2013). En la Educación Superior puede verse que casi todas las universidades del mundo cuentan con una página oficial en Facebook, mientras que algunas de ellas poseen una cuenta en Twitter. En la primera red social, hay universidades que adicionalmente presentan canales de atención para consultas. Asimismo, se reportan casos en que los propios estudiantes y docentes adoptan algunas funcionalidades de Facebook como una herramienta de aprendizaje colaborativo (Arteaga Sánchez et al., 2019; Peeters, 2019). De todas formas, es relevante acotar que el uso de redes sociales por parte de las universidades no se limita únicamente a Facebook o Twitter, sino también en los últimos años aplicaciones como Instagram, Pinterest o Snapchat han sido utilizadas en experiencias de aprendizaje (Manca, 2020).

De esta forma, el estudio se orienta a determinar si los contenidos generados en redes sociales sobre Calidad y Acreditación (como parte de las estrategias de comunicación con grupos de interés externos) se diferencian entre dos tipos de universidades (acreditadas y no acreditadas). El problema que da origen a este trabajo se basa en un hallazgo empírico de carácter cualitativo que surgió en el área de marketing de una universidad peruana, en donde se percibió que los mensajes publicados en redes sociales por las distintas instituciones educativas del país no exhibían grandes diferencias (desde una perspectiva de análisis del discurso), enfocándose en ciertos aspectos relacionados a la calidad del servicio educativo (mejora continua, resultados, procesos etc.). Por tal motivo, este estudio se basa en técnicas cuantitativas a fin de demostrar si este hallazgo reportado tiene sustento.

Si bien el principal objetivo de los medios de comunicación online es informar a la comunidad universitaria respecto a las actividades y noticias generadas en la institución (Kimmons et al., 2017) en los últimos años el contenido generado en las redes sociales también tiene un segundo objetivo: atraer a los futuros postulantes, difundir información comercial (Constantinides & Stagno, 2011), así como brindar asesoría en tiempo real a través de herramientas como los *chatbots* (Agus Santoso et al., 2018). De este modo, la continua generación de contenidos digitales en redes sociales por parte de las universidades presenta una oportunidad de analizar la información desde múltiples perspectivas. En este sentido, es importante entender cómo las estrategias de comunicación de una organización son parte importante de un sistema de calidad. En la norma técnica ISO 21001:2018, se plantea que la comunicación debe ser “pertinente, precisa y oportuna, coherente con la estrategia de la organización”. Esta definición es de gran relevancia para el alcance del estudio, a fin de apreciar el rol de la comunicación como aspecto que puede marcar diferencias en el posicionamiento de una institución educativa.

Un aspecto que adquiere una gran importancia en las universidades de Latinoamérica es el Aseguramiento de la Calidad de la Educación. Distintas reformas en Educación Superior han sido introducidas en algunos de los países de la región, con orientación a aspectos como el financiamiento de la Educación Superior, la descentralización en las políticas públicas, la introducción de modelos de aseguramiento de la calidad o la inclusión, entre otros temas (Bernasconi & Celis, 2017). Dichas reformas se materializan en el establecimiento de dos categorías de universidades: aquellas que cumplen estándares de calidad a un nivel determinado (acreditadas) y las que no alcanzan este umbral de calidad (no acreditadas).

De acuerdo a Benavides et al. (2019), en el Perú las reformas introducidas a partir del año 2014 implicaron la transformación del sistema universitario peruano, dentro de lo que destaca como resultado la no acreditación y el cierre gradual de instituciones que no demostraron el cumplimiento de estándares mínimos de calidad. En Chile, se han introducido reformas de manera progresiva, siendo la primera de ellas referente a la gratuidad (Espinoza et al., 2019), continuando con la posterior promulgación de la Ley de Educación Superior en el año 2018, en donde destaca la creación de la Subsecretaría y la Superintendencia de Educación Superior, además de modificaciones sustanciales a la composición de la Comisión Nacional de Acreditación.

De esta forma, el presente estudio se orienta al análisis del contenido digital relacionado con Calidad y Acreditación generado por universidades de Perú y Chile, a fin de determinar si existen diferencias en la cantidad de publicaciones vinculadas a la temática en dos categorías de universidades: en primer término las *tradicionales*, que cuentan con una presencia consolidada, liderando los distintos rankings universitarios, además de una alta cantidad de acreditaciones, y/o certificaciones de calidad tanto institucionales como programáticas, y en un segundo nivel las denominadas “*striving universities*” (O’Meara, 2007), casas de estudio que se encuentran desplegando los primeros pasos en torno a la calidad educativa, que en la actualidad no cuentan con acreditaciones y/o certificaciones a nivel institucional y de programas, o de manera

alternativa (para el alcance de este artículo) poseen acreditaciones aunque en los niveles más bajos. El objetivo principal de las striving universities, de acuerdo a Gonzales (2015), es la mejora en los rankings universitarios. En esta misma línea, Kezar (2014) plantea que si bien la posición en un ranking es importante, la motivación para el cambio se sustenta en las demandas de la sociedad civil, las empresas y los gobiernos, por lo que la mejora continua es obligatoria en un ambiente en constante movimiento.

MATERIALES Y MÉTODOS

Machine Learning (también denominado *Data Science* o *Data Mining* de acuerdo al contexto) abarca una serie de técnicas de modelamiento orientadas a distintos requerimientos de las organizaciones (predicción de resultados, análisis de textos, visión artificial, reconocimiento de voz, entre otras), las cuales pueden ser aplicados en distintas tareas en entornos empresariales (Allmendinger et al., 2017; Yang et al., 2019). En el caso del sector educativo, el uso de estas técnicas puede apreciarse en una serie de trabajos (Castrillón et al., 2020; Etxeberria et al., 2017; Ezz & Elshenawy, 2019) orientados a la predicción del rendimiento académico de estudiantes, mientras que Salloum et al. (2018) presenta un enfoque de minería de textos para extraer información de artículos de investigación. Por su parte, De La Hoz et al. (2019) describe el uso de algoritmos de clasificación para predecir usuarios en ambientes virtuales de educación. En el tema de Acreditación y Calidad, se puede encontrar el trabajo de Fontalvo-Herrera et al. (2018) que utiliza minería de datos para clasificar programas acreditados de ingeniería industrial en Colombia.

El análisis del contenido generado por cada universidad en redes sociales vinculado a la temática de Acreditación ha sido realizado únicamente en términos cuantitativos, es decir, contar la cantidad de términos claves presentes en las publicaciones realizadas por cada universidad. Para tal efecto, se utilizará una técnica de inteligencia artificial denominada Minería de Textos (Text Mining en inglés), la cual se inserta en las técnicas de Procesamiento de Lenguaje Natural (NLP por su denominación en inglés). De acuerdo a Alvarado et al. (2013), Minería de Textos es el proceso de encontrar patrones implícitos, previamente desconocidos, que pueden ser útiles desde un repositorio de textos.

Para recolectar los datos que se utilizaron en el estudio, se consideraron dos fuentes: Facebook y Twitter, por dos razones: (a) el tipo de datos que se genera en estas redes sociales es factible de analizar a través de NLP, a diferencia de otras redes como Instagram en que el contenido se compone mayoritariamente de datos no estructurados (imágenes o videos), y (b) el volumen de datos generados en ambas redes sociales permite alcanzar niveles mínimos para realizar análisis complejos de los datos, a través de Machine Learning. En efecto, se pudo detectar que tanto en Facebook como Twitter, las universidades generan contenidos que publican cada 2 o 3 días (exceptuando fines de semana y festivos).

La extracción de los datos se efectuó por dos vías: en el caso de Twitter, la obtención de los *tweets* publicados por las universidades se hizo a través de la API de dicha red social, utilizando la librería *tweepy* en Python 3.7, lo cual permite realizar una extracción automática, implicando tiempos reducidos en esta tarea. En cambio, la obtención del contenido escrito de las publicaciones en Facebook se realizó de manera manual, en cuanto no existe una herramienta específica que pueda automatizar dicha tarea. Sin embargo, al tener una muestra relativamente pequeña de cuentas de las cuales se extrajeron sus publicaciones (60 universidades), no constituyó un obstáculo crítico para el desarrollo del estudio.

El espacio de tiempo considerado para extraer las publicaciones escritas de las universidades fue de 12 meses, comprendidos entre los días 31 de diciembre de 2018 y 31 de diciembre de 2019. Luego de completar la extracción de datos, las publicaciones en Facebook y Twitter fueron unidas en una sola base de datos para cada institución, la cual se constituyó en el archivo que sería procesado para realizar los análisis. Es importante señalar que la unión de ambas fuentes se explica por el hecho de que las publicaciones en Facebook no necesariamente fueron las mismas que en Twitter, motivadas principalmente por la limitación en los caracteres (280) que tiene esta última red social.

De esta manera, se utilizarán dos procedimientos estándares de NLP denominados tokenización y limpieza. El primero se orienta a transformar un texto en una lista de palabras, también denominadas tokens. La limpieza elimina aquellos tokens que no añaden valor en el análisis, como los signos de puntuación o palabras poco representativas ("*stopwords*") como conjunciones, preposiciones o verbos. Para ejecutar la etapa de limpieza es necesario también contar con un diccionario, o listado de términos claves. Este diccionario fue construido en base a tres documentos de referencia: (1) "Modelo de Acreditación Institucional para Universidades" elaborado en el año 2017 por SINEACE (Sistema Nacional de Evaluación, Acreditación y Certificación de la Calidad Educativa) en el Perú, (2) "Guía para la Evaluación Externa de Universidades", elaborado en el año 2020 por la Comisión Nacional de Acreditación de Chile, además de (3) "Estándar Internacional ISO 21001 – Sistemas de Gestión para Organizaciones Educativas", publicado por ISO en el año 2018.

Para la elección de las universidades involucradas en el estudio, se escogieron 30 universidades peruanas y 30 universidades chilenas, en donde en cada país se generaron dos conjuntos de universidades a estudiar, en virtud de sus características como universidad “tradicional” o como universidad “*striving*” (denominadas Grupo A y Grupo B respectivamente). Es decir, son cuatro conjuntos de universidades, los que se describen en la Tabla 1. Posteriormente, en la Tabla 2 se presentan las universidades que fueron analizadas en el estudio, de acuerdo a la pertenencia a cada grupo.

Tabla 1: Criterios para la selección de universidades de Perú y Chile en los distintos grupos

	Perú	Chile
Grupo A	Poseer licenciamiento institucional otorgado por SUNEDU, y contar con al menos una acreditación a nivel de programas, otorgada por SINEACE.	Entre 5 a 7 años de acreditación institucional otorgada por el CNA
Grupo B	No tener acreditaciones institucionales (nacionales o internacionales)	Entre 2 y 4 años de acreditación institucional otorgada por el CNA, o bien sin acreditación institucional.

Tabla 2: Listado de universidades consideradas en el estudio

Grupo	Perú	Chile
A	Pontificia Universidad Católica del Perú Universidad Católica de Santa María Universidad Católica San Pablo Universidad Continental Universidad de Piura Universidad de San Martín de Porres Universidad ESAN Universidad Nac. Mayor de San Marcos Universidad Nac. de Ingeniería Universidad Nac. Agraria La Molina Universidad Peruana de Cs. Aplicadas Universidad Privada Antenor Orrego Universidad Privada del Norte Universidad Ricardo Palma Universidad San Ignacio de Loyola	Pontificia Universidad Católica de Chile Universidad Adolfo Ibáñez Universidad Andrés Bello Universidad Austral de Chile Universidad Católica del Norte Universidad de Chile Universidad de Concepción Universidad de La Frontera Universidad de Los Andes Universidad de Santiago de Chile Universidad de Talca Universidad de Tarapacá Universidad de Valparaíso Universidad del Desarrollo Universidad Diego Portales
B	UNID Universidad Alas Peruanas Universidad de Chiclayo Universidad de Lambayeque Universidad Latinoamericana CIMA Universidad José Carlos Mariátegui Universidad Marítima del Perú Universidad Peruana Simón Bolívar Universidad Peruana Las Américas Universidad Politécnica Amazónica Universidad Privada A. Guillermo Urrelo Universidad Privada San Carlos Universidad Privada Sergio Bernales Universidad San Andrés Universidad TELESUP	UNICIT Universidad Adventista de Chile Universidad Arturo Prat Universidad Arcis Universidad Bernardo O'Higgins Universidad Bolivariana Universidad Católica Silva Henríquez Universidad Central de Chile Universidad de Las Américas Universidad del Pacífico Universidad Gabriela Mistral Universidad La República Universidad Pedro de Valdivia Universidad SEK Universidad UCINF

Por otra parte, luego de hacer uso de NLP, se aplicará otra técnica de Machine Learning denominada clasificación binaria, la cual se orienta a agrupar un set de datos en dos categorías, en donde cada una de estas categorías deben ser lo menos similares entre sí de acuerdo a sus características. Para realizar la clasificación, se utilizará el algoritmo Random Forest (*bosque aleatorio*), introducido por Breiman (2001), el cual consiste en una técnica para segmentar entidades a partir de una colección de datos según criterios estadísticos, tomando como base métodos empleados en el uso de árboles de decisión.

Una de las ventajas de Random Forest, es el mejor rendimiento y precisión que tiene en comparación a otros algoritmos como Redes Neuronales, Support Vector Machines o Árboles de Decisión, además de contar con aplicaciones directas en el sector de la educación superior (Beaulac & Rosenthal, 2019) las cuales se orientan a la predicción de resultados en procesos como admisión, deserción y selección de cursos. De este modo, puede señalarse que Random Forest constituye una herramienta de gran efectividad a la hora de predecir distintos fenómenos. Sin embargo, cuando los datos incluyen características que contienen varias categorías, Random Forest no entrega resultados tan precisos debido a que el algoritmo tiende a parcializar en favor de las características con más niveles (Altmann et al., 2010). De todas formas, como en este caso de estudio no se encuentran variables de entrada categóricas, los resultados no tienen esta pérdida de efectividad en la predicción.

Para realizar la clasificación binaria se trabajó con la librería *scikit-learn* (Varoquaux et al., 2015) de Python 3.7, en donde la validación del clasificador se realizó bajo el enfoque *k-fold cross validation*, en donde $k = 10$. Tomando en cuenta que el objetivo del estudio no es realizar predicciones futuras de la categoría de acreditación de las universidades, los datos se trabajaron en un conjunto único de entrenamiento. De este modo, se utilizarán métricas tradicionales de rendimiento (accuracy, balanced accuracy, recall, precision y F-measure), además de la representación obtenida a través de la curva ROC (Fawcett, 2006).

RESULTADOS Y DISCUSIÓN

A continuación, se presentan en la Tabla 3 los resultados obtenidos en la primera etapa del estudio (minería de textos), es decir, la frecuencia en la que cada uno de los tokens seleccionados apareció en las publicaciones efectuadas por las universidades en un periodo de 12 meses. Posteriormente, en la Tabla 4 se muestra la desviación estándar obtenida en cada grupo, presentada como un porcentaje de la media.

Tabla 3: Cantidad promedio de veces que cada palabra (token) fue utilizada por cada universidad en las publicaciones en redes sociales.

Grupo	Calidad	Servicio	Gestión	Proceso	Resultado	Mejora	Personas	Educación	Total
A - Perú	161	93	82	63	48	45	52	31	575
A - Chile	142	114	106	89	39	41	44	43	618
B - Perú	123	109	97	91	39	43	30	51	583
B - Chile	131	88	110	54	21	36	62	33	535

Tabla 4: Desviación estándar del promedio de veces que cada palabra (token) fue utilizada por cada universidad en las publicaciones en redes sociales.

Grupo	Calidad	Servicio	Gestión	Proceso	Resultado	Mejora	Personas	Educación
A - Perú	5.1%	6.7%	4.3%	4.7%	5.3%	6.1%	7.1%	3.1%
A - Chile	6.3%	5.9%	6.7%	6.6%	6.1%	7.1%	5.2%	4.4%
B - Perú	11.5%	12.7%	15.1%	16.4%	12.3%	12.9%	21.3%	10.5%
B - Chile	7.4%	7.9%	5.8%	5.9%	6.4%	7.3%	4.9%	3.9%

A partir de ambas tablas puede apreciarse que no existe una tendencia clara que permita asociar niveles determinados de frecuencia de uso de los tokens seleccionados a alguna categoría de universidades, esto es, no es posible concluir (por ejemplo) que las universidades de alguno de los grupos tienen una mayor frecuencia promedio de uso de palabras en comparación al otro grupo. Si bien en las tablas se presentan los resultados agregados para cada grupo, al realizar el análisis a nivel individual los datos no permiten establecer que exista una diferencia apreciable en el nivel de uso de palabras claves en las publicaciones en redes sociales relacionadas a Calidad y Acreditación.

En este mismo sentido, al efectuar un análisis cruzado entre países, tampoco se puede establecer un patrón claro de mayor o menor uso entre las universidades de Chile y Perú. Por ejemplo, el grupo que en promedio utilizó la mayor cantidad de tokens fue el A de Chile (universidades “tradicionales”), mientras que el grupo que en promedio hizo uso de la menor cantidad de tokens fue el B de Chile. En Perú, ambos grupos tienen un promedio prácticamente similar, en virtud de que la diferencia fue de 8, cifra que apenas supera el 1% de todos los tokens utilizados en el periodo de estudio. Ahora bien, si la cantidad de conjuntos de estudio se modificara y quedara en 2 (30 universidades de Perú y 30 universidades de Chile sin separar por categoría), los resultados resultan aún más coincidentes, en la medida que en promedio, las universidades peruanas utilizaron 579 veces los tokens (palabras claves) en sus publicaciones efectuadas en Facebook y Twitter, cifra que contrasta con las 571.5 veces que las instituciones chilenas utilizaron las mismas palabras.

No obstante, vale la pena destacar la dispersión intragrupo que se encuentra en la cantidad de palabras utilizadas por las universidades pertenecientes al grupo B de Perú, en la medida que es muy superior a las desviaciones estándar encontradas en los otros tres grupos. Una característica que distingue al grupo B peruano, es que la mayor parte de las universidades que pertenecen a este grupo tienen una vida institucional menor o igual a 20 años, aunque el denominador común que tienen las 15 universidades de este grupo es la percepción de “baja calidad” que la mayor parte de ellas tenía previo a la introducción de la Ley Universitaria (Ipsos, 2013). En efecto, durante el transcurso de este estudio a algunas de estas universidades les fue denegada la licencia para prestar el servicio educativo, por lo que a partir del año 2022 no seguirán prestando servicios educativos en el país. Esta situación puntual permitiría confirmar los resultados que se obtienen en esta etapa de minería de textos en donde se aprecia que no existiría una relación aparente entre la cantidad de publicaciones en redes sociales relacionadas a Calidad y Acreditación, y los resultados y reconocimientos obtenidos a partir de la calidad del servicio educativo.

Con la finalidad de validar los resultados obtenidos en la etapa de minería de textos, se utilizará el algoritmo Random Forest para efectuar una clasificación binaria de las universidades para confirmar la confiabilidad de los resultados obtenidos por medio de la minería de textos. En el estudio, los datos no fueron divididos entre grupos de entrenamiento (train) y evaluación (test) en la medida que se el volumen es reducido (60 instancias). A continuación, se presentan en la Tabla 5 los resultados obtenidos por el modelo predictivo de clasificación binaria:

Tabla 5: Resultados del modelo predictivo de clasificación binaria, basado en el algoritmo Random Forest

	Accuracy (A)	Balanced Accuracy (B)	Recall (R)	Precision (P)	F-Measure (F)
Muestra completa (60 universidades)	67.1%	64.9%	61.8%	65.4%	63.6%
Chile (30 universidades)	61.8%	63.3%	65.9%	60.9%	63.4%
Perú (30 universidades)	69.7%	62.9%	60.3%	62.3%	59.1%

Al apreciar los resultados de la efectividad del modelo predictivo, se puede ver que el rendimiento del clasificador es bajo, en comparación a otros trabajos en donde la efectividad supera el 80% (Rodríguez-Galiano et al., 2012; Stumpf & Kerle, 2011). En efecto, la mejora que exhibe el clasificador en comparación a un clasificador aleatorio, es baja. De manera complementaria se presenta en la Figura 1 la curva ROC (Receiver Operating Characteristic) del clasificador para mostrar cómo es el rendimiento en términos de la correcta asignación de las categorías. Al visualizar la forma de la curva se concluye que el clasificador no tiene un rendimiento sobresaliente, en la medida que la distancia a un clasificador aleatorio (recta verde que va desde (0,0) hasta (1,1)) es corta, en comparación a otros clasificadores binarios de mejor desempeño predictivo.

A partir de los resultados obtenidos tanto en la primera etapa (minería de textos) como en la segunda (Clasificación Binaria con Random Forest) puede apreciarse que no se pueden encontrar marcadas diferencias entre los grupos A y B de ambos países, en relación a la cantidad de publicaciones relacionadas a Calidad y Acreditación. Tanto al aplicar un conteo simple de términos como en la aplicación de un modelo predictivo, los resultados no son concluyentes por lo que puede señalarse que para la muestra de estudio de Perú y Chile, la variable “publicaciones en redes sociales, referentes a Calidad y Acreditación” no posee un impacto directo sobre los resultados que se obtienen, medidos como Acreditaciones (institucionales y de programas), y en el caso peruano, también como autorizaciones de funcionamiento.

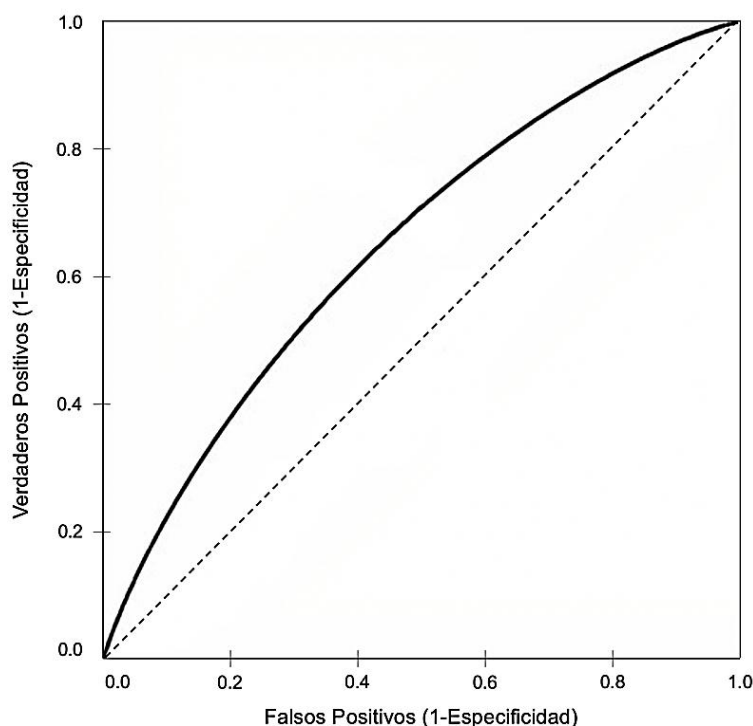


Fig. 1: Curva ROC del clasificador binario para la muestra completa

DISCUSIÓN FINAL

A partir de los resultados de este trabajo, puede apreciarse que no existen grandes diferencias en la cantidad promedio de veces que cada token fue utilizado en los distintos grupos de universidades, lo cual puede explicarse por el hecho que tanto en Perú como en Chile las dinámicas actuales de regulación universitaria configuran escenarios en los que cada universidad debe asignar diversos recursos a la mejora continua del servicio educativo. En el caso peruano, la existencia del proceso de licenciamiento institucional (autorización de funcionamiento) funciona como una especie de “*test de esfuerzo*” en donde las universidades que superan este test cuentan con el respaldo del Estado como instituciones que prestan un servicio con ciertos estándares de calidad. De este modo, en la medida que cada universidad despliega estrategias para la obtención de su licencia, genera contenidos que sean coherentes con dichas estrategias. Un resultado interesante que se generó en el caso peruano, fue que todas las universidades incluidas en el grupo B no obtuvieron la licencia de funcionamiento, lo cual presenta un antecedente claro de que la generación de contenidos *per se* no tuvo una relación directa con el cumplimiento de la regulación.

Por otra parte, en el sistema universitario chileno las exigencias regulatorias de la calidad se vinculan con los procesos de acreditación de calidad, en donde las universidades enfocan su actividad a obtener la mayor cantidad de años de acreditación a nivel institucional y de programas. En el caso de las universidades privadas que están en el grupo B, deben efectuar esfuerzos importantes para alcanzar la acreditación en la medida que su estructura de financiamiento depende de las becas y/o créditos con garantía del Estado.

Al hacer un análisis cualitativo de los textos y contenidos generados por las universidades, es posible apreciar distintos contextos de uso de cada término, en la medida que en las universidades del grupo A (en ambos países) se orientaban más a destacar mejoras puntuales en distintos procesos, en el caso de las universidades del grupo B el uso de los términos clave se orientó a presentar transformaciones más complejas y de largo plazo, algunas de las cuales se encontraban en etapas iniciales.

CONCLUSIONES

De acuerdo al trabajo presentado y a los resultados obtenidos, se pueden plantear las siguientes conclusiones principales:

1.- Este trabajo permite evidenciar empíricamente que las publicaciones sobre Acreditación y Calidad generadas en redes sociales por universidades peruanas y chilenas de dos categorías de acreditación no tienen grandes diferencias en términos de contenidos, por lo que no existe un alto nivel de diferenciación entre las distintas instituciones en sus estrategias de comunicación externa. Esta situación permite demostrar que no existe una relación entre los contenidos publicados y las posiciones en los rankings de universidades.

2.- Se aporta nuevo conocimiento en la gestión educativa en torno a la utilización de herramientas de minería de datos en la gestión de la educación superior, en base al manejo específico de técnicas de minería de textos para el análisis del discurso, los cuales complementan otros casos de uso en educación superior.

3.- Los resultados aportan antecedentes para la toma de decisiones relacionadas a las estrategias de comunicación externa que tienen las dos categorías de universidades, a fin de lograr un adecuado posicionamiento en virtud a su posición en rankings y acreditaciones de calidad. Para ello, es primordial el desarrollo de un relato único que sea reflejo de la identidad de cada institución y que logre marcar diferencias con respecto a las demás universidades.

REFERENCIAS

- Agus Santoso, H., Anisa sri Winassih, N., y otros siete autores, Dinus Intelligent Assistance (DINA) Chatbot for University Admission Services, <https://doi.org/10.1109/ISEMANTIC.2018.8549797>, *Proceedings – 2018 International Seminar on Application for Technology of Information and Communication: Creative Technology for Human Life, ISemantic 2018*, 417-423 (2018)
- Allmendinger, R., Emmerich, M. T. M., y otros tres autores (2017). Surrogate-assisted multicriteria optimization: Complexities, prospective solutions, and business case, <https://doi.org/10.1002/mcda.1605>, *Journal of Multi-Criteria Decision Analysis*, 24(1-2), 5-24 (2017)
- Altmann, A., Toloşi, L., Sander, O., y Lengauer, T., Permutation importance: A corrected feature importance measure. *Bioinformatics*, 26(10), 1340-1347. <https://doi.org/10.1093/bioinformatics/btq134> (2010)
- Alvarado, J., y Cuervo, C., Extracción de Funciones de un Cargo usando Minería de Texto en Correos Electrónicos, <https://dx.doi.org/10.4067/S0718-07642013000500008>, *Información Tecnológica*, 24(5), 61-68 (2013)
- Arteaga Sánchez, R., Cortijo, V., y Javed, U., Factors driving the adoption of Facebook in higher education, <https://doi.org/10.1177/2042753019863832>, *E-Learning and Digital Media*, 16(6), 455-474 (2019)
- Beaulac, C., y Rosenthal, J. S., Predicting University Students' Academic Success and Major Using Random Forests, <https://doi.org/10.1007/s11162-019-09546-y>, *Research in Higher Education*, 60(7), 1048-1064 (2019)
- Benavides, M., Arellano, A., y Zárate Vásquez, J. S., Market and government-based higher education reforms in Latin America: the cases of Peru and Ecuador, 2008-2016, <https://doi.org/10.1007/s10734-018-0317-3>, *Higher Education*, 77(6), 1015-1030 (2019)
- Bernasconi, A., y Celis, S., Higher education reforms: Latin America in comparative perspective, <https://doi.org/10.14507/epaa.25.3240>, *Education Policy Analysis Archives*, 25, 67 (2017)
- Breiman, L., Random Forests. *Machine Learning*, 45(1), 5–32 (2001)
- Castrillón, O. D., Sarache, W., y Ruiz-Herrera, S., Predicción del rendimiento académico por medio de técnicas de inteligencia artificial, <https://doi.org/10.4067/s0718-50062020000100093>, *Formación Universitaria*, 13(1), 93-102 (2020)
- Constantinides, E., y Stagno, M. C. Z. (2011) Potential of the social media as instruments of higher education marketing: A segmentation study, <https://doi.org/10.1080/08841241.2011.573593>, *Journal of Marketing for Higher Education*, 21(1), 7-24 (2011)
- De-La-Hoz, E., E. De-La-Hoz y T. Fontalvo. (2020). Metodología de Aprendizaje Automático para la Clasificación y Predicción de Usuarios en Ambientes Virtuales de Educación, <http://dx.doi.org/10.4067/S0718-07642019000100247>, *Información Tecnológica*, 30, 247–254 (2020)
- Espinoza, O., & González-Fiegehen, L. E., y Granda, M. L., Avances y desafíos que enfrentan los procesos reformistas en la educación superior de Chile y Ecuador: La perspectiva sobre el acceso y el financiamiento, <https://doi.org/10.22201/iisue.20072872e.2019.27.339>, *Revista Iberoamericana de Educación Superior*, X, 25-50 (2019)
- Etxeberria, P., Alberdi, E., Eguía, I., y García, M. L., Análisis del Rendimiento Académico en Relación al Perfil de Ingreso del Alumnado e Identificación de Carencias Formativas en Materias Básicas de dos Grados de Ingeniería, <https://doi.org/10.4067/S0718-50062017000400007>, *Formación Universitaria*, 10(4), 67-74 (2017)
- Ezz, M., y Elshenawy, A., Adaptive recommendation system using machine learning algorithms for predicting student's best academic program, <https://doi.org/10.1007/s10639-019-10049>, *Education and Information Technologies*, 25, 2733-2746 (2019)
- Fontalvo-Herrera, T.J., Delahoz, E.J., y Mendoza-Mendoza, A.A., Aplicación de Minería de Datos para la Clasificación de Programas Universitarios de Ingeniería Industrial Acreditados en Alta Calidad en Colombia, <https://dx.doi.org/10.4067/S0718-07642018000300089>, *Información Tecnológica*, 29(3), 89-96 (2018)
- Gonzales, L., Faculty agency in striving university contexts: mundane yet powerful acts of agency, <https://doi.org/10.1002/berj.3140>, *British Educational Research Journal*, 41(2), 303-323 (2015)
- Fawcett, T., An introduction to ROC analysis, <https://doi.org/10.1016/j.patrec.2005.10.010>, *Pattern Recognition Letters*, 27(8), 861-874 (2006)

- Kezar, A., Higher education change and social networks, <https://doi.org/10.1080/00221546.2014.11777320>, *The Journal of Higher Education*, 85(1), 91-125 (2014)
- Kimmons, R., Veletsianos, G., y Woodward, S., Institutional Uses of Twitter in U.S. Higher Education, <https://doi.org/10.1007/s10755-016-9375-6>, *Innovative Higher Education*, 42(2), 97-111 (2017)
- Leonardi, P. M., Huysman, M., y Steinfield, C., Enterprise social media: Definition, history and prospects for the study of social technologies in organizations, <https://doi.org/10.1111/jcc4.12029>, *Journal of Computer-Mediated Communication*, 19(1), 1-19 (2013)
- Manca, S., Snapping, pinning, liking or texting: Investigating social media in higher education beyond Facebook, <https://doi.org/10.1016/j.iheduc.2019.100707>, *Internet and Higher Education*, 44, 100707 (2020)
- O'Meara, K., Striving for what? Exploring the pursuit of prestige. *Higher Education*, Vol. XXII, 121-179 (2007)
- Peeters, W., The peer interaction process on Facebook: a social network analysis of learners' online conversations, <https://doi.org/10.1007/s10639-019-09914-2>, *Education and Information Technologies*, 24(5) (2019)
- Rodríguez-Galiano, V. F., Ghimire, B. y otros tres autores, An assessment of the effectiveness of a random forest classifier for land-cover classification, <https://doi.org/10.1016/j.isprsjprs.2011.11.002>, *ISPRS Journal of Photogrammetry and Remote Sensing*, 67(1), 93-104 (2012)
- Salloum, S. A., Al-Emran, M., Monem, A. A., y Shaalan, K., Using text mining techniques for extracting information from research articles, https://doi.org/10.1007/978-3-319-67056-0_18, *Studies in Computational Intelligence*, 740, 373-397 (2018)
- Stieglitz, S., Mirbadaie, M., Ross, B., y Neuberger, C. Social media analytics – Challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management*, 39(December 2017), 156-168. <https://doi.org/10.1016/j.ijinfomgt.2017.12.002> (2018)
- Stumpf, A., y Kerle, N., Object-oriented mapping of landslides using Random Forests, <https://doi.org/10.1016/j.rse.2011.05.013>, *Remote Sensing of Environment*, 115(10), 2564-2577 (2011)
- Varoquaux, G., Buitinck, L., Louppe, G., Grisel, O., Pedregosa, F., y Mueller, A. Scikit-learn, <https://doi.org/10.1145/2786984.2786995>, *Get Mobile: Mobile Computing and Communications*, 19(1), 29-33 (2015)
- Yang, Y., See-To, E. W. K., y Papagiannidis, S. You have not been archiving emails for no reason! Using big data analytics to cluster B2B interest in products and services and link clusters to financial performance. <https://doi.org/10.1016/j.indmarman.2019.01.016>, *Industrial Marketing Management*, July 2018, 1-14. (2019)

Página en blanco