



MEMORIA

CONGRESO INTERNACIONAL DE INVESTIGACIÓN E INNOVACIÓN

M U L T I D I S C I P L I N A R I O

ISSN 2448-6035

AÑO 4 - NÚMERO 1 - Abril 2022
1ra. Parte

COMITÉ EDITORIAL

Dr. J. Artemio Pérez Muñoz
Director

Dr. Florentino Vázquez Puente
Secretario General

Dr. Esteban Tolentino Calderón
Corrector Editorial

Lic. José Antonio Arzate Barbosa
Diseño Gráfico

COMITÉ ARBITRAL

Dr. Jesús Artemio Pérez Muñoz (México)

Dr. Florentino Vázquez Puente (México)

Dr. Esteban Tolentino Calderón (México)

Dr. Angel Nava Chirinos (Venezuela)

Dra. Misleida Coromoto Nava Chirinos (Venezuela)

M.C. Julio Valentín Santana Cruz (Cuba)

MSc. Rolan Nava Chirinos (Venezuela)

Esp. Maria Nava de Güere (Venezuela)

M.C. Juan Francisco Novoa Acosta (México)

M.C. Alma Delia Navarrete Leal (México)

M.C. Jesús Mendoza León (México)

“ La Universidad como
Formadora de Investigadores,
Generadora de Conocimientos
e Innovación Tecnológica ”





Percepción del uso e impacto de las redes sociales en los estudiantes de informática de la Universidad Nacional a raíz del COVID-19

Rubén Ramírez Salas
Investigación y Desarrollo
INCAE
Alajuela, Costa Rica
ruben.ramirez@incae.edu

Irene Hernández Ruiz
Escuela de Informática
Universidad Nacional
Heredia, Costa Rica
irene.hernandez.ruiz@una.cr

Luis Alejandro Fallas Carvajal
Escuela de Informática
Universidad Nacional
San José, Costa Rica
luis.fallas.carvajal@est.una.ac.cr



CONGRESO
INTERNACIONAL
DE INVESTIGACIÓN
E INNOVACIÓN
MULTIDISCIPLINARIO 2022
ISSN 2448-6035



Abstract: The following work wants to publicize the perception of the use and impact of social networks in students of the School of Computer Science of the National University in regards of COVID-19. For this, an exploratory study was carried out and the results were analyzed with PCA (Principal Component Analysis) as a tool to reduce the information obtained into the fundamental features that describe responses variation. Also, with clustering segmentation it was possible to describe which student segments can be differentiated. The results show, in addition to the main characteristics exhibited by the sample, a polarization in the use of social networks with respect to the effects of pandemic on student behaviors.

Palabras Clave: redes sociales, PCA, Clustering, X-Means, COVID19.

I. Introducción

La Real Academia Española, define el concepto de red en el área de informática como un conjunto de computadoras o de equipos informáticos conectados entre sí y que pueden intercambiar información [1]. A partir del año 1997 aparece el concepto de red social, estas son instrumentos en los cuales la comunicación y conexión entre personas se da diariamente, y en ella se manifiestan prácticamente todas las emociones como son el miedo, la ira, la alegría, la sorpresa, la tristeza y las derivadas de estas [2].

Actualmente, las redes sociales se han convertido en un medio de comunicación instantáneo y eficaz, mediante el cual aumenta la participación de la población de un modo social, cívico y político [3].

Para el año 2016, se registra que el 78% de los costarricenses usan redes sociales, según señala un estudio del Instituto para la Integración de América Latina (Intal) y de esta forma es el segundo país latinoamericano que más utiliza plataformas digitales. El informe publicado por El País de España muestra que Costa Rica, únicamente es superada por Paraguay en donde el 83% de los ciudadanos registran cuentas en alguna red social. [4]



CONGRESO
INTERNACIONAL
DE INVESTIGACIÓN
E INNOVACIÓN
MULTIDISCIPLINARIO 2022
ISSN 2448-6035



El estudio Red 506, realizado por la empresa Unimer para el periódico El Financiero en el año 2013, indica que la cantidad promedio de permanencia en internet de las 800 personas entrevistadas fue de 2.8 horas al día, e indicó que las personas de 12 a 17 años son los que utilizan más las redes sociales con un 95% y el segundo grupo que más la utilizan es el de 18 a 24 años con un 92% [5]

La Nación [6], señala que según el informe publicado por Corporación Latinobarómetro y replicado por Statista (portal de Estadísticas) en el 2018, Costa Rica se destaca como el país latinoamericano que más utiliza WhatsApp y Facebook.

Statista [7] indica que los sitios web de redes sociales más populares en Costa Rica en el primer semestre del 2020, según porcentaje de visitas Facebook presenta un 31,57%, Pinterest 24,76%, Youtube 23,88%, Twitter 14,15%, Instagram 2,65%. Todas estas, incluidas en esta investigación.

Al mismo tiempo que Internet se ha posicionado como una herramienta ventajosa, aparecen en diferentes localidades informes sobre las consecuencias del abuso y descontrol en su utilización, conocidos son ya los términos, “desorden de adicción a Internet”, “uso compulsivo de Internet”, “uso patológico de Internet” o simplemente “adicción a Internet” [8]

Kraut [9], señala que los efectos del uso de Internet serían positivos o negativos en relación con la personalidad, el apoyo social, la edad y el tipo de uso, ya que, en el estudio realizado por él, parece ser que el tipo de uso tuvo una fuerte influencia en los efectos. Además, se encontró una diferencia del tipo de uso entre extravertidos e introvertidos, entre personas con mayor apoyo social inicial, y entre adultos y adolescentes-será diferente en la gente extrovertida y con apoyo social.

Además, existen cuestionarios que permiten evaluar la adicción a Internet, para lo cual se han elaborado en base a los criterios del DSM-IV para el juego patológico y la dependencia de sustancias. El Internet Addiction Test-IAT- [10] evalúa el grado en que el uso de Internet afecta la rutina diaria, vida social, productividad, sueño y sentimientos.

Para este trabajo, se trabajará en el uso que le dan los jóvenes a las redes sociales, el tipo de publicación que realizan, la cantidad de tiempo que permanecen en redes sociales, facilidad de uso, el sentimiento que presentan al ingresar a las redes sociales así como la utilidad de las noticias que se encuentran en las publicaciones. Para ello se utilizó una muestra por conveniencia de 100 estudiantes de la carrera de ingeniería en Sistemas de Información con una edad promedio de 20 años, pertenecientes a todos los niveles de la carrera para el campus central de la universidad. En este estudio participaron 42 mujeres estudiantes y 59 hombres estudiantes, y el mismo se realizó durante el segundo semestre del año 2020 cuando la Universidad Nacional a raíz de la pandemia de la COVID-19 decidió implementar la presencialidad remota para impartir sus clases. El siguiente trabajo



CONGRESO
INTERNACIONAL
DE INVESTIGACIÓN
E INNOVACIÓN
MULTIDISCIPLINARIO 2022
ISSN 2448-6035



se divide en: descripción de la Escuela de Informática, descripción del estudio, análisis de los resultados y conclusiones.

Los estudiantes que formaron parte de este estudio son estudiantes de la Escuela de Informática de la Universidad Nacional. Esta tiene sus primeros graduados en el área de informática a partir de 1981, cuando la carrera de Diplomado pertenecía al Centro de Cómputo. En el año de 1989, la Universidad decide que es tiempo de consolidar una nueva escuela y la adscribe a la Facultad de Ciencias Exactas y Naturales y nace lo que hoy se conoce como Escuela de Informática y Computación.

La misma tiene como misión: "formar profesionales altamente competitivos en tecnologías de la Información y comunicación, e informática educativa y contribuir a la capacitación de la población en el uso de la tecnología de la información para propiciar el desarrollo humano de la sociedad".

Debido a la situación de la Pandemia del Covid-19, se consideró importante conocer hábitos de uso de las redes sociales de los estudiantes y conocer si existe un agrupamiento o clústeres entre ellos según las características del estudio que se explican en el apartado iii. descripción del estudio.

II. Metodología

El estudio es no probabilístico y se contó con una muestra por conveniencia de 100 participaciones. Para este estudio, se realizó un cuestionario en Google Forms, el cual se elaboró con preguntas cerradas y abiertas, las cuales incluyen escalas Linkert. Para este trabajo se utilizó parte del cuestionario de El Internet Addiction Test (IAT, por sus siglas es en inglés). Este es un test que se creó en el año de 1998 que posee 20 ítems que permiten medir la presencia de la adicción al internet de las personas, el cual fue desarrollado por el Dr. Kimberly Young quien fue profesor de la Universidad de Bonaventure y director del Centro de Recuperados acerca de la adicción a internet [11].

Se logró obtener un total de $n = 100$ participantes de los cuales 59 son hombres, 42 son mujeres, lo que es estadísticamente representativo a la población, ver Israel (1992) [12], con un error $\approx$ e aproximado de 10% de acuerdo a la fórmula para tamaño de muestra con una población de $N = 1500$ estudiantes:

$$n = \frac{N}{1 + Ne^2}$$



III. Resultados

Para el análisis de resultados, la intención es analizar las características de los estudiantes en cuanto a la respuesta frente al COVID-19 en sus hábitos de uso de las redes sociales además de encontrar un posible agrupamiento que segmente la población en clasificadores o clústeres que permita una mayor comprensión de los efectos de la pandemia en sus comportamientos. Para el primer objetivo, se utiliza la técnica de PCA (del inglés Principal Component Analysis), ver Abdi y Williams (2010) [13], que es una técnica ampliamente utilizada en casi todas las disciplinas científicas para analizar observaciones en los datos y encontrar variables interrelacionadas, en otras palabras, encontrar patrones y similitud entre las variables de la encuesta para establecer los rasgos que caracterizan las respuestas de los estudiantes. Para el segundo objetivo, después de analizar y encontrar los componentes principales de la información de la encuesta por medio de PCA, se procede a utilizar un algoritmo de segmentación o clustering que permite agrupar la muestra de estudiantes en uno grupos afines.

A. Análisis de Componentes Principales (PCA)

La técnica de PCA además de obtener los patrones de las variables de la encuesta, es un mecanismo ampliamente utilizado para reducción de dimensiones. Como explica Kolenikov y Angeles (2004) [14], esta técnica permite resumir la variabilidad en la información en los componentes principales que explican dicha variación.

De esta forma, sea x en este caso una matriz con la información recopilada, la técnica de PCA localiza los valores que maximizan la variabilidad, sean estos $a_1, a_2, \dots, a_k$ descritos de la siguiente manera.

$$a_k = \arg \max V[a'x]$$

$$a: ||a|| = 1$$

$$a \perp a_1, \dots, a_{k-1}$$

El razonamiento es que cada valor maximiza la variabilidad que encuentra en los datos en una dirección en particular, mientras que el siguiente componente vuelve a maximizar la variabilidad representada, pero es ortogonal al anterior (ortogonal es el equivalente a perpendicular en un espacio euclídeo), para que así se trate de otro rasgo que explica la variabilidad, distinto al anterior.

Al utilizarlo en el análisis de resultados permite obtener los eigenvalues los cuales son los valores que nos permiten priorizar de mayor a menor los rasgos más característicos de la información de la encuesta, y además los eigenvectors que permiten describir la composición de esos componentes principales con respecto al



valor de la variable en la encuesta. El aspecto de reducción de dimensiones se explica cómo los valores de eigenvalores ordenados de mayor a menor explican un porcentaje de la variación en los datos, lo permitiría resumir así la información en menos variables que las originales, como se describirá en la sección de resultados.

Al resumir la variabilidad, lo que se hace es explicar qué es lo que los valores de la encuesta tienen como patrones que permiten describir los rasgos principales, en este caso de las respuestas de los estudiantes.

B. Segmentación (Clustering)

Para la segmentación en grupos de estudiantes, se utilizó la técnica de segmentación en clústeres o grupos utilizando el algoritmo de X-Means, el cual se utiliza para determinar el número óptimo de clústeres.

Como describen Pelleg y Moore (2002) [15], para segmentación en clústeres, es ampliamente utilizado el algoritmo de K-means, en donde se puede agrupar la información en K clústeres. Sin embargo, no es un algoritmo óptimo en el sentido que requiere que el usuario del algoritmo proporcione el número de clústeres en el que quiere segmentar la información, sin embargo, su propuesta de X-means selecciona el número X de clústeres que la misma información suministrada sugiere por medio de la optimización de parámetros que indica el mejor modelo, en este caso, el mejor parámetro para optimizar es la cantidad de clústeres.

Debido a la naturaleza exploratoria del estudio, se prefiere el algoritmo X-means para que determine el número óptimo para la segmentación adecuada de la información y se obtengan la cantidad de clústeres que mejor representen las respuestas de la encuesta.

C. Resultados

Los variables de la encuesta pasan por un proceso de transformación hacia valores dummy o numéricas, lo cual permite posteriormente aplicar el análisis con PCA, cuando los datos ya están representados de forma numérica, sin categorías o variables cualitativas.

Este proceso de codificación de las variables resulta en las siguientes variables de medición:

- Zona (rural o urbana)
- Género del estudiante
- Nivel de carrera universitaria
- Horas mínimas diarias de uso de redes sociales
- Indicador de cambio en el patrón de sueño.
- Tipo de publicación en redes sociales preferida.
- Indicador positivo o negativo del impacto de redes sociales
- Indicador de utilidad de las redes sociales



- Descriptor de si usa redes sociales para compra o venta de productos.
- Frecuencia en la que usa las redes sociales más de lo que pretendía.

Se utilizó la herramienta Rapid Miner, Mierswa y otros (2006) [16], para el procesamiento de los resultados y se obtuvieron los siguientes componentes principales los cuales se muestran en la tabla 1, donde de las 19 variables analizadas en la encuesta, el método de PCA logra resumir en los primeros componentes principales la mayor proporción de la variación de los datos, donde el componente principal PC 1 por ejemplo, logra describir %19.4 de la variación. De esta forma, se logra resumir la información en un %80 con apenas los primeros 8 componentes principales. Lo que caracteriza la información y a su vez reduce las 19 variables iniciales a solo 8, una reducción del 57,8%. Por conveniencia, se decide utilizar solo los 8 primeros componentes para el análisis de segmentación.

TABLA 1. COMPONENTES PRINCIPALES

Componente Principal	Proporción de Varianza	Varianza acumulada
PC 1	0.194	0.194
PC 2	0.131	0.325
PC 3	0.122	0.447
PC 4	0.113	0.56
PC 5	0.073	0.633
PC 6	0.063	0.696
PC 7	0.052	0.748
PC 8	0.048	0.796
PC 9	0.045	0.841
PC 10	0.036	0.877
PC ... 19	...	0.999

D. Principales rasgos de los datos encuestados (PCA)

Los componentes principales ofrecen los valores que, asociados a cada una de las variables de la encuesta utilizada, identifican el rasgo subyacente de la información que el componente está resumiendo.

Así, el PC1 caracteriza estudiantes que mayormente indican que en redes sociales venden, compran, publican fotos y consideran las noticias como de utilidad. A su vez, por el lado negativo, son estudiantes contrarios a indicar "no" en la variable de cambio de hora de dormir, en otras palabras, que si han modificado su horario de sueño atribuido al uso de redes sociales. Con esta información, el



CONGRESO
INTERNACIONAL
DE INVESTIGACIÓN
E INNOVACIÓN
MULTIDISCIPLINARIO 2022
ISSN 2448-6035



PC1 se puede caracterizar como un rasgo de usuario activo de redes sociales que percibe de utilidad las noticias sobre COVID-19 y alteró sus hábitos de sueño. De forma análoga, la información de PCA permite caracterizar los demás rasgos principales y brinda descriptores consolidados de los patrones observados en los estudiantes mediante distintas variables.

E. Clustering

Por medio de la herramienta Rapid Miner también, se utilizaron los 8 componentes principales como variables para utilizar en el algoritmo de segmentación X-Means, el cual resuelve utilizar una segmentación óptima de $K = 2$ (parámetro óptimo), o sea, dos clústeres que segmentan de la mejor manera la muestra de estudiantes que participó en la encuesta.

Al ejecutar el algoritmo, la herramienta provee información sobre los centroides, que son una descripción del centro o valor más característico de cada clúster, lo cual permite describir los dos grupos segmentados de estudiantes.

F. Análisis de Clusters

Con la información de los centroides se pueden describir los clústeres como:

- Clúster 1: usuario extensivo de redes sociales que usualmente publica su estado, que compra y vende en esas plataformas y cambió su hora de dormir a raíz del COVID-19.
- Clúster 2: usuario menos extensivo de las redes, no comercia, considera las noticias de COVID-19 de utilidad y no se ve afectado en sus horarios de sueño con respecto a su consumo de noticias.

Adicionalmente, el hecho de haber reducido las dimensiones por medio de PCA, permite visualizar la segmentación en un espacio multidimensional (con N variables) la separación espacial de los clústeres en sus dos componentes principales con mayores diferencias, que de otra forma no hubiese sido posible visualizar.

Finalmente, al analizar los clústeres se puede ver una segmentación de 42 estudiantes en el Clúster 1 y 58 estudiantes en el Clúster 2, lo cual refleja una segmentación casi simétrica de la información, o una polarización de la población estudiantil en cuanto a su respuesta en comportamiento con respecto al uso de redes sociales relacionado con el efecto del COVID19.



Figura 1: Visualización de Clústeres

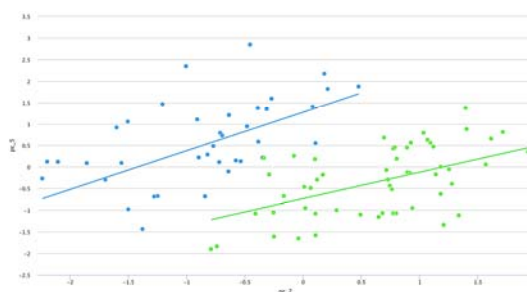


Figura 1

Fuente: Elaboración propia por medio del software Rapid Miner

IV. Conclusiones

Las redes sociales juegan un papel significativo en la población estudiantil representada en el estudio, tanto por el acceso a internet como penetración de uso de redes sociales. Los cambios de hábitos en el contexto de la educación virtual también generan interés en medir el impacto en otras áreas asociadas a un uso más extensivo de internet y su afectación, particularmente por fenómenos sociales a raíz de la situación sanitaria causada por la COVID-19.

La reducción de dimensiones por medio de PCA en este tipo de análisis de resultados permite identificar los mayores componentes que resumen la variación, no solo para caracterizar los rasgos de la muestra sino como insumo para segmentación.

La segmentación que permite el algoritmo X-Means identifica los grupos sobre los cuales se podrían realizar acciones concretas, así como una descripción de segmentos que en este caso muestra una polarización entre la población con respecto al impacto de la pandemia y el uso de las redes sociales.

La aplicación de estas técnicas combinadas sugiere una vía para el análisis efectivo de comportamiento y datos de encuestas, sobre todo cuando la cantidad de variables o dimensiones se vuelve extensiva y no brinda información a priori de los factores predominantes de la población, así como de los grupos afines subyacentes.



V. Bibliografía

- [1] Real Academia Española, Available in <https://dle.rae.es/red?m=form>
- [2] B. Acuña and M. Vega, "Gestión de la emoción en la comunicación mediada por ordenador (CMO)", *Estudios sobre el Mensaje Periodístico*, 2013, 19, 905-913. http://dx.doi.org/10.5209/rev_ESMP.2013.v19.42174
- [3] M. García, J. Seco and M. Del Hoyo, M, "La participación de los jóvenes en las redes sociales. Análisi: Quaderns de comunicació i cultura", 2013, 48, 95-110. https://ddd.uab.cat/pub/analisi/analisi_a2013m9n48/analisi_a2013m9n48p95.pdf
- [4] ABC, "Adictos a las redes sociales y muy vulnerables: así son los niños españoles en internet", 2016. [Online]. Available: <http://www.abc.es/tecnologia/redes/abci-adictos-redes-sociales>
- [5] El Financiero, "La conectividad en Costa Rica", 2013. [Online]. Available, https://www.elfinanciero.cr.com/gnfactory/ELF/INF/2013/09/26/0001/index.html#mod_usoInternetParte1
- [6] La Nación, "Costa Rica es el país latinoamericano que más utiliza WhatsApp y Facebook", 2013, [Online]. Available, <https://www.nacion.com/tecnologia/redes-sociales/costa-rica-es-el-pais-latinoamericano-que-mas/SVJOPNN5ZRCVZKQWAZYC3A6JE4/story/>
- [7] Statista "Number of global social network users 2010-2020", 2016, [Online]. Available: <http://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>
- [8] N. Lam- Figueroa, H. Contreras-Pulache, E. Mori-Quispe, , M. Nizama-Valladolid, C. Gutiérrez, W. Hinostroza-Camposano, W. Hinostroza-Camposano, "Adicción a Internet: Desarrollo y Validación de un instrumento en escolares adolescentes de Lima", *Peru. Per Med Exp Salud Publica*, XXVIII (3), 2011 462-469
- [9] K. Rohrs, "Audience: Marketing in the Age of Subscribers, Fans & Followers". Hoboken NJ: John Wiley & Sons, 2014
- [10] K. Young, "Internet Addiction: The emergence of as new clinical disorder". *Cyberpsychology & Behavior*, 1998, 1, 237-244.
- [11] M. Ecurra and E. Salas Blas, Edwin, "Construcción y Validación del Cuestionario de Adicción a redes sociales (ARS)", *Liberabit. Revista Peruana de Psicología*, 20(1), 73-91, 2014, ISSN: 1729-4827. , [Online]. Available: <https://www.redalyc.org/articulo.oa?id=686/68631260006>
- [12] G. Israel, "Determining Sample Size. University of Florida Cooperative Extension Service", *Institute of Food and Agriculture Sciences, EDIS*, Florida, 1992.
- [13] H. Abdi, and L. Williams, "Principal component analysis" *WIREs Comp Stat*, 2010, 2: 433-459. , [Online]. Available: <https://doi.org/10.1002/wics.101>
- [14] S. Kolenikov and G. Angeles, "The use of discrete data in PCA: theory, simulations, and applications to socioeconomic indices". 2004. *Measure Evaluation* October 20th.
- [15] D. Pelleg, and A. Moore, "X-means: Extending K-means with Efficient Estimation of the Number of Clusters. *Machine Learning*", 2002.
- [16] I. Mierswa, M. Wurst, R. Klinkenberg, M. Scholz, and T. Euler, "YALE: rapid prototyping for complex data mining tasks". In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '06)*. 2006. Association for Computing Machinery, New York, NY, USA, 935-940. DOI: <https://doi.org/10.1145/1150402.1150531>