Una comparación empírica de algoritmos de aprendizaje automático versus aprendizaje profundo para la detección de noticias falsas en redes sociales

An empirical comparison of machine learning versus deep learning algorithms for fake news detection in social networks

Luis Rojas Rubio[1*] Claudio Meneses Villegas[1]

Recibido 06 de enero de 2022, aceptado 17 de mayo de 2022
Received: January 06, 2022 Accepted: May 17, 2022

RESUMEN

Las redes sociales se han convertido en uno de los principales canales de información del ser humano debido a la inmediatez e interactividad social que ofrecen, permitiendo en algunos casos publicar lo que cada usuario considere pertinente. Esto ha traído consigo la generación de noticias falsas o Fake News, publicaciones que solo buscan generar incertidumbre, desinformación o sesgar la opinión de los lectores. Se ha evidenciado que el ser humano no es capaz de identificar en su totalidad si un artículo es realmente un hecho o bien una Fake News, debido a esto es que surgen modelos que buscan caracterizar e identificar artículos basados en minería de datos y machine learning. Este artículo compara empíricamente distintos esquemas de machine learning y deep learning en la tarea de identificar fake news. Para ello se utilizan conjuntos de datos extraídos desde el estado del arte. Los resultados obtenidos en base a la técnica de muestreo utilizado y la representación vectorial *Tf-Idf* del corpus, indica una mejora significativa en el accuracy en contraste a los resultados obtenidos en el estado del arte considerando el repositorio FakeNewsNet.

Palabras clave: Fake News, fakeNewsNet, redes sociales, *tf-idf*, clasificación.

ABSTRACT

Social networks have become one of the leading information channels for human beings due to the immediacy and social interactivity they offer, allowing, in some cases, to publish what each user considers relevant. This usage has brought with it the generation of false news or Fake News, publications that only seek to generate uncertainty, misinformation, or skew the readers' opinion. It has been shown that the human being is not able to fully identify whether an article is actually a fact or a Fake News; due to this, models that seek to characterize and identify articles based on data mining and machine learning emerge. This article empirically compares different machine learning and deep learning schemes to identify fake news; data sets extracted from state of the art are used to accomplish this. The results obtained based on the sampling technique used and the Tf-Idf vector representation of the corpus indicate a significant improvement in accuracy in contrast to the results obtained in the state of the art considering the FakeNewsNet repository.

Keywords: Fake News, fakeNewsNet, social networks, tf-idf, classification.

[1] Universidad Católica del Norte. Departamento de Ingeniería de Sistemas y Computación. Antofagasta, Chile.
E-mail: larr014@gmail.com; cmeneses@ucn.cl
* Autor de correspondencia: larr014@gmail.com

INTRODUCCIÓN

Desde el inicio de los tiempos el ser humano ha buscado la forma de comunicarse con el resto, ya sea por sonidos, gestos o alguna manera más gráfica como dibujos o símbolos, teniendo como problema su susceptibilidad a sesgos intrínsecos a cada individuo. En el contexto de las comunicaciones nos encontramos con las noticias falsas o *Fake News* (FN) cuya intención principal, entre otras, es generar incertidumbre, desinformación e influir en la opinión de quien las lee [1].

En el último tiempo, a raíz del estallido social registrado en octubre del 2019 en Chile y a nivel mundial con el Covid-19, la cantidad de FN registradas en redes sociales han aumentado considerablemente, sumado al hecho que la credibilidad de los medios oficiales ha disminuido y por lo tanto su consumo [2]. Por otra parte, las redes sociales como Facebook o WhatsApp han aumentado en relación a su consumo y confianza [2].

De acuerdo al estudio "On Deception and Deception Detection: Content Analysis of Computer-MediatedStated Beliefs" realizado por Victoria Rubin [3], los seres humanos tenemos una precisión de un 54% al discernir la veracidad de un artículo. Ante esto, los algoritmos relacionados al lenguaje natural son idóneos para el análisis y reconocimiento de la veracidad de una publicación.

En este artículo se comparan dos enfoques de *machine learning*: el enfoque clásico cuya etapa de selección de características (*feature selection*) es fundamental para obtener mejores resultados, en contraste al enfoque de *deep learning* donde la complejidad interna del modelo se encarga de codificar y decodificar las características que no son obvias. El contraste de ambos enfoques surge a partir de los resultados que se evidencian en el estado del arte, donde posicionan a los algoritmos de *deep learning* con resultados prometedores en tareas de clasificación; sin embargo, estos algoritmos funcionan como una caja negra, es decir, se desconoce la forma en la cual está clasificando. Por otra parte, los algoritmos de *machine learning* tradicionales generalmente no ofrecen los mejores resultados y los modelos que generan suelen tener problemas de escalabilidad; sin embargo, si es posible inferir como está clasificando. El desarrollo de este tipo de comparativas permite ofrecer una base inicial para seleccionar objetivamente algoritmos a implementar en trabajos futuros relacionados a la clasificación de autenticidad de una publicación. Más aún, este estudio puede servir como base para realizar mejoras tanto a los algoritmos utilizados y sus configuraciones, el preprocesamiento, incorporación de nuevas características, fusiones de distintos conjuntos de datos, entre otros. En materia de políticas públicas en Latinoamérica, existe una carencia sustancial de normativas respecto al uso de internet, por lo que desarrollar métodos que permitan caracterizar una *fake news* aporta a la verificación de publicaciones mediante *fact checkers*, las mismas redes sociales o directamente a los mismos usuarios.

El artículo se estructura como sigue. En la primera sección se presentan aspectos generales sobre las *Fake News*, en la segunda sección se abordan antecedentes respectos a eventos relacionados a *fake news* y conceptos afines. En la tercera sección se presenta una revisión del estado del arte. En la cuarta sección se da a conocer el desarrollo del trabajo donde se aborda el conjunto de datos utilizado, método de recopilación de *Tweets*, preparación del conjunto de datos, representación del corpus y muestreo. En la quinta sección se exponen los resultados obtenidos y la discusión de los mismos. Finalmente, se presentan las conclusiones del trabajo realizado.

ANTECEDENTES

Eventos Relevantes

El origen de las *Fake News* no es reciente, si consideramos que una *Fake News* es una información que no puede ser contrastada con la realidad de manera oportuna, podría datar de la Antigua Grecia [4-5] mediante la retórica persuasiva. Sin embargo, si la relacionamos a noticias estas comienzan con la imprenta en 1439 [6]. En el siglo XIX gracias al rápido crecimiento de los periódicos ayudado por las tecnologías emergentes de la época, se logran dar los primeros atisbos de las *Fake News*, un hecho histórico surgió en el año 1835 donde el New York Sun alertaba sobre la existencia de la vida en la Luna [7] o la transmisión realizada por la emisora CBS la noche del 30 de octubre de 1938, donde fue dramatizada la novela "La Guerra de los Mundos" de H.G. Wells [8]. Esta dramatización generó el pánico en millones de radioescuchas que creyeron que lloverían meteoritos con naves alienígenas

en su interior para someter al planeta con rayos y gases venenosos.

Otras noticias han generado mayores problemas como en 1844 en los periódicos anticatólicos de Filadelfia, donde afirmaron falsamente que los irlandeses robaban biblias en las escuelas públicas, provocando una serie de disturbios violentos y ataques contra las iglesias católicas [9]. Un caso más reciente a raíz del coronavirus fue la teoría de que las redes 5G expandieron el virus, en Reino Unido fue tal el revuelo que sus ciudadanos boicotearon y sabotearon algunas antenas de telefonía [10], según el medio "The Guardian" se registraron más de veinte incendios y destrozos en antenas telefónicas [11].

En pleno siglo XXI gracias a los avances tecnológicos se han digitalizado la mayoría de los canales informativos, sin embargo, han sido las redes sociales las principales fuentes de información, esto también ha aumentado la cantidad de *Fake News*. Se han hecho evidentes campañas de desinformación a gran escala relacionadas con el cambio climático, las vacunas, los transgénicos, la nutrición, el origen de la vida, la salud, el origen de enfermedades o el impacto de la inmigración. Pero el hecho más relevante sucedió el año 2016, dando cuenta de una campaña de difamación para el proceso electoral de Estados Unidos, donde se generaron un total de 115 *Fake News* favorables a Donald Trump compartidas 30 millones de veces, en contraste a 41 *Fake News* compartidas 7.6 millones de veces beneficiando a Hillary Clinton [12]. De acuerdo con BrandWatch [13] en el 2020 se han registrado más de 400.000 artículos relacionados al cambio climático y al coronavirus, con publicaciones que registran ser compartidas más de 4.2 millones de veces solo en Facebook.

Terminología

Uno de los problemas aún presentes en la clasificación de *Fake News* es la identificación de la misma, debido a lo general de su definición. De acuerdo con el conjunto de definiciones que ofrece Oxford, una *fake news* es un reporte falso de un evento en una página web, alguna noticia que es difícil de verificar y que crea confusión pública sobre un evento [14]. Según el diccionario de Cambridge es una historia falsa que aparenta ser una noticia, se propaga en Internet o en otros medios digitales, usualmente creados para influenciar políticamente [15].

En el estado del arte podemos encontrar algunos conceptos con los cuales se relaciona una *Fake News*, los que se detallan a continuación para posteriormente ser comparados en relación a su autenticidad e intención (ver Tabla 1).

- Satire: Forma de criticar a las personas de manera humorística, generalmente utilizado en temas políticos. Se caracteriza por el uso de ironías. No tiene una mala intención [16].
- Disinformation: Difusión de información falsa cuyo único fin es engañar a la gente [17].
- Misinformation: Falta de información sobre un tema, generalmente utilizado para generar desinformación [18].
- Hoax: Busca farsear sobre una temática con el fin de engañar a alguien [19].
- Rumor: Noticia no verificada la cual se difunde rápidamente de persona a persona [20].

ESTADO DEL ARTE

El uso de redes sociales y la presencia de *Fake News* han generado un impacto negativo en la sociedad [21], generando incertidumbre y desinformación, anexo a esto se ha evidenciado que el ser humano no es capaz de identificar en su totalidad si un artículo es realmente un hecho o bien un *Fake News*, debido a esto es que surgen modelos que buscan caracterizar y predecir la veracidad de artículos o noticias, basados en métodos de minería de datos y *machine learning*.

El estudio de publicaciones consideradas como *Fake News* aún resulta ser un desafío, diversos trabajos proponen desde la caracterización hasta la predicción e intervención de la misma. Algunos de los trabajos relacionados detallan a continuación [1].

En [22] se utiliza un modelo supervisado basado en el clasificador bayesiano dando cuenta que con un

Tabla 1. Comparación de conceptos.

	Autenticidad	**Intención**
Satire	Desconocida	Ironizar
Disinformation	Falsa	Engañar
Misinformation	Falsa	Desconocida
Hoax	Desconocida	Desconocida
Rumor	Desconocida	Desconocida

Fuente: Adaptado de [1].

conjunto de datos pequeños y un modelo simple se pueden obtener buenos resultados, en este caso se logran identificar un 71,73% de *Fake News*.

Continuando con modelos supervisados en [23] se caracterizan las *Fake News* en un conjunto de características, a saber, características textuales, que involucran sintaxis, léxico, semántica y subjetividad, características asociadas a la fuente que establecen polaridad y credibilidad de la misma, y características del medio, refiriendo a patrones temporales de las publicaciones y como la red interactúa (*likes*, comentarios, número de veces compartido, entre otros). De esta serie de características se implementan cinco clasificadores: *k-Nearest Neighbor* (KNN), *Naive Bayes* (NB), *Support Vector Machine* (SVM), *Random Forest* (RF) y *XGboost* (XGB), siendo estos dos últimos los de mayor *accuracy* con 85% y 86% respectivamente.

En [24] se presentan dos modelos de clasificación, el primero de ellos aborda una clasificación en base al texto, utilizando NB, SVM, *Convolutional Neural Network* (CNN) y *Recurrent Neural Network* (RNN) con *Long Short Term Memory* (LSTM), siendo estas dos últimas, CNN y LSTM, las de mayor *accuracy* con 92% y 93%, respectivamente. El segundo modelo aborda la clasificación de *Fake News* según el usuario que realiza la publicación, para esto se utiliza el método de probabilidad en base a un historial de publicaciones y un método de ajuste de parámetros, obteniendo un *accuracy* de 75% y 80% respectivamente.

Debido al tamaño del conjunto de datos y lo que puede aportar cada clasificador de manera independiente, en [25] se propone un modelo de aprendizaje en conjunto, el cual toma un conjunto de datos y los divide entre cinco clasificadores: SVM, CNN, LSTM, KNN y NB. El estudio permite dar cuenta que en una clasificación conjunta se pueden obtener mejores resultados que de manera independiente y/o mejores caracterizaciones. Para este modelo se obtiene un *accuracy* de un 85%.

En [27] se utilizan tres modelos de redes neuronales: *Deep Neural Network* (DNN), CNN y RNN. La finalidad del estudio reside en las variaciones respecto a la representación vectorial de entrada, a saber, *Bag of Words* (BoW), *Term Frequency-Inverse Document Frequency* (*TF-IDF*) y *Word2Vec* (W2V). Para un modelo DNN se obtiene que la representación vectorial, cuyo mejor resultado obtenido de *accuracy* corresponde a *TF-IDF* (94,31%), seguido por BoW (89,23%) y por último W2V (75,67%).

Utilizando un conjunto de datos relacionado al terremoto del año 2010 ocurrido en el sur de Chile, en [28] se evidencia como los modelos tradicionales de *Machine Learning* ofrecen mejores resultados que los modelos de *Deep Learning*. El modelo SVM y NB alcanzan los mejores resultados, con 89,34% y 89,06% de *accuracy* respectivamente.

Con el fin de mejorar la clasificación de noticias falsas se han implementado modelos neuronales híbridos, [29] se presenta una red CNN+BiLSTM con la cual alcanzan un accuracy de 88% para publicaciones obtenidas de Twitter. Así mismo en [30] se implementa una red LSTM-CNN para un conjunto de datos que incorpora imágenes y texto, la red obtiene un accuracy de 80,38%, en contraste a una red LSTM (82,29%) y una red LSTM con dropout(73,78%).

DESARROLLO

Datasets

FakeNewsNet es un repositorio que a diferencia de otros conjuntos de datos como BuzzFeed, LIAR o CREDBANK cuenta con tres aspectos fundamentales en el estudio de Fake News: (1) contenido de la noticia, tanto lingüístico como visual; (2) un contexto social que abarca usuarios, publicaciones, respuestas y su red; y (3) información espacio temporal. Este conjunto de datos se conforma de dos conjuntos de datos, "PolitiFact" y "GossipCop" abordando el dominio de la política y rumores relacionados a celebridades. Para el desarrollo de este trabajo se utiliza GossipCop el cual cuenta con 16.817 temáticas asumidas como *noFake* y 5.323 temáticas como *Fake*, cada una de estas temáticas está asociada a múltiples identificadores de Twitter a modo de respetar sus políticas y no exponer directamente las publicaciones.

Recopilación de Tweets

Tal como se mencionó en el apartado anterior, GossipCop contiene temáticas relacionadas a un conjunto de identificadores con los cuales es posible recuperar una publicación en específico de Twitter. Para esto es necesario contar con credenciales de

acceso y uso de una API, por simplicidad se decide utilizar Tweepy logrando recuperar 1.250.954 *noFake Tweets* y 465.309 *Fake Tweets*.

Preparación del Conjunto de Datos

Dada la cantidad de Tweets recuperados se realiza una preparación del conjunto de datos (Ver Figura 1), iniciando por la eliminación de URLs ya que estas no aportan elementos incidentes a la clasificación. Posterior a ello, se procede a eliminar los duplicados reduciendo en dos tercios el conjunto de datos.

Para concluir la preparación de los datos se convierte todo a minúsculas, se eliminan las *stopwords* en inglés, se aplica el método de *Stemming* de Porter a modo que cada palabra vuelva a su raíz, para finalmente volver a eliminar los duplicados.

Representación del Corpus

Existen diversos métodos de representación para trabajar con textos, los más comunes son del tipo vectorial, los cuales se detallan a continuación:

- **One hot Encoding**: Construye a partir de un corpus de documentos un vector binario cuyo tamaño es igual al número de palabras únicas de dicho corpus. Esta representación solo indica la presencia o no de una palabra en un documento.
- **Bag of Words (BoW)**: Construye a partir de un corpus de documentos un diccionario en el cual contabiliza cuantas veces aparece dicha palabra en cada documento. Esta representación evidencia la frecuencia de una palabra en un documento.
- **Word2Vec**: Es un modelo cuyo objetivo es entrenar una red neuronal utilizando un corpus de documentos para que, dada una palabra *w*, indique la probabilidad que una palabra sea vecina de *w* según una ventana (número de palabras cercanas).
- ***Tf-Idf***: Esta representación se compone de dos partes: la primera Tf que representa la frecuencia de un término con respecto a un documento, al igual que BoW, y una segunda parte Idf, la cual busca discriminar también la rareza de una palabra con respecto a la totalidad del corpus de documentos, esto hace que su representación vectorial indique la frecuencia de una palabra en un documento y cuán importante es en relación con una serie de documentos que conforman el corpus. Dado una colección de documentos *D*, una palabra *w* y un documento *d* existente en *D*, *Tf-Idf* puede ser calculado como se aprecia en la ecuación 1 [26]. Donde $f_{w,d}$ es el número de veces que *w* aparece en *d*, |*D*| el tamaño del corpus y $f_{w,D}$ el número de documentos en los cuales *w* aparece en *D*.

$$TfIdf\left(d\right) = f_{w,\,d} * \log\left(\frac{|D|}{f_{w,\,D}+1}\right) \qquad (1)$$

Se opta por representar el corpus mediante un vector *Tf-Idf*. En [27] se realiza una comparativa entre las principales formas de representación, es decir *Tf-Idf*, BoW, y Word2Vec. La comparación final se realiza en contraste a los mejores resultados obtenidos en distintas variaciones a los modelos, con los cuales la representación Tf-Idf se posiciona como la que obtiene mejor *accuracy* (94,31%), seguido de la representación BoW con 92,46% de *accuracy*, finalmente, Word2Vec obtiene 88,46% de *accuracy*, siendo la de *accuracy* más bajo.

Para la vectorización del corpus se comienza por no limitar el tamaño del diccionario, pero debido a la cantidad de palabras y documentos con las cuales se cuentan se opta por limitarlo, estableciendo de manera empírica 20.000 palabras.

Muestras de datos

Debido a la vectorización, el número de documentos y la capacidad de cómputo, se opta por dividir el conjunto de datos. En primera instancia se obtiene de

Fuente: Elaboración propia.

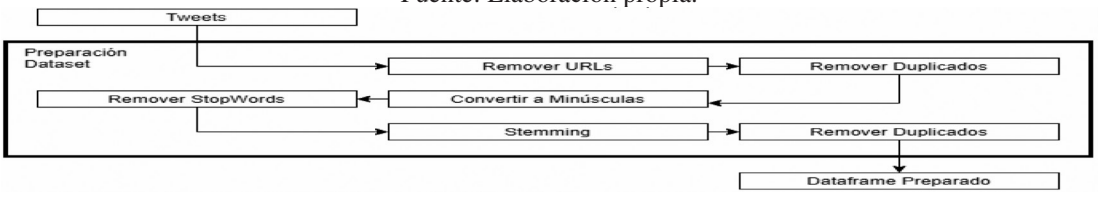

Figura 1. Proceso de preparación del conjunto de datos.

manera aleatoria pero estratificada el 70% del total de los datos, de esta muestra se obtienen 3 muestras distintas de un 50% las cuales son divididas en un 49% para el conjunto de entrenamiento con 73.567 publicaciones, 30% para el conjunto de validación con 45.042 publicaciones y finalmente 21% para el conjunto de prueba con 31.529 publicaciones (Ver Figura 2).

En la Figura 3 se muestra un gráfico de nube en término de las palabras más frecuentes en cada una de las muestras. En tanto la Figura 4, Figura 5 y Figura 6 evidencian las 20 palabras de mayor frecuencia en cada una de las muestras posterior a la eliminación de *StopWords*. Se puede apreciar en base a las figuras antes descritas que las muestras presentan las mismas palabras, variando solo en la frecuencia con la cual se presentan en cada muestra. Esto demuestra que el método de muestreo utilizado representa la totalidad del conjunto y, por ende, carece de sesgo.

Métricas de Rendimiento

Para poder contrastar los resultados se usan las siguientes métricas de rendimiento: *accuracy* (Ecuación 2), *precisión* (Ecuación 3), *recall* (Ecuación 4) y *F1-Score* (Ecuación 5); donde *TP* es la cantidad de verdaderos positivos, *TN* es la cantidad de verdaderos negativos, *FP* es el número de falsos positivos y FN es el número de falsos negativos.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FN} \tag{3}$$

$$recall = \frac{TP}{TP + FN} \tag{4}$$

$$F1 - Score = \frac{2 * precision * recall}{precision + recall} \tag{5}$$

RESULTADOS

Cada uno de los algoritmos implementados están configurados con sus parámetros predeterminados de acuerdo a *scikit-learn* excepto por los que se

Fuente: Elaboración propia.

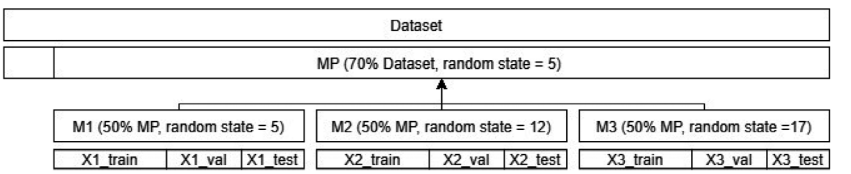

Figura 2. Partición de los datos en muestras M1, M2 y M3.

Fuente: Elaboración propia.

Figura 3. Términos más frecuentes en cada muestra.

Fuente: Elaboración propia.

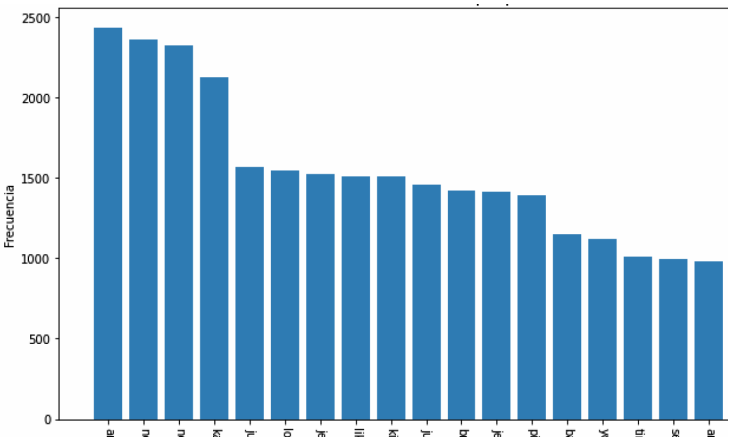

Figura 4. Top 20 palabras de mayor frecuencia - Muestra 1.

Fuente: Elaboración propia.

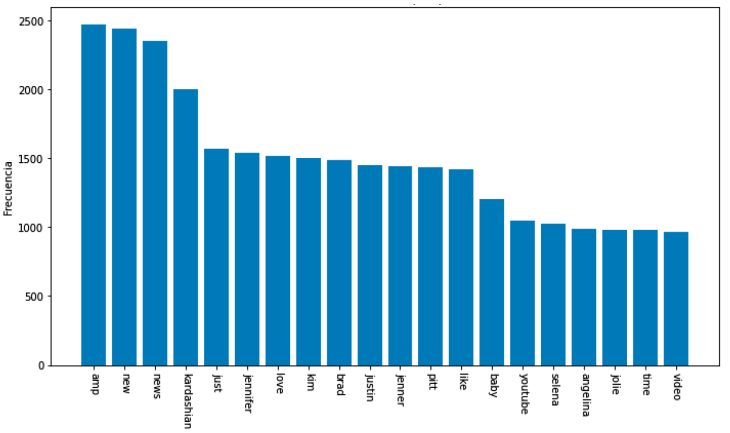

Figura 5. Top 20 palabras de mayor frecuencia - Muestra 2.

Fuente: Elaboración propia.

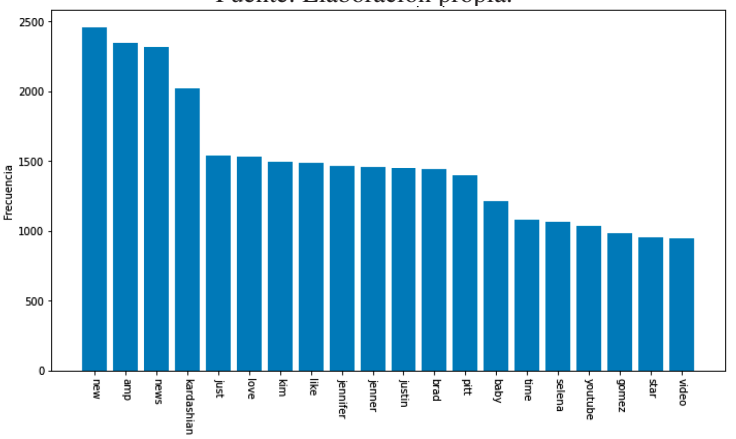

Figura 6. Top 20 palabras de mayor frecuencia - Muestra 3

muestran en la Tabla 2. Por otra parte, los algoritmos de aprendizaje profundo como la red densa se componen de una capa oculta de 1000 neuronas y una capa de salida de 2 neuronas, con una función de activación *relu* y *softmax*, respectivamente. En tanto la red LSTM se configura con una capa oculta de 100 neuronas y una capa de salida de 2 neuronas con una función de activación *softmax*. Cabe destacar que ambas redes fueron construidas usando *keras* y fueron configuradas empíricamente bajo un entrenamiento de 20 épocas (Ver Figura 7), seguido de un ajuste en base al conjunto de validación, para finalmente evidenciar sus resultados mediante el conjunto de prueba.

Tabla 2. Configuración de parámetros de los algoritmos utilizados.

Algoritmo	Parámetro	Valor
Perceptron	Tol	1e-3
	random_state	0
	Verbose	2
MLP	random_state	0
	Verbose	2
RF	random_state	0
	n_estimators	100
	n_jobs	8
	Verbose	2
DT	random_state	0
KNN	n_neighbors	2
	Algorithm	kd_tree
SVM	learning_rate	Constant
	eta0	0,1
	n_jobs	4

Fuente: Elaboración propia.

Fuente: Elaboración propia.

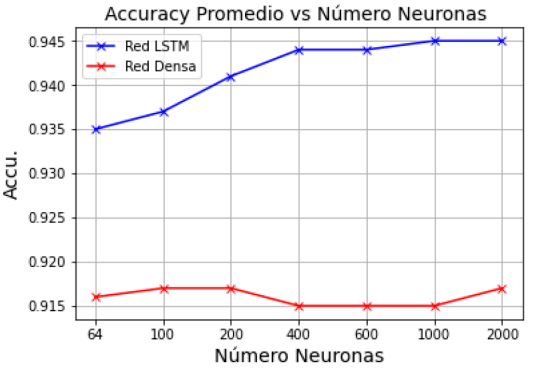

Figura 7. Accuracy promedio vs Número de neuronas.

En la Tabla 3 se pueden apreciar los resultados de cada algoritmo según la muestra utilizada. En negritas se puede apreciar el *accuracy* más alto por modelo, y subrayado los tres *accuracy* más altos, evidenciando el buen desempeño de la red densa, *Random Forest* y por último la red LSTM. En tanto a la Figura 8 se puede observar de manera gráfica como las tres muestras presentan un *accuracy* con diferencias de ± 0,006 en cada uno de los algoritmos, a excepción del *accuracy* del Perceptrón en la muestra 2 y la máquina de vector de soporte (SVM) en la muestra 3. Lo que demuestra que la diferencia de los resultados se basa en las características de cada algoritmo y no de la técnica de muestreo. En general la muestra 1 logra los mejores resultados.

En la Tabla 4 se evidencian los promedios de clasificación por algoritmo, subrayados se pueden apreciar los tres mejores resultados por métrica, y

Tabla 3. Resultados por modelo.

	Modelo 1				Modelo 2				Modelo 3			
	Acc	**P**	**f1**	**Recall**	**Acc**	**P**	**f1**	**Recall**	**Acc**	**P**	**f1**	**Recall**
Perceptron	**0,857**	0,858	0,857	0,870	0,793	0,860	0,793	0,749	0,854	0,857	0,854	0,868
MLP	0,915	0,914	0,915	0,910	0,914	0,912	0,914	0,910	**0,916**	0,913	0,916	0,913
RF	**0,931**	0,931	0,931	0,926	0,928	0,928	0,928	0,924	0,930	0,930	0,930	0,926
DT	0,864	0,859	0,864	0,860	0,865	0,860	0,865	0,862	**0,866**	0,862	0,866	0,862
KNN	0,879	0,909	0,879	0,855	0,885	0,914	0,885	0,862	**0,881**	0,912	0,881	0,856
SVM	**0,890**	0,887	0,890	0,887	0,887	0,891	0,887	0,876	0,837	0,874	0,837	0,806
NB	0,898	0,898	0,898	0,890	0,900	0,900	0,900	0,893	**0,901**	0,901	0,901	0,894
LSTM	**0,919**	0,917	0,915	0,919	0,916	0,914	0,914	0,916	0,917	0,915	0,915	0,917
Dense	**0,946**	0,946	0,942	0,946	0,945	0,946	0,941	0,945	0,945	0,944	0,942	0,945

Fuente: Elaboración propia.

Fuente: Elaboración propia.

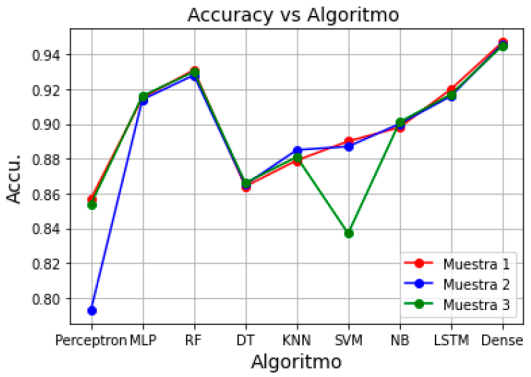

Figura 8. Accuracy vs Algoritmos.

Tabla 4. Promedio por muestras.

	Acc.	**Precision**	**f1**	**Recall**
Percetron	0,835	0,858	0,835	0,829
MLP	0,914	0,913	0,914	0,910
RF	0,930	0,930	0,930	0,925
DT	0,865	0,860	0,865	0,861
KNN	0,882	0,912	0,882	0,858
SVM	0,871	0,884	0,871	0,856
NB	0,900	0,900	0,900	0,892
LSTM	0,917	0,915	0,915	0,917
Dense	**0,945**	**0,945**	**0,942**	**0,945**

Fuente: Elaboración propia.

en negrita el mejor resultado de cada una. De dicha tabla también se puede observar que los algoritmos tradicionales de *machine learning* ofrecen un *accuracy* entre 0,865 para un modelo de árbol de decisión (DT) y hasta un 0,931 de *accuracy* para el algoritmo Random Forest (RF), estos resultados se le pueden atribuir a que el conjunto de entrada se basa en un vector *TF-IDF* compuesto por un vocabulario de 20.000 palabras, donde ciertas palabras de dicho vector pudiesen no estar aportando a la clasificación o incluso generando ruido. Con respecto a los algoritmos de Deep Learning se obtienen desde 0,914 de *accuracy* para el perceptrón multi capa (MLP) y hasta un 0,945 de *accuracy* para un modelo neuronal denso (Dense), en este caso el tamaño y las palabras del vocabulario no influyen, ya que el mismo algoritmo de *deep learning* es capaz de discriminar y autoajustarse.

Tomando en cuenta la matriz de confusión de los tres algoritmos con mejores resultados (Red Densa, Random Forest y Red LSTM), la red Densa en promedio presenta un error de 7.87% en el etiquetado de publicaciones *noFake* (Ver Figura 9) en contraste al mismo etiquetado con el algoritmo Random Forest (Ver Figura 10) con un error promedio de 9,91% y el modelo de red LSTM (Ver Figura 11) con 10,22% de error. Con respecto al etiquetado de publicaciones *fake*, la

Fuente: Elaboración propia.

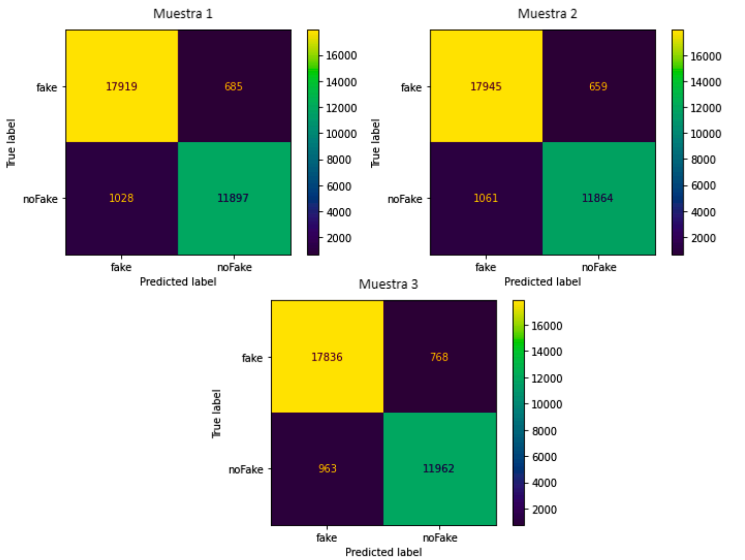

Figura 9. Matriz de confusión Red Densa por muestra.

Fuente: Elaboración propia.

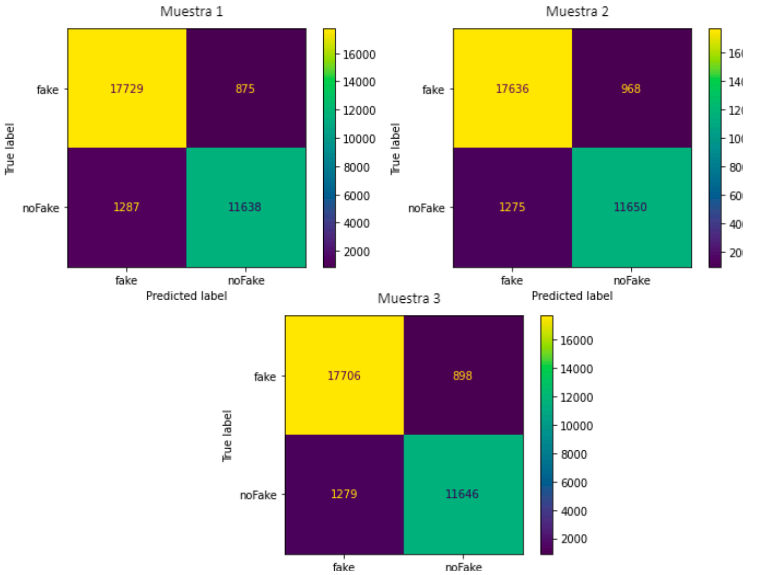

Figura 10. Matriz de confusión Random Forest por muestra.

Fuente: Elaboración propia.

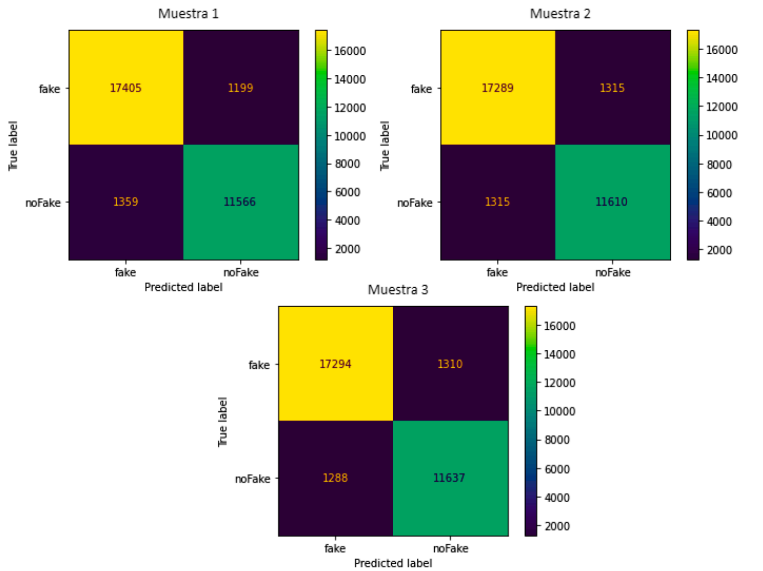

Figura 11. Matriz de confusión red LSTM por muestra.

red Densa obtiene un error promedio de 3,78% en comparación al algoritmo Random Forest y la red LSTM, que en promedio obtienen un error de 4,91% y 6,85% respectivamente. Es decir, el algoritmo de red Densa tiene un mayor *accuracy* tanto en el etiquetado de publicaciones *noFake* como *Fake*, en tanto el algoritmo de red LSTM obtiene el menor *accuracy* de los tres modelos.

En [31] se realiza un análisis desde 3 perspectivas, la primera de ellas aborda el contenido de las publicaciones, para lo cual se utilizan SVM, Regresión logística, NB

y CNN con las configuraciones predeterminadas de la librería *scikit-learn* y sin sintonización de parámetros. Adicionalmente se incluye SAF/S, un modelo presentado en [32] por los mismos autores, en los cuales se utiliza *autoencoder* para aprender *features* de artículos de noticias para clasificar nuevos artículos como falsos o reales. La segunda perspectiva utiliza el contexto social a través del modelo SAF/A, el cual es una variación del modelo anterior en que el *autoencoder* aprende *features* relacionadas a la propagación de la noticia. Finalmente, la tercera perspectiva utiliza las dos anteriores mediante el modelo SAF que combina SAF/S y SAF/A utilizando *autoencoder* con celdas LSTM. Cabe destacar que cada uno de los modelos anteriormente descritos utilizan como entrada un vector *one hot encoding*.

De los modelos descritos en el párrafo anterior se obtienen los rendimientos presentes en la Tabla 5, donde el modelo desarrollado con CNN obtiene un *accuracy* de 0,723, siendo el modelo que mejores resultados presenta. Por otra parte, la máquina vector de soporte (SVM) y NB obtienen un *accuracy* de 0,497 y 0,624 respectivamente, en contraste a los resultados presentados en este trabajo donde se obtuvo un *accuracy* de 0,871 para SVM y 0,900 para NB. Anexo a esto, el proceso de clasificación utilizado mediante la red neuronal densa logra obtener un 0,945 de *accuracy*, superando en más de un 30% los resultados obtenidos por los autores del conjunto de datos (0,723 de accuracy modelo CNN) [31].

CONCLUSIÓN

Este artículo se enfoca en la comparativa de algoritmos tradicionales de *Machine Learning* encontraste a algoritmos de *Deep Learning* en la clasificación de publicaciones presentes en el dataset FakeNewsNet, específicamente el conjunto GossipCop utilizando la técnica de vectorización *Tf-Idf*. Realizando la comparativa de los resultados obtenidos por los autores del datasets, se logra obtener mejoras de un 14% en el peor de los casos (Perceptrón) y de un 30% en el mejor de los casos con el modelo de red neuronal densa. La diferencia de estos resultados pudiese situarse por la representación vectorial utilizada, en [31] se utiliza la técnica *one hot encoding* en contraste a la representación *Tf-Idf* de este artículo. Otro punto diferenciador para considerar es el muestreo utilizado, ya que en [31] se aborda la totalidad del conjunto de datos, en comparación a este trabajo donde se realiza un muestreo del conjunto de datos. Sin embargo, como se ha evidenciado en este trabajo, los resultados obtenidos de cada una de las muestras resultan ser representativas de la totalidad del conjunto. Por otra parte, los resultados obtenidos muestran que los algoritmos tradicionales logran entregar sobre 0,8 de *accuracy*, e incluso logran resultados similares a los algoritmos de *Deep Learning*, el modelo generado con Random Forest se posiciona como el segundo modelo de mayor *accuracy*, con una diferencia de 0,015 de *accuracy* por debajo del mejor resultado de *Deep Learning* obtenido con la red neuronal densa.

AGRADECIMIENTOS

Agradecimientos a los autores Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee y Huan Liu que han dispuesto de manera gratuita el acceso al repositorio FakeNewsNet.

REFERENCIAS

[1] X. Zhou and R. Zafarani. “Fake news: A survey of research, detection methods, and

Tabla 5. Rendimiento por modelo FakeNewsNet.

Modelo	Accuracy	Precision	Recall	f1
SVM	0,497	0,511	0,713	0,595
Logistic regression	0,648	0,675	0,619	0,646
Naive Bayes	0,624	0,631	0,669	0,649
CNN	**0,723**	0,751	0,701	0,725
SAF /S	0,689	0,671	0,738	0,703
SAF /A	0,635	0,589	0,882	0,706
SAF	0,689	0,656	0,792	0,717

Fuente: Adaptado de [31].

opportunities". Vol. 53 Issue 5, pp. 1-40. 2018. DOI: 10.1145/3395046.

[2] Comunidad Cadem. "Estudio medios de comunicación". Fecha de consulta: 03 junio de 2021. URL: https://www.cadem.cl/wp-content/uploads/2020/01/Estudio-rol-de-los-medios_VFP.pdf

[3] V. Rubin. "On deception and deception detection: Content Analysis of Computer-Mediated Stated Beliefs". Proceedings of the American Society for Information Science and Technology. Vol. 47 Nº 1, pp. 1-10. 2010. DOI: 10.1002/meet.14504701124.

[4] R. Fonseca and L. Prieto de Alizo. "Las emociones en la comunicación persuasiva: desde la retórica afectiva de Aristóteles". Quorum académico. Vol. 7 Nº 1, pp. 78-94. 2010.

[5] T. de los Santos, E. Smith and M. Cohen. "Targeting truth: How museums can collaboratively address social issues. Journal of museum education". Journal of Museum Education. Vol. 43 Nº 2, pp. 104-113. 2018. ISSN: 1690-7582. DOI: 10.1080/10598650.2018.1457842.

[6] L. Fernández. "Desinformación y comunicación organizacional: estudio sobre el impacto de las fake news". Revista latina de comunicación social. Nº 74, pp. 1714-1728. 2019. ISSN: 1138-5820. DOI: 10.4185/RLCS-2019-1406.

[7] H. Allcott and M. Gentzkow. "Social media and fake news in the 2016 election. Journal of economic perspectives". Journal of economic perspectives. Vol. 31 Nº 2, pp. 211-36. 2017. DOI: 10.1257/jep.31.2.211.

[8] ABC. "El día que Orson Welles sembró el pánico con la guerra de los mundos". Fecha de consulta: 04 de junio de 2021. URL: https://www.abc.es/cultura/20131030/abci-aniversario-orson-welles-guerra-201310300614.html

[9] B. Bharali and A. Goswami. "Fake news: Credibility, cultivation syndrome and the new age media". Media Watch. Vol. 9 Nº 1, pp. 118-130. 2018. ISSN: 2249-8818. DOI: 10.15655/mw/2018/v9i1/49277.

[10] ABC. "La verdad sobre la loca teoría de que las redes 5G expandieron el coronavirus". Fecha de consulta: 04 de junio de 2021. URL: https://www.abc.es/tecnologia/informatica/soluciones/abci-verdad-sobre-loca-teoria-redes-expandieron-coronavirus-202004081505_noticia.html

[11] The Guardian. "5G, coronavirus and contagious superstition". Fecha de consulta: 04 de junio de 2021. URL: https://www.theguardian.com/world/2020/apr/26/5g-coronavirus-and-contagious-superstition

[12] P. Valero and L. Oliveira. "Fake news: una revisión sistemática de la literatura". Observatorio (OBS*). Vol. 12 Nº 5. 2018. DOI: 10.15847/obsOBS12520181374.

[13] Brandwatch. "Brandwatch investigates fake news in 2020". Fecha de consulta: 07 de junio de 2021. URL: https://www.brandwatch.com/reports/fake-news-2020/view/

[14] Oxford Learner's Dictionaries. "Fake-news noun - definition, pictures, pronunciation and usage notes". Fecha de consulta: 07 de junio de 2021. URL: https://www.oxfordlearnersdictionaries.com/us/definition/english/fake-news

[15] Cambridge Dictionary. "Significado de fake news en el diccionario Cambridge inglés". Fecha de consulta: 07 de junio de 2021. URL: https://dictionary.cambridge.org/es/diccionario/ingles/fake-news

[16] Cambridge Dictionary. "Significado de satire en el diccionario Cambridge inglés". Fecha de consulta: 07 de junio de 2021. URL: https://dictionary.cambridge.org/es/diccionario/ingles/satire

[17] Cambridge Dictionary. "Significado de disinformation en el diccionario Cambridge inglés". Fecha de consulta: 07 de junio de 2021. URL: https://dictionary.cambridge.org/es/diccionario/ingles/disinformation

[18] Cambridge Dictionary. "Significado de misinformation en el diccionario Cambridge inglés". Fecha de consulta: 07 de junio de 2021. URL: https://dictionary.cambridge.org/es/diccionario/ingles/misinformation

[19] Cambridge Dictionary. "Significado de hoax en el diccionario Cambridge inglés". Fecha de consulta: 07 de junio de 2021. URL: https://dictionary.cambridge.org/es/diccionario/ingles/hoax

[20] Cambridge Dictionary. "Significado de rumor en el diccionario Cambridge Inglés". Fecha de consulta: 07 de junio de 2021. URL: https://dictionary.cambridge.org/es/diccionario/ingles/rumor

[21] D. Martínez. "Las fake news sobre la covid-19 en plataformas digitales durante las dos primeras semanas de abril 2020 y sus efectos en los estudiantes del décimo semestre de Psicología Clínica de la Universidad Técnica de Babahoyo". Tesis de Magister. Universidad católica de Santiago de Guayaquil. Guayaquil, Ecuador. 2021.

[22] M. Granik and V. Mesyura. "Fake news detection using naive Bayes classifier". 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), pp. 900-903. 2017. DOI: 10.1109/UKRCON.2017.8100379.

[23] J. Reis, A. Correia, F. Murai, A. Veloso and F. Benevenuto. "Supervised learning for fake news detection". IEEE Intelligent Systems. Vol. 34 Nº 2, pp. 76-81. 2019. DOI: 10.1109/MIS.2019.2899143.

[24] P. Harjule, A. Sharma, S. Chouhan and S. Joshi. "Reliability of news". 2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE), pp. 165-170. 2020. DOI: 10.1109/ICETCE48199.2020.9091751.

[25] A. Agarwal and A. Dixit. "Fake news detection: An ensemble learning approach". 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1178-1183. 2020. DOI: 10.1109/ICICCS48265.2020.9121030.

[26] K. Chen, Z. Zhang, J. Long and H. Zhang. "Turning from *TF-IDF* to TF-IGM for term weighting in text classification". Expert Systems with Applications. Vol. 66, pp. 245-260. 2016. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2016.09.009.

[27] A. Thota, P. Tilak, S. Ahluwalia and N. Lohia. "Fake news detection: A deep learning approach". SMU Data Science Review. Vol. 1 Nº 3, pp. 10. 2018.

[28] Abdullah-All-Tanvir, E.M. Mahir, S. Akhter and M.R. Huq. "Detecting fake news using machine learning and deep learning algorithm". 2019 7th International Conference on Smart Computing & Communications (ICSCC), pp. 1-5. 2019. DOI: 10.1109/ICSCC.2019.8843612.

[29] N. Rani, P. Das and A.K. Bhardwaj. "A hybrid deep learning model based on CNN-BiLSTM for rumor detection". 2021 6th International Conference on Communication and Electronics Systems (ICCES), pp. 1423-1427. 2021. DOI: 10.1109/ICCES51350.2021.9489214.

[30] O. Ajao, D. Bhowmik and S. Zargari. "Fake news identification on twitter with hybrid CNN and RNN models". In Proceedings of the 9th international conference on social media and society, pp. 226-230. 2018. DOI: 10.1145/3217804.3217917.

[31] K. Shu, D. Mahudeswaran, S. Wang, D. Lee and H. Liu. "Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media". Big Data. Vol. 8 Nº 3, pp. 171-188. 2020.

[32] K. Shu, D. Mahudeswaran and H. Liu. "FakeNewsTracker: a tool for fake news collection, detection, and visualization". Computational and Mathematical Organization Theory. Vol. 25 Nº 1, pp. 60-71. 2019.