



Title: Automatic identification of false opinions in social networks

Authors: GUZMAN-CABRERA, Rafael, HERNÁNDEZ-RAYAS, Angelica, PRASAD-MUKHOPADHYAY, Tirtha and RUIZ-PINALES, José

Editorial label ECORFAN: 607-8695

BECORFAN Control Number: 2023-03

BECORFAN Classification (2023): 111213-0301

Pages: 07

RNA: 03-2010-032610115700-14

MARVID - Mexico

Park Pedregal Business. 3580-
Adolfo Ruiz Cortines Boulevard –
CP.01900. San Jerónimo Aculco-
Álvaro Obregón, Mexico City
Skype: MARVID-México S.C.
Phone: +52 1 55 6159 2296
E-mail: contact@marvid.org
Facebook: MARVID-México S. C.
Twitter: @Marvid México

www.marvid.org

Holdings

Mexico	Colombia	Guatemala
Bolivia	Cameroon	Democratic
Spain	El Salvador	Republic
Ecuador	Taiwan	of Congo
Peru	Paraguay	Nicaragua

Resumen

En este trabajo se presenta el problema de la detección de opiniones falsas en redes sociales, también llamadas “opinión spam”, describiendo cómo las mentiras pueden detectarse automáticamente usando diferentes métodos. Estudios particulares sobre psicología social y comunicaciones muestran que las tasas de precisión de las habilidades de las personas para detectar el engaño están en el rango de 55 a 58%, es decir, ligeramente mejor que el azar. En este trabajo se aborda el problema específico de la detección del engaño en la comunicación.

Finalmente, nos centramos en la identificación de opinión spam, cuya detección es muy importante para una minería de opinión fiable. Se presentan resultados obtenidos utilizando distintos métodos de aprendizaje automático. Los resultados obtenidos permiten ver la viabilidad de la metodología propuesta para llevar a cabo la detección de opiniones falsas en redes sociales al obtener valores de precisión superiores al 80%.

Introducción

La opinión es un acto natural del ser humano y le permite a éste discernir la realidad que lo rodea para luego tomar acción sobre ella. El que las personas estén recibiendo información falsa, no es algo nuevo y exclusivo de nuestra era, sin embargo, se ha popularizado por el uso de foros, blogs y en general en las redes sociales.

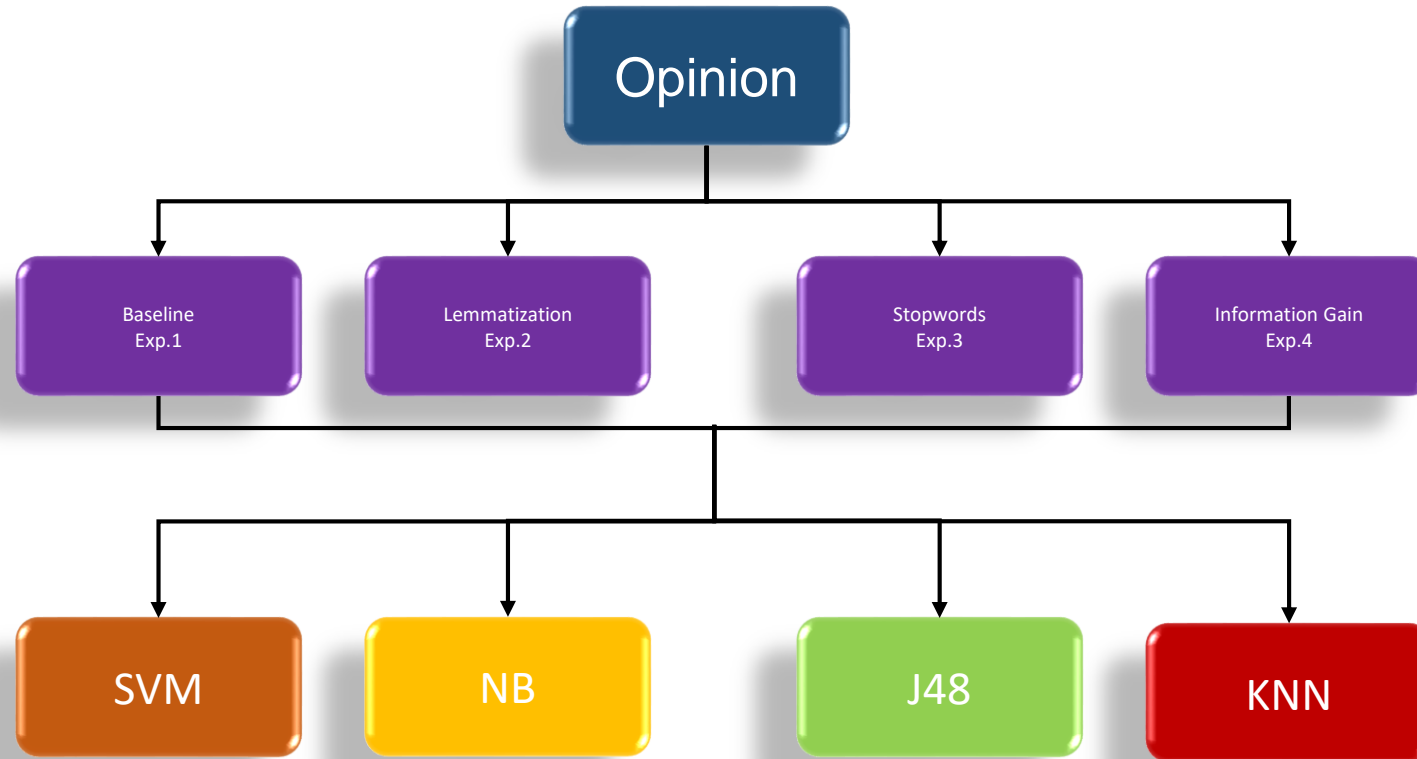
En la época actual por medio de las comunicaciones digitales es posible adquirir casi cualquier producto y contratar toda clase de servicios, sin haber tenido que cruzar una sola palabra con persona alguna. El problema de la detección de opiniones en textos no estructurados consiste en detectar las opiniones que no siguen una estructura como tal o un formato ya establecido. Esto se puede observar claramente de las opiniones que se dan en redes sociales como Facebook, Instagram, Twitter, etc.

Introducción

Todo esto conlleva a un gran problema de las opiniones falsas (Opinion spam), que son escritas deliberadamente para promover o desacreditar un producto o servicio. Existen diferentes técnicas que pueden ayudar a dar solución a este problema como los son el procesamiento del lenguaje natural y el aprendizaje automático.

En este trabajo se presentan resultados de cuatro experimentos realizados utilizando el corpus “Deceptive Opinion Spam” que consta de 1.600 opiniones en total, las opiniones son sobre el servicio en hoteles, divididas en dos categorías principales: opiniones "truthful" (genuinas) y opiniones "deceptive" (engañosas). Cada categoría tiene 800 documentos. En la parte de metodología se describen cada uno de los cuatro experimentos realizados.

Metodología

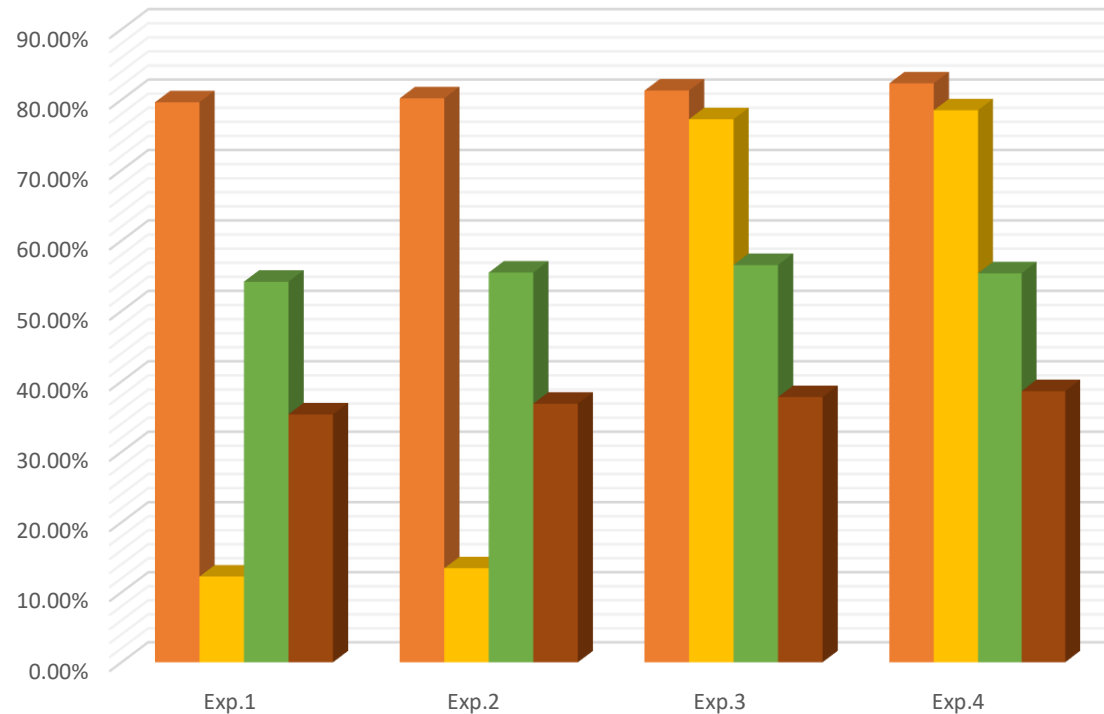


Resultados

Los resultados obtenidos para los cuatro experimentos realizados se muestran en la tabla y figura siguiente:

	SVM	NB	J48	KNN
<i>Exp.1</i>	79.63%	12.25%	54.16%	35.32%
<i>Exp.2</i>	80.20%	13.43%	55.46%	36.84%
<i>Exp.3</i>	81.30%	77.23%	56.53%	37.78%
<i>Exp.4</i>	82.32%	78.50%	55.35%	38.62%

Resultados Experimentales



Conclusiones

La metodología implementada en el presente trabajo aborda la detección de opiniones falsas en el ámbito hotelero utilizando el corpus Deceptive Opinion Spam. Se utilizaron cuatro métodos distintos de aprendizaje automático y se presentan resultados utilizando varios preprocesamientos, incluyendo una lematización de palabras. Los mejores resultados en la métrica de evaluación utilizada se obtienen con SVM en la detección de opiniones falsas y engañosas. Este nivel de precisión es competitivo y sugiere que el enfoque utilizado en este estudio tiene potencial para detectar opiniones engañosas de manera efectiva.

Es importante destacar que la lematización de palabras aplicada en el preprocesamiento de los datos ha demostrado ser una técnica efectiva para mejorar la precisión de la detección de opiniones falsas y engañosas. Al reducir las palabras a su forma base, se logra una representación más generalizada y se capturan mejor las características esenciales de las opiniones.

Referencias

1. Fitzpatrick, E., J.C. Bachenko, and T. Fornaciari, *Automatic detection of verbal deception*. 2015.
2. Ren, Y. and D. Ji, *Learning to detect deceptive opinion spam: A survey*. IEEE Access, 2019. **7**: p. 42934-42945.
3. Ott, M., C. Cardie, and J.T. Hancock. *Negative deceptive opinion spam*. in *Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: human language technologies*. 2013.
4. Salunkhe, A., *Attention-based Bidirectional LSTM for Deceptive Opinion Spam Classification*. arXiv preprint arXiv:2112.14789, 2021.
5. SujithraKanmani, R. and B. Surendiran, *Boosting credibility of a Recommender System using Deep Learning Techniques-An Empirical Study*.
6. Fusilier, D.H., et al. *Using PU-learning to detect deceptive opinion spam*. in *Proceedings of the 4th workshop on computational approaches to subjectivity, sentiment and social media analysis*. 2013.



© MARVID-Mexico

No part of this document covered by the Federal Copyright Law may be reproduced, transmitted or used in any form or medium, whether graphic, electronic or mechanical, including but not limited to the following: Citations in articles and comments Bibliographical, compilation of radio or electronic journalistic data. For the effects of articles 13, 162,163 fraction I, 164 fraction I, 168, 169,209 fraction III and other relative of the Federal Law of Copyright. Violations: Be forced to prosecute under Mexican copyright law. The use of general descriptive names, registered names, trademarks, in this publication do not imply, uniformly in the absence of a specific statement, that such names are exempt from the relevant protector in laws and regulations of Mexico and therefore free for General use of the international scientific community. BECORFAN is part of the media of MARVID-Mexico., E: 94-443.F: 008- (www.marvid.org/booklets)