

Método para detección de emociones desde foros utilizando Text Mining

Method for detecting emotions from forums using Text Mining

Laura Aballay¹, Silvana Aciar¹, Eliseo Reategui²

¹ Universidad Nacional de San Juan, Argentina

² Universidad Federal de Río Grande do Sul (UFRGS), Brasil

lnaballay@gmail.com , silvanav.aciar@gmail.com , eliseoreategui@gmail.com

RESUMEN. La deserción escolar es un fenómeno presente a nivel internacional y es notable el aumento en la tasa de fracaso y abandono de alumnos en cualquier modalidad de educación. Es por ello que es importante analizar y comprender el estado anímico de los estudiantes, tratando de averiguar que emociones presenta éste durante el cursado.

En este trabajo se propone un método para detectar automáticamente emociones en el texto escrito en los comentarios de un foro de alumnos. El objetivo es ofrecer esta información al profesor para que pueda hacer un seguimiento del estado anímico de sus estudiantes, y así decida si debe mejorar los contenidos o estructura de su curso con el fin de mantener a los estudiantes comprometidos, aumentar el interés, y, probablemente, mejorar su satisfacción para mejorar y fortalecer el aprendizaje. Este método fue aplicado con la herramienta RapidMiner y posteriormente los resultados se contrastaron con un método manual, utilizando encuestas.

ABSTRACT. The dropout is a phenomenon internationally and the rate of failure and abandonment of children have a marked increase in any level of education. That is why it is important to analyze and understand the mood of students, trying to figure out which emotions it presents during the studied.

This paper presents a method for detecting emotions in forum comments of students in a course that can be classified into positive and negative. The goal is that the teacher can track the mood of his students, and improve the content or structure of your course. The method has been implemented with the RapidMiner tool. The purpose of this work is to show corroboration for results of the method using a manual method using surveys.

PALABRAS CLAVE: Análisis de sentimiento, E-Learning, Minería de Texto, Detección de emociones, Léxico afectivo, Aprendizaje, Foro virtual, Educación, Tecnología.

KEYWORDS: Sentiment analysis, E-Learning, Text Mining, Emotions detection, Affective lexicon, Learning, Virtual forum, Learning, Technology.

1. Introducción

Un tema que preocupa a docentes, profesores y tutores es la deserción escolar. Este problema está presente a nivel internacional (Binali et al., 2010) y es notable el aumento en la tasa de fracaso y abandono de alumnos en cualquier modalidad de educación. Se estudian diversos modos para retenerlos, intentando motivarlos, mejorando las estrategias de enseñanza. Es por ello que es importante conocer y analizar el estado anímico de los estudiantes, tratando de averiguar que emociones experimenta durante el cursado (Binali et al., 2009; Shivhare et al., 2015).

Tradicionalmente, estas percepciones se evalúan a través de encuestas de satisfacción y una vez que el curso ha finalizado. Esta forma de evaluar la satisfacción de alumnos en un curso no detecta las emociones en el momento preciso que se presentan, y no se llega a saber por qué motivo el alumno se sintió frustrado o molesto.

Las emociones pueden expresarse y detectarse a través de diversos medios, tales como el habla, expresiones faciales, gestos y datos textuales (Cowie et al., 2008). Recopilar y estudiar gestos y habla implica hacer una gran inversión monetaria, sin embargo detectar por medio de texto escrito es más viable económicamente. Siendo el texto escrito el principal modo de comunicación (Cowie et al., 2008).

Los foros educativos virtuales son espacios de interacción que permiten el intercambio de conocimiento de una temática específica, de forma asincrónica, y que permiten buscar soluciones a problemáticas por medio de la opinión de los participantes de un curso. Todo esto con una finalidad académica (Área, 2004).

En este trabajo se propone un método para detectar automáticamente emociones en el texto escrito en los comentarios de un foro de alumnos, para que el profesor pueda hacer un seguimiento del estado anímico de sus estudiantes, y así decida si debe mejorar los contenidos o estructura de su curso con el fin de mantener a los estudiantes comprometidos, aumentar el interés, y, probablemente, mejorar su satisfacción para mejorar y fortalecer el aprendizaje. Este método fue aplicado con la herramienta RapidMiner y posteriormente los resultados se contrastaron por medio de método manual, utilizando encuestas.

La estructura del artículo es la siguiente: en la sección 2 se presentan los trabajos relacionados a la detección de emociones en texto, mencionando el uso de bases léxicas, los modelos de emociones, análisis de sentimientos. En la sección 3, se detalla el método propuesto para analizar las emociones en alumnos. Luego en la sección 4 se presenta un ejemplo en el que se aplicó el método propuesto, siguiendo en la sección 5 se muestran los resultados de la aplicación del método propuesto, donde se incluye una validación del mismo por medio de encuestas a los alumnos y comparación de los resultados. Para finalizar, en la sección 6, se plantean las conclusiones y el trabajo a futuro.

2. Trabajos relacionados

Para llevar a cabo la detección de emociones en texto, existen diversas técnicas (Amato et al., 2015; Chopade, 2015; Hirat y Mittal, 2015; Vinola y Vimaladevi, 2015), entre las que se destaca el uso de bases de datos léxicas. Actualmente para el análisis de sentimientos en texto existen corpus como Word Net, Word Net Affect (Valitutti, 2004), basada en el modelo de emociones de OCC (Clore y Ortony, 2013), SentiwordNet (Esuli y Sebastiani, 2006). La estrategia es recoger primero un pequeño conjunto de palabras, comparar este conjunto mediante la búsqueda en bases léxicas, como por ejemplo Word Net, por sus sinónimos y antónimos. Este enfoque es utilizado por Gaspar y Macedo, Loia y Senatore (Gaspar y Macedo, 2013) y Colace y otros (Colace et al., 2014). Estos léxicos de emociones se han desarrollado para diferentes idiomas, entre ellos inglés y portugués, pero existen pocos trabajos en los que se utilice un léxico en castellano.

Las emociones a interpretar suelen basarse en algún modelo teórico de emociones. Por ejemplo, Potena y Diamantini (Potena y Diamantini, 2010); Consoli (Consoli, 2010); Gaspar y Macedo (Gaspar y Macedo, 2013) y Binali (Binali et al., 2009) se basan en el modelo de emociones en forma categórica de Paul Ekman (Ekman

et al., 1972).

Muchos investigadores han indagado en el campo del análisis de sentimiento en e-learning, Rodríguez (Rodríguez et al., 2012); Binali (Binali et al., 2010); Colace (Colace et al., 2014) y Jackson (Jackson et al., 2008). En estos trabajos se ha podido ver como las emociones pueden afectar el proceso de e-learning.

3. Método propuesto

Existen diversos enfoques metodológicos para abordar esta problemática (Chopade, 2015; Vinola y Vimaladevi, 2015), el método propuesto sigue el enfoque basado en léxico (Amato et al., 2015). Este enfoque depende de una lista predefinida o léxico de palabras emotivas con cierta polaridad, que luego por medio de herramientas de análisis de texto, se buscarán esas palabras, contándolas y midiendo la polaridad total del texto.

Para construir el léxico se llevó a cabo una revisión de artículos, y se llegó a la conclusión de utilizar el diccionario de (Baca-Gomez et al., 2016) para emociones en debido a que está en español. Este diccionario es una traducción al español de Word Net Affect, y tiene en cuenta la polaridad establecida en esa base.

El léxico de palabras afectivas elegido se separó en dos clases:

- “negativa”, en la cual se agrupan las palabras emotivas que indican emociones desfavorables para el aprendizaje, son aquellas cuya polaridad es negativa.
- “positiva”, a la que pertenecen las palabras emotivas que denotan emociones positivas y favorables para el aprendizaje. Este grupo de palabras tiene polaridad positiva.

Las dos clases que componen el léxico se almacenan en dos archivos de texto diferentes.

Este léxico puede ser modificado simplemente modificando estos archivos. Se supone que el léxico irá creciendo a medida que se realicen pruebas con el mismo. Mientras más grande sea el léxico, mayor será la precisión obtenida por el método.

Por lo tanto se utilizan dos diccionarios que componen el léxico afectivo, uno para emociones positivas y otro para emociones negativas, tal como se puede observar en la figura 1.



Figura 1. Diccionarios de palabras emotivas. Fuente: Elaboración propia.

Tal como se observa en la figura 2, el método propuesto puede utilizarse para el análisis de los mensajes de un foro de alumnos. Estos mensajes son agrupados por alumno y se extraen en forma de texto plano.

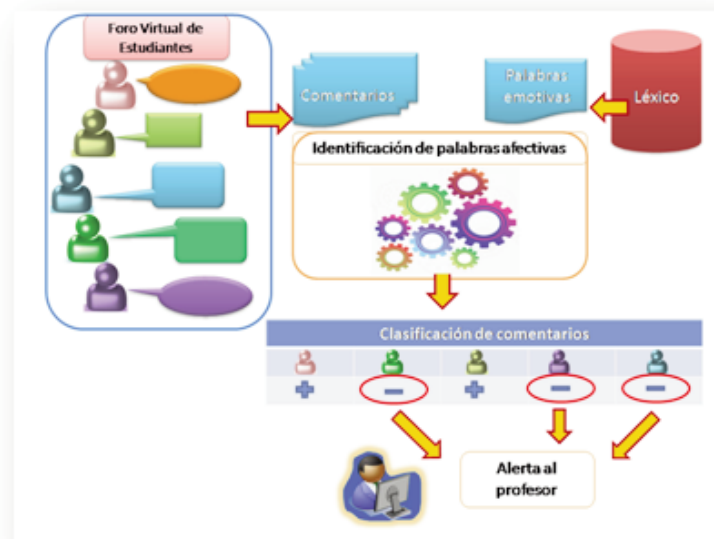


Figura 2. Método para extraer emociones. Fuente: Elaboración propia.

Para detectar las emociones se recorre cada comentario en busca de alguna palabra que coincida con las listas de palabras emotivas y determinar si el comentario es negativo o positivo, etiquetando al comentario con la polaridad que acumule mayor cantidad. Luego se reúnen todos los comentarios por cada alumno obteniendo el valor emocional total del estudiante. Si un comentario contiene igual cantidad de palabras emotivas negativas que positivas, éste se considera negativo ya que el foco de este trabajo está en percibir la insatisfacción de los estudiantes.

4. Ejemplo de Aplicación

Como ejemplo de aplicación se tomó un foro en el que participaron alumnos de un curso universitario. Se trata de una materia con modalidad presencial donde se dictan clases de teoría y se entregan apuntes de cátedra. Como complemento se dictan clases prácticas, donde se les da ejercitación por medio de actividades prácticas. Para regularizar la materia se toman dos parciales y por último se les pide que desarrollen un trabajo final para integración de conceptos.

Se diseñó un foro virtual, en el que los estudiantes se registraron y fueron dejando sus comentarios sobre las diferentes perspectivas del curso, se les pidió que participaron opinando sobre el curso, qué problemáticas tuvieron durante el cursado, qué dificultades encontraron o qué aspectos positivos deseaban destacar.

Se tomaron seis perspectivas de la materia, de acuerdo a las actividades planteadas en la cátedra: Clases Teóricas, Clases Prácticas, Apuntes, Prácticas, Parciales y Trabajo Final. A cada perspectiva le corresponde un tema dentro del foro, por lo que se generaron seis temas, tal como se puede observar en la figura 3. De ésta manera se obtuvo suficiente texto escrito por cada alumno y separado por temas, para analizarlo intentando estimar sus emociones.

Como los mensajes son agrupados por diferentes aspectos de la materia y por alumno, es posible identificar

a qué alumno y respecto a qué aspecto de la materia pertenece el comentario analizado.

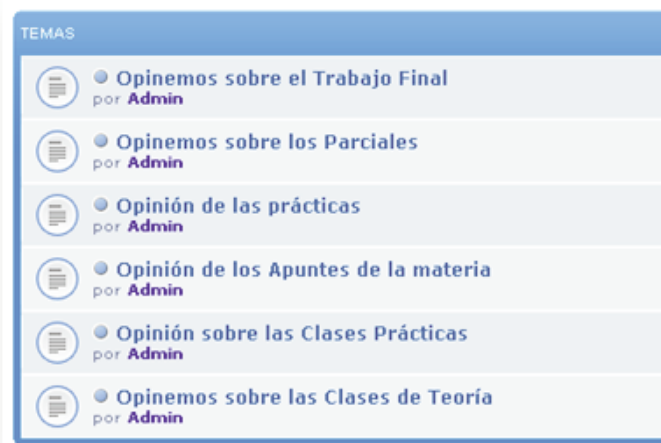


Figura 3. Temas del Foro. Fuente: Elaboración propia.

Posteriormente, se almacenó cada comentario en archivos de texto separados para poder analizarlos y aplicar el método propuesto.

Para conseguir el propósito, en este caso, se utilizó la herramienta RapidMiner ©, para analizar los comentarios de los alumnos.

El proceso tuvo como entradas los archivos de texto de comentarios por un lado y los listados de palabras afectivas (figura 4).

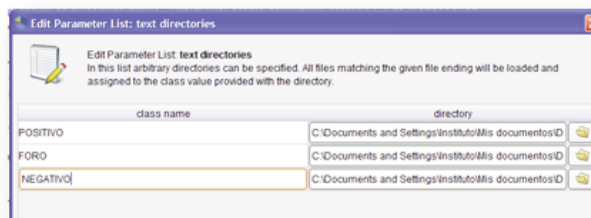


Figura 4. Archivos de Entrada para RapidMiner. Fuente: Elaboración propia.

A estos documentos de texto se los transformó en vectores, es decir, se convirtió cada documento en un vector de palabras donde cada comentario de un alumno será una fila y tendrá tantas columnas como el conjunto de palabras de todo el texto. Esto se lleva a cabo de manera automática con el operador “Process Document” de RapidMiner, el cual a su vez hizo un pre procesamiento: transformó todo el texto a minúscula, separó el texto en palabras (esto se llama tokenización en minería de texto), se descartaron palabras irrelevantes tales como artículos, preposiciones, etc. (extracción de stopwords).

El mismo procedimiento se hizo con los listados de palabras que denotan emociones, se lo transformó en documento de texto, se pre procesó y se vectorizó.

A continuación se confrontaron los documentos de comentarios con los listados de palabras emotivas y se analizó la distancia entre ellos con el operador “Cross Distance” de RapidMiner, el cual aplica la técnica de semejanza o similitud de documentos, por medio del coseno del ángulo que forman los vectores de los documentos. Mientras ese valor se acerca a 1, mayor será la similitud.

5. Resultados

Como resultado se consiguió un listado indicando si el comentario es más cercano a ser positivo o a negativo, esto se puede visualizar en la figura 5.

Row No.	request	distance	valencia
1	A mi me gustó la clase de teoría del diagrama de estado, fué divertido el ejemplo que	0.096	POSITIVO
2	Yo estuve satisfecho con todo lo que nos dio la profesora, muy buenos apuntes y bie	0.115	POSITIVO
3	Lo que la materia era los días lunes en la mañana temprano, me sentía desganado	0.060	NEGATIVO
4	Huy si, yo siempre estaba desanimado en la teoría porque era muy larga la clase.	0.062	NEGATIVO
5	A mi el trabajo final me resultó demasiado extenso, estaba harto...pero por suerte lo	0.087	NEGATIVO
6	Umm en la teoría me sentía aburrido.	0.126	NEGATIVO
7	Como aún no las terminé por problemas de salud, me siento frustrado y enojado, a	0.166	NEGATIVO
8	Me sentía nervioso, temeroso al hacerlos, aunque estuvieron bien, me desilusioné n	0.167	NEGATIVO

Figura 5. Resultados de RapidMiner. Fuente: Elaboración propia.

Con esto es posible identificar al alumno que ha expresado tal comentario y determinar su estado emotivo respecto a un aspecto de la materia, para que el profesor arbitre los medios necesarios para tratar de revertir esto.

Resumiendo los resultados obtenidos con el proceso de RapidMiner, se confeccionaron los siguientes gráficos (figura 6 y figura 7).

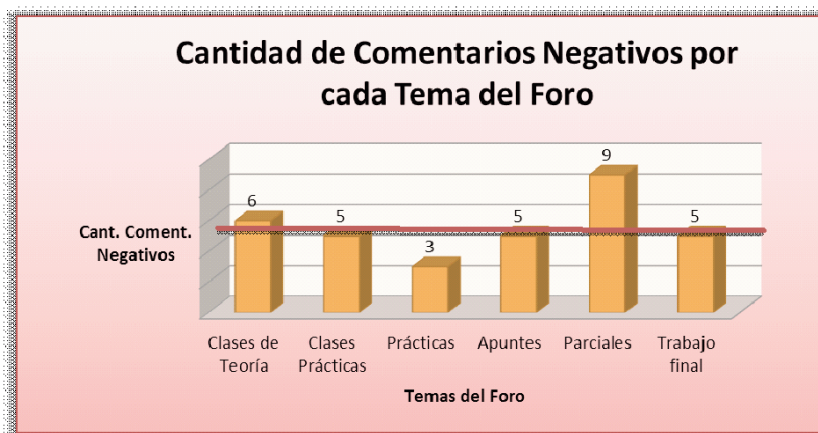


Figura 6. Comentarios negativos por cada Tema del Foro. Fuente: Elaboración propia.

Tal como se observa en la figura 6, los temas del foro donde los alumnos han indicado estar más desconformes, es decir que la mayoría se ha expresado con palabras emotivas negativas, son Clases de Teoría y Parciales. Desde otro punto de vista, en la figura 7, el alumno con mayor cantidad de comentarios negativos es el número 10, seguido de los alumnos 3, 4 y 9. Esto indica que hay que analizar los problemas que pueden existir en estos aspectos de la materia, y por otro lado hacer un seguimiento a los estudiantes más insatisfechos.

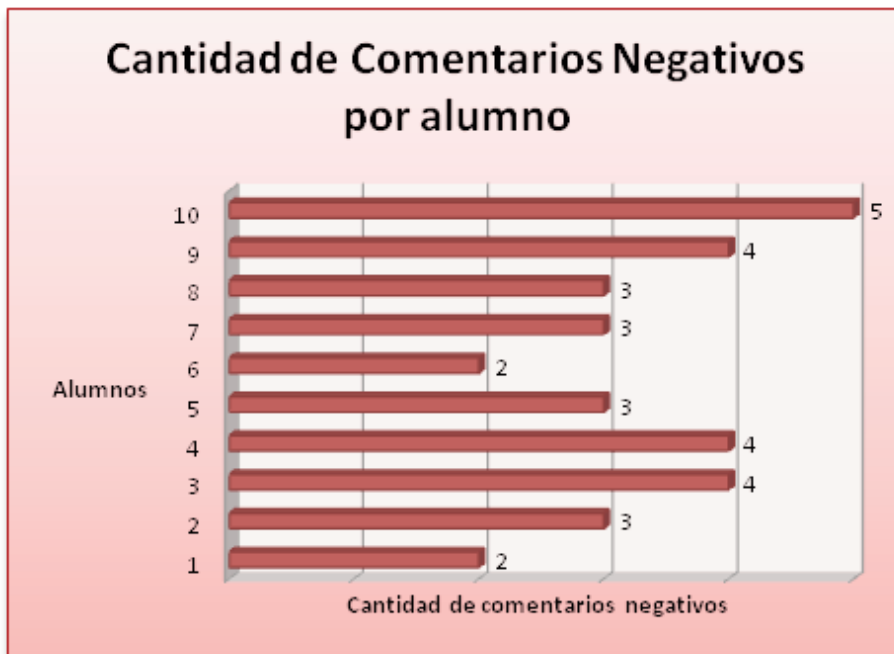


Figura 7. Comentarios negativos por Alumno. Fuente: Elaboración propia.

5.1. Validación de los Resultados

La validación de los resultados del método se realizó encuestando a los mismos alumnos que participaron del foro de la materia. Posterior a la participación del foro se les solicitó a los alumnos que completen una encuesta, resumiendo emociones percibidas respecto a cada aspecto de la materia. Se utilizó como base la encuesta PrEmo (Desmet, 2005) y se la adaptó en este caso para interpretar emociones respecto a un curso. Luego de llevar a cabo las encuestas, se recopilaron y analizaron por cada alumno y para los diferentes aspectos definidos para la materia.

Los resultados del método aplicando la herramienta RapidMiner, se corroboraron con un método manual, utilizando las encuestas PrEmo. Esta corroboración se realizó encuestando a los mismos alumnos que participaron del foro de la materia. Posterior a la participación del foro se les solicitó a los alumnos que completen una encuesta, resumiendo emociones percibidas respecto a cada aspecto de la materia. Se utilizó como base la encuesta PrEmo y se la adaptó en este caso para interpretar emociones relativas al aprendizaje, que coinciden con las palabras de los léxicos emotivos utilizados en la aplicación del método. La encuesta utiliza 14 emociones, 7 placenteras y 7 no placenteras, para que el encuestado indique cuál de ellas ha percibido al momento de realizar las diferentes actividades planteadas en la cátedra.

De la misma manera que en el foro, se plantearon seis temas correspondientes con las mismas seis perspectivas de la materia: Clases Teóricas, Clases Prácticas, Prácticas, Apuntes, Parciales y Trabajo Final.

Luego de llevar a cabo las encuestas, se recopilaron y analizaron por cada alumno y para los diferentes aspectos definidos para la materia. Los detalles de los análisis se pueden observar en las figuras 8 y 9.

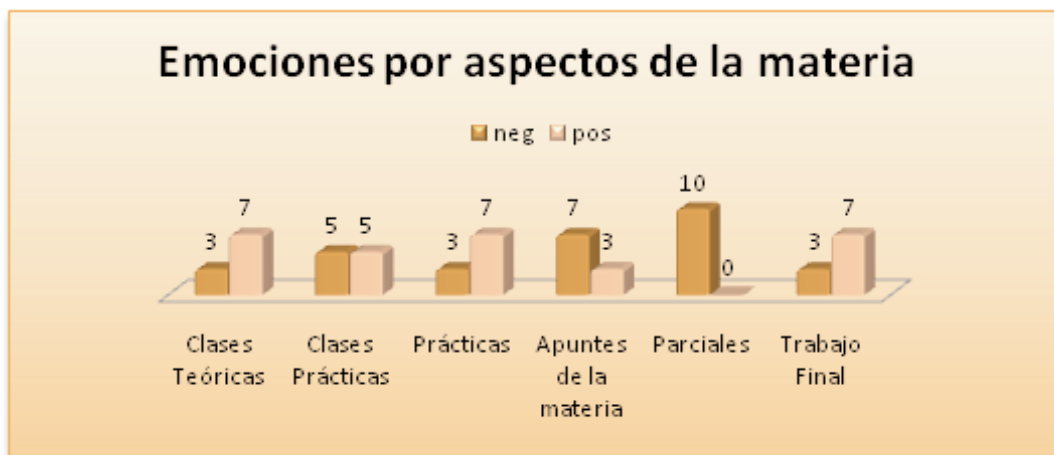


Figura 8. Emociones Negativas y Positivas por aspectos de la materia. Fuente: Elaboración propia.

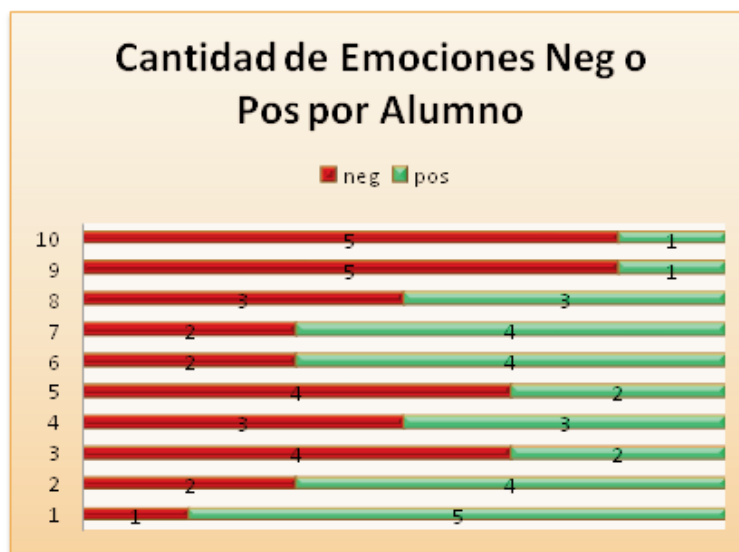


Figura 9. Emociones Negativas y Positivas por alumno. Fuente: Elaboración propia.

5.2. Comparación de resultados obtenidos con RapidMiner y encuestas

Como una alternativa de corroborar el procedimiento realizado con RapidMiner, se utilizaron encuestas. Haciendo un recuento de las coincidencias (y transformándolas en porcentaje) entre los resultados de aplicar el método con la herramienta RapidMiner y los resultados de la encuesta se obtuvo que (ver figura 10): en general hubo gran porcentaje de coincidencia respecto a cada tema del foro (perspectivas en que se dividió la materia). Pues todos los valores son iguales o mayores a 70%, siendo las Clases Teóricas el más bajo con 70% de coincidencia, siguiendo con Prácticas, Apuntes y Trabajo Final, con 80%; luego los Parciales con 90% y por último las Clases Prácticas hubo coincidencia total entre lo que se obtuvo con RapidMiner y el resultado de la encuesta.

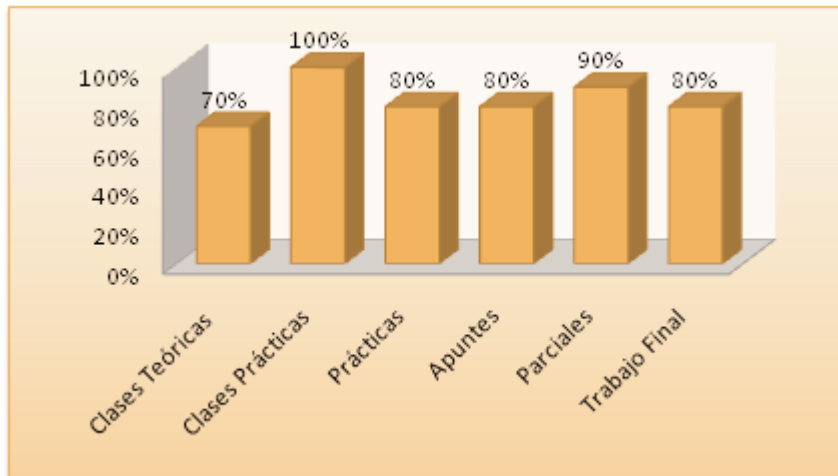


Figura 10. Porcentajes de coincidencia RapidMiner y encuesta. Fuente: Elaboración propia.

Haciendo un análisis general, uniendo los resultados de todos los temas del foro, se puede apreciar que un 83% de los resultados de RapidMiner coinciden con los resultados de la encuesta.

Las causas por la que los resultados fuesen diferentes se puede deber a varios motivos, como por ejemplo, que el comentario haya sido ambiguo y el método en RapidMiner no haya conseguido detectar la emoción, que el comentario haya utilizado palabras emotivas que no están en el diccionario, o no haber usado palabras emotivas en el foro y en la encuesta si, por citar algunos.

6. Conclusiones y trabajo a futuro

El objetivo de este trabajo es interpretar las emociones expresadas en forma escrita por estudiantes que participan interactuando en un foro educativo virtual, con el fin de que el profesor tenga conocimiento de esto y realice acciones en su materia tendientes a mejorar el estado anímico del alumno, y así mejorar su predisposición para el aprendizaje. Pensando en la satisfacción de los estudiantes y contribuyendo a disminuir el abandono por parte de los mismos, situación que preocupa en todos los niveles educativos.

Se diseñó un método para detectar emociones y se lo aplicó a un foro educativo virtual, obteniendo de manera automática si el alumno se sintió negativa o positivamente respecto a algún aspecto de la materia.

Este método utiliza listados de emociones en español, para poder compararlos con los comentarios de los alumnos. Pero se puede adaptar fácilmente a otro idioma cambiando estos diccionarios.

Este método fue aplicado con la herramienta RapidMiner y posteriormente los resultados se contrastaron por medio de método manual, utilizando encuestas.

Para corroborar estos resultados, se usó otro tipo de detección de emociones, encuestando a los mismos alumnos que participan del foro.

Luego de analizar las encuestas se concluye que un 83% de los resultados de RapidMiner coinciden con los resultados de la encuesta. Esto indica que el estado anímico del alumno detectado con el método aplicado con RapidMiner es bastante cercano a lo que el mismo alumno expresa en las encuestas. Si bien son resultados provisorios, hay que continuar afinando el método y buscar otras formas de validación.

Los diccionarios de palabras afectivas utilizados, son listados reducidos de palabras emotivas que pueden

ser extendidos o modificados con futuras iteraciones para obtener mejores resultados en la precisión del método. Además, es conveniente aclarar que no se tienen en cuenta las palabras abreviadas, ni con errores de ortografía, así como tampoco se analizan en este trabajo los emoticones introducidos por los alumnos.

Actualmente se está trabajando en la incorporación de detección de negaciones, palabras que invertirían la polaridad de una emoción.

Cómo citar este artículo / How to cite this paper

Aballay, L.; Aciar, S.; Reategui, E. (2017). Método para detección de emociones desde foros utilizando Text Mining. *Campus Virtuales*, 6(1), 89-98. (www.revistacampusvirtuales.es)

Referencias

- Amato, F.; Colace, F.; Greco, L.; Moscato, V.; Picariello, A. (2015). A Quick Survey on Sentiment Analysis Techniques: a lexical based perspective. (<http://doi.org/10.18293/VLSS2015-020>)
- Área, M. (2004). Las tecnologías de la información y comunicación en el sistema escolar: Una revisión de las líneas de investigación.
- Baca-Gomez, Y. R.; Martinez, A.; Rosso, P.; Estrada, H.; Farias, D. I. H. (2016). Web service SWePT: A hybrid opinion mining approach. *Journal of Universal Computer Science*, 22(5), 671-690.
- Binali, H. H.; Wu, C.; Potdar, V. (2009). A new significant area: Emotion detection in E-learning using opinion mining techniques. In 3rd IEEE International Conference on Digital Ecosystems and Technologies, 2009. DEST '09 (pp. 259-264). DOI: <http://doi.org/10.1109/DEST.2009.5276726>
- Binali, H.; Wu, C.; Potdar, V. (2010). Computational approaches for emotion detection in text. In 2010 4th IEEE International Conference on Digital Ecosystems and Technologies (DEST) (pp. 172-177). DOI: <http://doi.org/10.1109/DEST.2010.5610650>
- Chopade, C. R. (2015). Text Based Emotion Recognition : A Survey.
- Clare, G. L.; Ortony, A. (2013). Psychological Construction in the OCC Model of Emotion. *Emotion Review*, 5(4), 335-343. DOI: <http://doi.org/10.1177/1754073913489751>
- Colace, F.; De Santo, M.; Greco, L.; Guerriero, G. (2014). Sentiment Analysis and E-Learning: a Proposal.
- Consoli, D. (2010). Textual Emotions Recognition With An Intelligent Software Of Sentiment Analysis. Università Politecnica delle Marche, Ancona, Italy.
- Cowie, R.; Douglas-Cowie, E.; Karpouzis, K.; Caridakis, G.; Wallace, M.; Kollias, S. (2008). Recognition of Emotional States in Natural Human-Computer Interaction. In *Multimodal User Interfaces* (pp. 119-153). Springer Berlin Heidelberg, Berlin, Heidelberg. DOI: http://doi.org/10.1007/978-3-540-78345-9_6
- Desmet, P. (2005). Pieter Desmet Chapter 9. Funology: From Usability to Enjoyment, 111-123.
- Ekman, P.; Friesen, W. V.; Ellsworth, P. (1972). Emotion in the human face: Guidelines for research and an integration of findings (Vol. xii). Pergamon Press, Oxford.
- Esuli, A.; Sebastiani, F. (2006). SentiWordNet: A publicly available lexical resource for opinion mining. *Proceedings of Language Resources and Evaluation (LREC)*.
- Gaspar, Y.; Macedo, L. (2013). Towards Building User Affect Models from Tweets: a comparative study with various datasets, features and algorithms.
- Hirat, R.; Mittal, N. (2015). A Survey On Emotion Detection Techniques using Text in Blogposts. *International Bulletin of Mathematical Research*.
- Jackson, T.; Craig, S.; Morgan, B.; Chipman, P.; White, H.; Person, N.; ... ; Graesser, A. (2008). AutoTutor Detects and Responds to Learners Affective and Cognitive States.
- Potena, D.; Diamantini, C. (2010). Mining opinions on the basis of their affectivity. In 2010 International Symposium on Collaborative Technologies and Systems (CTS) (pp. 245-254). DOI: <http://doi.org/10.1109/CTS.2010.5478503>
- Rodríguez, P.; Ortigosa, A.; Carro, R. M. (2012). Extracting Emotions from Texts in E-Learning Environments. In 2012 Sixth International Conference on Complex, Intelligent and Software Intensive Systems (CISIS) (pp. 887-892). DOI: <http://doi.org/10.1109/CISIS.2012.192>
- Shivhare, S. N.; Garg, S.; Mishra, A. (2015). EmotionFinder: Detecting emotion from blogs and textual documents. In 2015 International Conference on Computing, Communication Automation (ICCCA) (pp. 52-57). (<http://doi.org/10.1109/CCAA.2015.7148343>)
- Valitutti, R. (2004). WordNet-Affect: an Affective Extension of WordNet. In *Proceedings of the 4th International Conference on Language Resources and Evaluation*. Citeseer.
- Vinola, C.; Vimaladevi, K. (2015). A survey on human emotion recognition approaches, databases and applications. *ELCVIA Electronic Letters on Computer Vision and Image Analysis*, 14(2), 24-44. DOI: <http://doi.org/10.5565/rev/elcvia.795>