



RIED. Revista Iberoamericana de
Educación a Distancia
ISSN: 1138-2783
ried@edu.uned.es
Asociación Iberoamericana de Educación
Superior a Distancia
Organismo Internacional

Gutiérrez Esparza, Guadalupe; Margain Fuentes, María de Lourdes; Ramírez del Real,
Tania Aglaé; Canul Reich, Juana

Un modelo basado en el Clasificador Naïve Bayes para la evaluación del desempeño
docente

RIED. Revista Iberoamericana de Educación a Distancia, vol. 20, núm. 2, 2017, pp. 293-
313

Asociación Iberoamericana de Educación Superior a Distancia
Madrid, Organismo Internacional

Disponible en: <http://www.redalyc.org/articulo.oa?id=331453132014>

- ▶ Cómo citar el artículo
- ▶ Número completo
- ▶ Más información del artículo
- ▶ Página de la revista en redalyc.org

Un modelo basado en el Clasificador Naïve Bayes para la evaluación del desempeño docente

A model based on the Naïve Bayes Classifier for teacher performance assessment

Guadalupe Gutiérrez Esparza
María de Lourdes Margain Fuentes
Tania Aglaé Ramírez del Real
Universidad Politécnica de Aguascalientes (México)

Juana Canul Reich
Universidad Juárez Autónoma de Tabasco (México)

Resumen

La evaluación del desempeño docente es un proceso de medición importante en las instituciones de educación superior en México y en el mundo, ya que retroalimenta el desempeño de los docentes con el fin de mejorar las clases y estrategias para beneficio de la educación de los estudiantes. En este trabajo se describe el desarrollo y evaluación de un Modelo Computacional denominado *SocialMining*, basado en el algoritmo Naïve Bayes, para apoyar el análisis de las opiniones de los estudiantes en el proceso de la evaluación del desempeño docente, llevada a cabo mediante dispositivos móviles. Esta propuesta considera el uso de dispositivos móviles para la recopilación de datos aprovechando su aceptación por parte de los estudiantes en el proceso de educación y aprendizaje. Asimismo, se describe el desarrollo de corpus de subjetividad, el cual consta de un conjunto de términos afectivos relevantes de la evaluación docente para apoyar al algoritmo Naïve Bayes en la clasificación de las opiniones de los estudiantes dentro de las clases: positivo, negativo y neutral. Para medir el desempeño del proceso de la clasificación del Modelo Computacional *SocialMining*, se utilizan métricas como la matriz de confusión, precisión y la curva de ROC. Se presenta además un caso de estudio, en el cual se recolectan nuevas opiniones de estudiantes de la Universidad Politécnica de Aguascalientes (México) con el fin de probar el desempeño del modelo propuesto en la clasificación. Los resultados obtenidos consideran factible el Modelo Computacional *SocialMining* para implementarse en instituciones de educación superior.

Palabras clave: minería de opiniones; clasificador bayesiano; conjunto de datos; subjetividad; análisis ROC; dispositivo móvil; evaluación docente; planeación de la educación; caso de estudio.

Abstract

The evaluation of teacher performance is an important measurement process in Mexico's higher education institutions and around the world, because it allows feedback on the teacher's performance to detect improvements in classes and propose strategies for the benefit of students' education. This paper describes the development and evaluation of a proposed computational model called *SocialMining*, which is based on the classifier algorithm Naïve Bayes to support the analysis of students' opinions from the process of teachers' performance evaluation, which is carried out through mobile devices. The mobile device revolutionizes processes in education; the proposal considers the use of this technology for the collection of data, taking advantage of processing capacity and acceptance by students in the process of education and learning. It also describes the development of a set of relevant affective terms of the teacher evaluation called corpus of subjectivity, which supports the Naïve Bayes algorithm to classify students' comments within the classes: positive, negative and neutral. To measure the comments classification performance of the *SocialMining* Computational Model, metrics such as the confusion matrix, precision, sensitivity, specificity and the ROC curve are used. Likewise, a study case is presented, which gathers new comments from students of the Polytechnic University of Aguascalientes (Mexico), in order to test the classification process performance of the proposed model. The results show that *SocialMining* Computational Model is feasible to be implemented in institutions to support Teacher Performance Assessment. Besides, our results show that Naïve Bayes can obtain a classification percentage very similar to those reported in recent works with related algorithms.

Keywords: Sentiment analysis; Naïve Bayes; corpus; subjectivity; ROC curve; mobile device; teacher performance assessment; educational planning; case study.

La evaluación docente contribuye al fortalecimiento de las prácticas de enseñanza, además de que permite orientar la formación continua y sirve como un marco de referencia del desempeño del docente. Las modalidades actuales para llevar a cabo la evaluación del desempeño docente en el aula consisten en aplicar cuestionarios a los estudiantes, realizar evaluaciones de pares académicos y observar la clase en el aula (Moreno 2004; Córdova 2015). Una de las estrategias más utilizadas consiste en captar las opiniones de los estudiantes respecto al desempeño del docente (Elizalde, 2008) mediante encuestas que permiten la inclusión de comentarios. Esta estrategia (evaluación del profesor por el alumno) en la actualidad puede ser realizada por medio de herramientas informáticas o redes sociales como Twitter. Esto ha propiciado la necesidad de contar con herramientas computacionales para efectuar un procesamiento más rápido y efectivo en el análisis de las opiniones de los estudiantes (Gewerc, 2014). Cabe resaltar que la evaluación docente resultante de las encuestas a los estudiantes se considera como una parte significativa del desempeño del profesorado, tomando en cuenta que esto solo ofrece un segmento de la calificación total (Loureiro, 2016).

El uso del Internet para aplicar la evaluación docente por parte de alumnos cada vez es más común, este método ha sido ampliamente divulgado y admitido por las escuelas de educación superior a nivel internacional (Ruiz, 2005).

Este trabajo presenta la descripción del desarrollo de un Modelo Computacional denominado *SocialMining* para apoyar en el proceso de la evaluación docente mediante el análisis de las opiniones de los estudiantes, empleando técnicas de minería de textos y el algoritmo Náïve Bayes.

El desarrollo de la etapa de experimentación fue llevada a cabo en la Universidad Politécnica de Aguascalientes, México (UPA). Los autores muestran el desarrollo del Modelo Computacional SocialMining considerando la importancia de la validación de un conjunto de datos, en el cual se recolectan comentarios de alumnos de la UPA durante dos períodos escolares con el fin de conformar un corpus de *términos* relevantes denominado corpus de subjetividad, que se apoya en el proceso de clasificación de comentarios en positivos o negativos. De igual manera, se presenta un caso de estudio con un nuevo grupo de estudiantes para comparar la precisión del Modelo Computacional *SocialMining* en clasificación de comentarios.

El Modelo Computacional *SocialMining* presenta ventajas frente al análisis de opiniones de estudiantes, debido a que comúnmente es realizado por personas, el apoyo con sistemas de información programados implica menor tiempo invertido y se tiene como resultado una evaluación cualitativa; esto permite implementar acciones correctivas y/o preventivas en la actividad docente en un tiempo más efectivo.

Las secciones del presente trabajo de aquí en adelante se organizan de la manera siguiente: se expone una revisión de trabajos relacionados, se presentan los materiales y métodos usados en esta investigación así como la arquitectura general del Modelo Computacional *SocialMining*, posteriormente se describen los experimentos y los resultados obtenidos así como el caso de estudio. Por último, se presentan las conclusiones y las referencias de apoyo para esta investigación.

TRABAJOS RELACIONADOS

En la década de los noventa algunos estudios de la literatura (Brusilovsky, 2001) hacen una revisión de las técnicas y métodos de un área denominada hipermedia adaptativa, la cual desde entonces buscaba ofrecer contenidos de páginas web hipermedias de acuerdo a las necesidades del usuario. Esta área es una de las impulsoras para apoyar la mejora de cursos de educación a distancia y ofrecer al usuario contenidos adaptables de acuerdo al estilo de aprendizaje y nivel de conocimiento, con base en los datos obtenidos mediante cuestionarios y actividades de interacción con el estudiante. Se podría decir que este tipo de aplicaciones son las que provocan la inclusión del área de análisis de emociones en la educación.

El análisis de emociones, definido como un área de estudio computacional de opiniones, sentimientos y emociones expresadas en textos (Liu, 2010), se ha combinado con técnicas de aprendizaje automático, minería de datos y procesamiento

de lenguaje natural. En el área de la educación se ha buscado aplicar el análisis de emociones con el fin de mejorar el proceso de enseñanza – aprendizaje. Existen investigaciones que demuestran las ventajas de usar redes sociales para fomentar la participación de estudiantes universitarios para expresarse libremente (Crovi 2014; González-Lizárraga 2016). Algunos otros estudios (Valencia, 2016) buscan medir por medio de ecuaciones estructurales la aceptación del aprendizaje móvil (m-learning o mobile-learning) por los estudiantes, otros (Zaldivar 2015; Guerrero 2016) se enfocan en determinar el impacto de las tecnologías móviles y redes sociales, como apoyo en el aprendizaje de los alumnos.

La evaluación docente se considera un recurso para conducir la labor del profesorado de acuerdo al desempeño obtenido y una fuente para medir su actuación es la valoración por parte de los estudiantes, el cual se denomina modelo basado en la opinión de los alumnos (Elizalde, 2008).

En el trabajo reportado por Loureiro (2016) realizan un análisis de los comentarios globales a los docentes, con lo cual concluyen que es un buen indicador, ya que revela cualidades del docente en su labor. Es importante destacar que este estudio lo ejecutan leyendo detalladamente cada una de las opiniones.

La mejora continua en el proceso de enseñanza-aprendizaje ha enfocado sus esfuerzos para concluir los componentes que determinan la labor docente ejecutada adecuadamente, instrumentos como lo son cuestionarios de evaluación de la docencia por los alumnos persiguen contar con la opinión por parte de los estudiantes, los cuales han figurado como una de las herramientas más examinadas y aplicadas para este propósito (Martínez, 2010).

En este trabajo el uso de las tecnologías móviles facilita el acceso a la información permite que los estudiantes participen en procesos de evaluación –a docentes– comentando diferentes aspectos sobre el aprendizaje. Como beneficio del uso de esta tecnología, el alumno interviene en el proceso de manera favorable. Expresan una opinión y se fomenta el trabajo colaborativo en la construcción del nuevo conocimiento mediante la captura de información que posteriormente requiere ser procesada. Las evaluaciones en línea permiten la retroalimentación continua, el estudiante puede opinar cuando lo deseé, así como el profesor consultar su desempeño (Tirado, 2007).

La recolección de opiniones mediante aplicaciones en dispositivos móviles implica ventajas, como lo pueden ser la disposición en línea. Esto último permite acceder desde cualquier lugar mientras que lo primero implica más opiniones recabadas así como la disponibilidad de datos para su procesamiento (Tirado, 2007).

Asimismo, hay investigaciones (Altrabsheh, 2014; Ortigosa, 2014) para monitorear y analizar en tiempo real los comentarios que los estudiantes hacen en redes sociales con el fin de mejorar las clases de los docentes. Además, algunos autores (Bravo, 2013) documentan el uso de la red social Twitter como una herramienta complementaria en la educación colaborativa y el aprendizaje informal en el área de ingeniería. De igual manera, otros autores (Novak, 2011) exponen el crecimiento

exponencial en el uso de redes sociales como Twitter por universidades a nivel mundial para mantener comunicación con los estudiantes y mejorar la enseñanza.

MATERIALES Y MÉTODOS

Materiales

En este trabajo se aplica un tipo de investigación experimental, siendo el diseño seleccionado el cuasi experimental, por lo que cabe resaltar que los grupos utilizados para este estudio son intactos.

Los datos analizados en este estudio se recolectaron utilizando dos instrumentos: Twitter y la Encuesta de Evaluación Docente (EED). En Twitter se analizaron alrededor de 800 comentarios de un grupo que evaluó a sus docentes durante un ciclo escolar. Los estudiantes evaluaron a sus profesores desde cuentas anónimas en Twitter. Utilizaron como hashtag la clave de la asignatura para identificar correctamente al profesor a evaluar. La EED es un cuestionario en el cual los alumnos evalúan al docente por medio de 20 ítems con base en un rango numérico del 1 al 10, siendo la máxima 10 (Muy satisfecho) y la mínima 1 (Muy insatisfecho); por último, en el ítem 21 se permite al alumno ingresar un comentario acerca de la clase o desempeño del profesor. La finalidad de la EED es medir las habilidades del docente mediante 20 ítems evaluando aspectos como: el conocimiento sobre el tema, la planeación de clases, la mediación, la aplicación de evaluaciones y la actitud del docente (ver tabla 1).

Tabla 1. Habilidades a medir en la EED

Habilidades	Preguntas																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Conocimiento	X																			
Planeación		X	X	X																
Mediación					X	X	X	X												
Evaluación									X	X	X									
Actitudes												X	X	X	X	X	X	X		
Evaluación general																			X	X

El acceso a la EED se realiza mediante el uso de dispositivos móviles para facilitar la evaluación por parte del estudiante, que es aplicada dos veces en un periodo escolar. Es importante resaltar que para asegurar la honestidad de los comentarios, los cuestionarios son confidenciales; por lo tanto, los profesores no relacionan las respuestas con los estudiantes.

En este trabajo los autores determinan que los datos de la EED sean almacenados en una base de datos usando la aplicación móvil como una herramienta para recolectar las opiniones de los estudiantes. Esto resulta con una serie de beneficios, incluyendo el tiempo de extracción dado la facilidad que ofrece la conexión y el acceso por medio de los dispositivos; es una ventaja al reducir tiempos en el proceso del tratamiento de la información. Otro beneficio otorgado al proyecto en el uso de los dispositivos móviles es la coordinación en el momento de captura. En este punto se lanza a los alumnos un aviso del periodo de captura. Dada la alta disponibilidad de acceso a los dispositivos móviles por parte de los alumnos el periodo se redujo de tres a un día.

La EED usada en este trabajo corresponde a la evaluación realizada por 1,505 estudiantes de la UPA a 229 profesores que impartieron clases desde el segundo periodo escolar del 2015 al primer periodo escolar del 2016, obteniendo un total de 9,054 comentarios.

Mediante Twitter y la EED se recolectaron 9,854 comentarios. En total participaron 21 grupos de estudiantes de seis carreras. Cabe resaltar que todos los estudiantes pertenecen a la misma institución, el perfil de carrera es Ingeniería (pregrado), las edades oscilan entre 18 y 22 años y, además, poseen preferencias hacia el uso de medios y herramientas tecnológicas. De los 9,854 comentarios recolectados de Twitter y la EED se consideraron sólo aquellos comentarios libres de ruido o spam (caracterizado en este estudio como textos con caracteres raros, espacios vacíos sin opinión o comentarios sin relación a la evaluación docente), quedando un total de 5,002 comentarios a analizar, los cuales conformaron el denominado corpus de comentarios. Una vez obtenido el corpus de comentarios, fue necesario realizar un proceso de etiquetado manual, el cual involucró la participación de profesores de la institución con experiencia en evaluaciones docentes. Se definió identificar comentarios en tres categorías: positivos, negativos y neutrales. Los profesores identificaron 2,696 comentarios negativos y 2,306 comentarios positivos. Cabe resaltar que en este proceso no hubo comentarios clasificados como neutrales. Se clasificó a aquellos comentarios positivos como aquellos donde el alumno felicitaba al profesor o mostraba interés por su clase, por ejemplo: *“es una maestra que aporta bastante conocimiento sobre la materia”*, *“buen maestro colabora al aprendizaje uniforme con asesorías”*. Para el caso de los comentarios negativos se consideraron comentarios que manifiestan características negativas del profesor, por ejemplo: *“me confunde su manera de explicar”*, *“Es un bipolar y cambia las evaluaciones a su beneficio”*. Los comentarios neutrales son aquellos que no expresan una emoción positiva o negativa.

Por otro lado, con el objetivo de medir el desempeño del Modelo Computacional *SocialMining* con datos desconocidos por este, se conformó un segundo corpus de comentarios. Para ello, se aplicó un muestreo por conveniencia y se seleccionó a dos grupos de quinto cuatrimestre de la carrera de Ingeniería en Sistemas de la misma institución, a quienes se les pidió realizar la evaluación docente de sus 11 profesores actuales. El nuevo corpus de comentarios para el caso de estudio se conformó por un

total de 623 comentarios, de los cuales 304 fueron etiquetados manualmente como comentarios negativos, 305 como comentarios positivos y 14 comentarios como neutrales.

Como se mencionó anteriormente, en este trabajo se cuidó la identidad de los estudiantes al momento de escribir las opiniones de los docentes, por lo que los resultados del análisis de las opiniones se consideran confiables, ya que no solo se están considerando ítems donde el estudiante otorga una calificación numérica (situación donde comúnmente responden de manera rápida para terminar lo antes posible). La finalidad de este estudio es que el Modelo Computacional *SocialMining* analice las opiniones continuamente y no solo dos veces en un periodo escolar. Esto debido a que un seguimiento a las opiniones (retroalimentación) de los estudiantes puede mejorar considerablemente el proceso de enseñanza del profesor para facilitar el aprendizaje.

Arquitectura del Modelo Computacional *SocialMining*

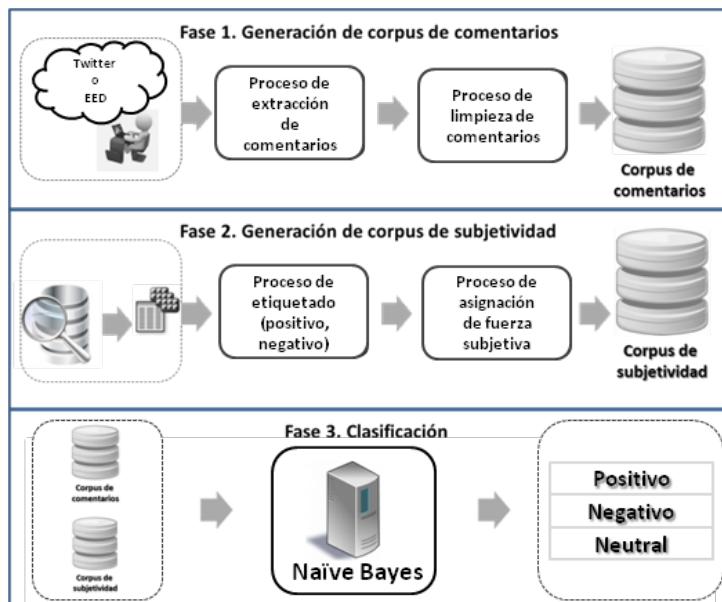
El Modelo Computacional *SocialMining* se compone de tres fases como se puede observar en la figura 1. Las cuales se explican enseguida.

Fase 1. Generación de corpus de comentarios

En esta fase se extraen los comentarios de los alumnos realizados en Twitter y la EED. La descarga de los comentarios de Twitter se realiza mediante el uso de la API de la misma red social. En cambio, los comentarios de la EED son obtenidos por medio de un archivo de la base de datos del sistema.

Una diferencia que se observa en los comentarios extraídos de ambos medios, es que la aplicación móvil EED no tiene restricciones en el número de caracteres, por lo cual es común encontrar comentarios con bastante ruido a diferencia de los comentarios de Twitter, por ejemplo: *“Creo que la maestra a veces se equivoca en reglas que son básicas en matemáticas. Además no me parece que su manera de explicar no es excelente ni de 10 porque SIEMPRE SE ENOJA MUCHO también muchas veces le da demasiadas vueltas al asunto de lo que debería: (y no siempre se va por el camino más fácil y práctico. Sin embargo, reconozco su disponibilidad y su motivación”*.

Figura 1. Arquitectura general del Modelo de *SocialMining*



Fuente: Gutiérrez, 2016 (elaboración propia)

Twitter es considerada una red social usada para redactar noticias e información propia de manera rápida y eficaz (Arrabal-Sánchez, 2016), permite al usuario usar solo 140 caracteres para realizar una publicación, por lo que puede generar publicaciones un poco más concretas que en otras redes sociales. Por esta razón y además por el servicio de un corrector ortográfico, se observa que la mayoría de las veces en esta red social los estudiantes comentan de manera concreta y puntual. No obstante, es posible encontrar el mismo tipo de ruido en los comentarios de Twitter que en la EED. Por lo cual, con el fin de minimizar este inconveniente y conformar los corpus de comentarios, se realiza un proceso de limpieza. Las actividades del proceso de limpieza se describen enseguida.

- **Revisión de ortografía.** En este proceso se hace una revisión de cada uno de los comentarios para corregir faltas de ortografía de acuerdo con el Diccionario de la Real Academia Española.
- **Proceso de tokenización.** Consiste en dividir los componentes (palabras) de un comentario en piezas, los cuales pasan a ser denominados como *tokens*. Por ejemplo el comentario: "Me gustaría que el profesor explicara más ejemplos", al pasar por el proceso de tokenización se representaría de la siguiente manera: "Me", "gustaría", "que", "el", "profesor", "explicara", "más", "ejemplos".

- Eliminación de *stop words*. Las palabras vacías comúnmente conocidas en el idioma inglés como *stop words*, son palabras sin significado como artículos, pronombres, sustantivos o preposiciones, entre otras.
- Conversión a minúsculas. En este proceso cada uno de los componentes del comentario es convertido a minúsculas para facilitar el proceso de comparación.

Fase 2. Generación de corpus de subjetividad

En esta fase se conforma el corpus de subjetividad basado en la metodología usada por Riloff (2003), quien realizó un corpus de este tipo para el idioma inglés constituido por 6,518 términos de diferentes enfoques.

En esta investigación todos los comentarios están escritos en el idioma español enfocados al contexto de la evaluación docente. Asimismo, se incluyen términos usuales conocidos como coloquialismos, correspondientes a la región donde se aplica el Modelo Computacional *SocialMining*. Por esta razón, se considera importante realizar un análisis de los comentarios con el fin de identificar las palabras o términos distintivos más usados, que apoyen a la clasificación de un comentario en positivo o negativo. Cabe resaltar que los términos usados para el corpus de subjetividad no necesariamente son aquellos que se repiten en una mayor cantidad de veces, sino aquellos que denotan una emoción, por ejemplo: excelente, fomenta, facilita, creativa, aburrida, impaciente, tolerante, entre otros términos.

Fase 3. Clasificación

En esta fase los datos de entrada, como puede observarse en la Figura 1, son los corpus obtenidos en las anteriores fases, el corpus de comentarios y el corpus de subjetividad. Para el proceso de clasificación en el Modelo Computacional *SocialMining* se aplica el algoritmo *Naïve Bayes* con una adaptación basada en el método de propuesto por Jurka (2012), quién incluye el cálculo de un valor (*score*) de cada uno de los términos del comentario que se encuentren en el corpus de subjetividad (Gutiérrez, 2016). La ecuación 1 representa el método propuesto por Jurka.

$$score = abs \left[\ln \left(\frac{scorefs * prior}{\sum count_{p,n}} \right) \right] \quad (1)$$

Donde: la variable posee el valor de la fuerza subjetiva que contiene el término, es decir, 0.5 para los términos etiquetados como *strongsubj* y 1.0 para los términos etiquetados como *weaksubj*, la variable es el valor numérico de la probabilidad, que en este trabajo es de 1.0 y, por último, la variable , es una variable tipo contador que

contiene el número de veces que aparecen los términos negativos o los términos positivos, dependiendo de la polaridad (positivo, negativo) del término evaluado. La figura 3 presenta el algoritmo para llevar a cabo el proceso de clasificación. En la siguiente sección se describen los métodos aplicados en este trabajo.

Métodos

Algoritmo de Naïve Bayes. El algoritmo de Naïve Bayes está basado en el teorema de Bayes (1763) y en la premisa de independencia de los atributos dada una clase. Asimismo, es uno de los métodos de aprendizaje supervisado más utilizados debido a que es posible adaptar para el análisis de emociones (Tan 2009; Kaur 2016).

El Modelo Computacional *SocialMining* implementa el método propuesto por Jurka (2012), el cual se basa en la ecuación 1. Este método utiliza una variable score para apoyar el proceso de clasificación, en lugar de utilizar sólo el método de la probabilidad. La figura 2 presenta el algoritmo del Modelo propuesto.

Figura 2. Algoritmo aplicado para ejecución del proceso de clasificación

- (1) Selección de corpus de comentarios a evaluar.
- (2) Selección de corpus de subjetividad.
- (3) Asignación de valores a variables *pstrong* y *pweak*.
- (4) Creación de matriz de comentarios (*matcomment*).
- (5) Lectura de corpus de subjetividad (*subjectivity*).
Repetir
 - (5.1) Match entre *matcomment* y *subjectivity*.
 - (5.2) Cálculo de score de términos en comentarios.
 - (5.3) Si *scores\$positive > scores\$negative* entonces.
 - (5.3.1) etiquetar comentarios como positivo.
 - (5.4) Si *scores\$positive < scores\$negative* entonces.
 - (5.4.1) etiquetar comentarios como negativo.
 - (5.5) Si *scores\$positive == scores\$negative* entonces.
 - (5.5.1) etiquetar comentario como neutral hasta que termine de recorrer matriz de comentarios.

Fuente: Gutiérrez, 2016 (elaboración propia)

Métricas para evaluar el desempeño en la clasificación

La matriz de confusión se considera el punto de partida para el cálculo de la medición del desempeño de un modelo predictivo; en este caso, presenta los resultados de clasificación del Modelo Computacional *SocialMining* con *Naïve Bayes*. Esta tabla (ver tabla 2) se deriva de la comparación de los resultados de un modelo predictivo contra los valores reales.

Tabla 2. Matriz de confusión para clasificación binaria

		Valor actual		
Valor predictivo	VP	FP		
	FN	VN		

Donde VP = verdadero positivo, FP = falso positivo, FN = falso negativo y VN = verdadero negativo.

Precisión (ACC). Se trata de una de las bases para evaluar el desempeño de un modelo predictivo. En una clasificación binaria, la precisión se calcula dividiendo el número de casos identificados correctamente entre el total de casos.

Precisión ponderada (WACC). Es una mejor estimación del rendimiento del clasificador cuando está presente en un conjunto de datos de una distribución desigual de las dos clases.

Margen de error. El margen de error mide la proporción de errores del clasificador en una serie de casos. Es el complemento de la precisión, es decir 100% - precisión.

Sensibilidad. La sensibilidad (VRP) mide la proporción de verdaderos positivos, que fueron identificados correctamente.

Especificidad. Mide la proporción de verdaderos negativos, que fueron identificados correctamente.

Curva ROC. El análisis ROC (*Receiver Operating Characteristic* por sus siglas en inglés) mide el rendimiento del clasificador binario. La curva ROC se considera una herramienta útil para medir el desempeño de algoritmos clasificadores (Spackman 1989).

EXPERIMENTACIÓN Y RESULTADOS

Experimentación

De acuerdo a los 5002 comentarios obtenidos en Twitter y la aplicación móvil EED explicado en la sección de Materiales y Métodos, una vez etiquetado el corpus de comentarios se procede a realizar un análisis mediante técnicas como la nube de palabras y el método de la frecuencia de ocurrencia del término en la colección de documentos, conocido por sus siglas en inglés como Tf-idf (Salton, 1986). Estos métodos son llevados a cabo mediante el programa R (R-Core-Team, 2013), con el fin de determinar los términos relevantes (features) a considerar para la conformación del corpus de subjetividad. La figura 3 presenta la nube de palabras generada del corpus de comentarios, una vez que este corpus ha pasado por el proceso de limpieza, omitiendo por lo tanto las *stop words*, sustantivos y términos considerados sin importancia. De esta manera, la nube de palabras denota términos distintivos

que pueden apoyar al proceso de clasificación, por ejemplo: debería, interesantes, amable, permite, apoya, paciencia, conoce entre otras.

Luego de identificar los términos distintivos que conforman el corpus de subjetividad, se etiqueta a cada uno con la polaridad correspondiente (positivo, negativo, neutral), asimismo se asigna la etiqueta de la fuerza subjetiva (*weaksubj* para los términos débiles y *strongsubj* para los términos fuertes) del término. Cabe resaltar que este proceso también es manual y realizado por profesores de la UPA. El corpus de subjetividad contiene en total 2,704 términos, de los cuales 1,014 son términos positivos, donde 598 son *strongsub* y 416 son *weaksubj*. De igual manera contiene 1,690 términos negativos, de los cuales 1,239 son *strongsub* y 451 son *weaksubj*.

En el proceso de clasificación se aplica el corpus de subjetividad con el algoritmo de Naïve Bayes para determinar el score de los términos que influyen en la clasificación del comentario. Por lo tanto el algoritmo de *Naïve Bayes*, además de considerar la cantidad de términos negativos y positivos en el corpus de subjetividad, también considera el *score* de cada término calculado. Una vez realizado el proceso de clasificación se calculan cada una de las métricas descritas en la sección II Materiales y Métodos.

Figura 3. Nube de palabras generada de corps de comentarios limpio



Fuente: Gutiérrez, 2016 (elaboración propia)

A lo largo de la investigación el Modelo Computacional *SocialMining* ha variado en sus resultados, por lo cual se considera que ha pasado por tres versiones, las cuales han variado en el uso de funciones y la actualización del corpus de subjetividad, provocando diferencias en los resultados de las métricas.

Resultados

El Modelo Computacional *SocialMining* ha pasado por tres versiones, la versión 3 del Modelo Computacional *SocialMining* es la más actual, en la cual se ha mejorado el proceso de clasificación. La tabla 3 presenta las matrices de confusión del proceso de clasificación de los comentarios, correspondiente a cada una de las versiones.

Tabla 3. Matrices de confusión de versiones de *SocialMining*

	Versión 1		Versión 2		Versión 3	
	Valor Actual					
Valor predictivo	2270	2000	2197	458	2173	363
	33	691	95	2226	118	2260
Total	4,994		4,976		4,914	

Fuente: Gutiérrez, 2016 (elaboración propia)

Es necesario indicar que en las matrices de confusión de cada una de las versiones no se completan los 5002 comentarios del corpus, esto se debe a la variación de resultados en la clasificación de comentarios por el Modelo Computacional *SocialMining*, el cual clasificó algunos comentarios como neutrales.

En la versión 1, resultaron 26 comentarios etiquetados como neutrales, en la versión 2, son 8 comentarios etiquetados como neutrales y, por último, en la versión 3 se etiquetaron 88 comentarios neutrales. Una de las razones en la variación de la clasificación de los comentarios, se debe a que en el modelo propuesto no se considera la semántica, por lo que no detecta aquellos comentarios que cambian su polaridad de positiva a negativa o de negativa a positiva, debido a un término de contraposición (por ejemplo: pero, sin embargo, no obstante, aunque, entre otros); en consecuencia, esto, provoca un error en la clasificación.

A pesar de que el Modelo Computacional *SocialMining* en la versión 3 etiqueta más comentarios neutrales a diferencia de la versión 1 y 2, hay una mejora en el desempeño de la precisión total y ponderada. La tabla 4 presenta los resultados de las métricas: sensibilidad (VRP), precisión total (ACC), precisión ponderada (WACC) y especificidad (SPC) de cada una de las versiones del Modelo Computacional *SocialMining*. En esta tabla se puede observar la mejora en resultados de clasificación de la versión 3, la cual alcanza poco más de 0.90 en las métricas de precisión total y ponderada, resultado que se considera aceptable en la literatura (Mejova, 2009; Prasad, 2010; Gupte, 2014).

Tabla 4. Resultados de métricas de cada versión

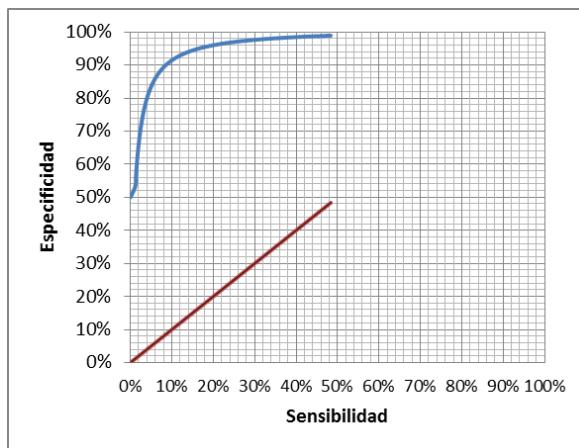
	Valor (versión 1)	Valor (versión 2)	Valor (versión 3)
VRP	0.985670864	0.958551483	0.948494107
SPC	0.256781865	0.829359165	0.861608845
ACC	0.592911494	0.888866559	0.902116402
WACC	0.743017907	0.893282329	0.903619834

Fuente: Gutiérrez, 2016 (elaboración propia)

Considerando los resultados de las métricas VRP y SPC de la versión 3 del Modelo Computacional *SocialMining*, se procede a obtener una relación de valores predictivos, con el propósito de construir la curva ROC, en la cual mediante la curva puede observarse el incremento del desempeño en la clasificación de comentarios del Modelo Computacional *SocialMining* (ver figura 4).

Según investigaciones de la literatura (Mejova 2009; Prasad 2010; Gupte 2014), la técnica de Náive Bayes aplicada en el análisis de emociones en comentarios puede lograr buenos resultados en procesos de clasificación; en este caso, el resultado de la precisión total y de la precisión balanceada del Modelo Computacional *SocialMining* fue de un 0.90, este resultado se considera factible y se atribuye al uso de Náive Bayes en combinación con el corpus de subjetividad.

Figura 4. Curva ROC obtenida con base a los valores predictivos del caso de estudio



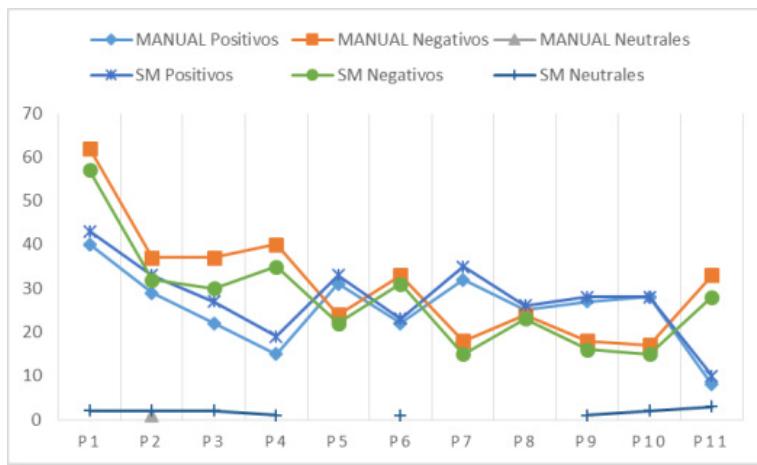
Fuente: Gutiérrez, 2016 (elaboración propia)

Caso de estudio

Una vez obtenidas las métricas del Modelo Computacional *SocialMining*, se realizó un caso de estudio con el fin de medir el desempeño del mismo modelo con un corpus de comentarios nuevo, el cual se conformó por estudiantes del 5º cuatrimestre de la carrera de Ingeniería en Sistemas que evaluaron a 11 profesores. Este corpus es descrito en la sección II de Materiales y Métodos.

Posteriormente al proceso de clasificación por el Modelo Computacional *SocialMining* se procedió a comparar contra el nuevo corpus de comentarios generado para este estudio. La figura 5 muestra los resultados de esta comparación; en la parte horizontal de la gráfica se ubican los 11 profesores, la parte vertical presenta el número de comentarios positivos o negativos obtenidos. Los comentarios etiquetados manualmente se identifican por la palabra “manual”, mientras que los comentarios etiquetados por el Modelo Computacional *SocialMining* se identifican por la palabra “SM”. El corpus utilizado en este caso de estudio es diferente, con el fin de comparar el desempeño en clasificación por el algoritmo. Tal como se explicó en la sección de Materiales y Métodos, está conformado por un total de 623 comentarios, de los cuales 304 son comentarios negativos, 305 son comentarios positivos y 14 son comentarios neutrales.

Figura 5. Gráfico de valores predictivos obtenidos en caso de estudio



Fuente: Gutiérrez, 2016 (elaboración propia)

Como se puede observar en la figura 5 los resultados de la clasificación de comentarios del Modelo Computacional *SocialMining* no varían substancialmente de los comentarios etiquetados manualmente; puesto que clasifica correctamente

273 comentarios negativos de 304 y 302 comentarios positivos de 305. El resultado de las métricas obtenidas, se presentan en la tabla 5.

Tabla 5. Resultados de la precisión y la sensibilidad

ACC	WACC	VPR	SPC
0.936482085	0.936338172	0.978494624	0.901492537

Fuente: Gutiérrez, 2016 (elaboración propia)

De igual manera en este caso de estudio se clasificaron 9 comentarios como neutrales falsos. Sin embargo, como puede observarse los valores de ACC y ACCW, incrementaron de un 0.90 (ver tabla 4) a un 0.93 (ver tabla 5), lo cual denota que el Modelo Computacional *SocialMining* puede mejorar con el uso del corpus de subjetividad en combinación con el algoritmo de Naïve Bayes.

CONCLUSIONES

La evaluación del profesor por el alumno mediante la aplicación de cuestionarios a estudiantes ha sido de gran utilidad para la retroalimentación de la labor del profesorado. Esto se facilita cuando existe una calificación cuantitativa en cada una de las dimensiones evaluadas mediante las preguntas. Sin embargo, esta tarea requiere mayor dedicación y análisis cuando el cuestionario involucra un apartado especial para la expresión de comentarios libres por parte de los alumnos.

Lo anterior tiene como consecuencia que éstos no se lleguen a considerar. El Modelo Computacional *SocialMining* propuesto ayuda a tener un valor cualitativo respecto a estas opiniones, determinando si el comentario es positivo o negativo, lo cual puede implicar un desempeño bueno o malo del profesor. Además, la recopilación y el análisis automático periódico pueden favorecer un cambio en la actuación docente.

El análisis de los comentarios recolectados en Twitter y alimentados en la aplicación móvil EED favorecen la adecuación del corpus de subjetividad mediante la identificación de términos distintivos. El corpus de subjetividad en combinación con el algoritmo de Naïve Bayes mejora el proceso de clasificación. Sin embargo, para la mejora continua en los resultados de clasificación se considera importante mantener actualizado el corpus de subjetividad, ya que, por ejemplo, los coloquialismos varían con el tiempo y la región; además, es necesario verificar que la asignación de futuras etiquetas en los términos corresponda con el significado de éstos.

El uso de dispositivos móviles en la aplicación móvil EDD arrojó beneficios al proceso completo tales como conectividad, ubicuidad y permanencia. En la literatura se ha documentado que “las tecnologías móviles han redibujado el panorama

educativo, aportando a la educación, no solo movilidad, sino también conectividad, ubicuidad y permanencia" (Cantillo, 2012, p. 3).

La aplicación del Modelo Computacional *SocialMining* como apoyo para la evaluación del profesor por el alumno resulta factible para diferentes aspectos a mejorar en el proceso de enseñanza. Asimismo, la generación de un corpus en español dedicado a la educación y específicamente enfocado a la evaluación del profesor por el alumno permite ser aplicado en diferentes instituciones de educación superior. La ventaja de la adecuación del corpus de subjetividad en combinación con Náïve Bayes puede derivar incluso en una implementación de un sistema recomendador, para los casos en los cuales sea detectada una debilidad en el desempeño del profesor, pudiendo ofrecer a este curso de formación de competencias, docentes que contribuyan a la mejora de sus clases.

De la misma forma, mediante la aplicación del Modelo Computacional *SocialMining* fue posible observar ciertos comportamientos de los estudiantes en sus comentarios, que pueden beneficiar a la mejora del Modelo mencionado. Una de las observaciones percibidas mediante el análisis de comentarios es que los estudiantes por lo regular usan la evaluación docente como un medio para quejarse de las clases, si existe un caso o situación que no les agrada, muy pocos aplauden el esfuerzo del profesor y sugieren estrategias o cambios para mejorar la clase. Por ejemplo, los comentarios de grupos donde el índice de aprovechamiento es alto sugieren al profesor cambios constructivos con el fin de aprender mejor e incluso existen comentarios donde reconocen el esfuerzo del profesor al impartir clases, contrariamente a lo que sucede en algunos otros grupos donde no muestran interés por evaluar a sus profesores y se observa un bajo rendimiento académico.

Otra observación del análisis de comentarios que se detecta como problema para el proceso de clasificación del Modelo Computacional *SocialMining* es que los estudiantes utilizan palabras negativas para expresar comentarios positivos, por ejemplo "*el profesor no es tan malo en sus clases*"; por lo cual, para mejorar el desempeño de la clasificación de comentarios en el Modelo Computacional *SocialMining*, se vuelve necesario explicar a los estudiantes cómo expresar correctamente un comentario. No obstante, esta observación puede ser considerada también como un área de oportunidad para implementar técnicas de Procesamiento de Lenguaje Natural y semántica con el fin de minimizar dicho problema.

A pesar de los problemas detectados en este trabajo con base en el análisis de comentarios, los resultados obtenidos en las métricas indican que el desempeño del Modelo Computacional *SocialMining* en el proceso de clasificación de comentarios de la evaluación por el alumno es factible. Sin embargo, se resalta la importancia de mantener actualizado el corpus de subjetividad, debido a la variedad de coloquialismos con la que se pueden expresar los estudiantes de las instituciones. De igual manera, es posible aplicar el Modelo propuesto a otro contexto diferente, siempre y cuando se adecúe el corpus de subjetividad al área correspondiente.

AGRADECIMIENTOS Y CONTRIBUCIÓN DE AUTORES

Las autoras contribuyeron de igual manera a este trabajo. Las autoras de este artículo agradecen el apoyo de PRODEP y CONACYT.

REFERENCIAS BIBLIOGRÁFICAS

- Altrabsheh, N., Cocea, M., y Fallahkhair, S. (2014). Learning Sentiment from Students' Feedback for Real-Time Interventions in Classrooms. *Adaptive and Intelligent Systems*. Volume 8779 of the series *Lecture Notes in Computer Science*, 40-49. doi: http://dx.doi.org/10.1007/978-3-319-11298-5_5
- Arrabal-Sánchez, G., y De-Aguilera-Monyano, M. (2016). Comunicar en 140 caracteres. Cómo usan Twitter los comunicadores en España. *Revista Científica de Educomunicación. Comunicar*, 46(XXIV). 9-17. Recuperado de <http://goo.gl/YOqybX>
- Bayes T. (1763). An Essay towards solving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. *Philosophical Transactions* (1683-1775), 370-418.
- Bravo, E., Pedraza, A., y Herrera, L. (2013). *Educación 2.0: Twitter como herramienta de aprendizaje de la Ingeniería*. Latin American and Caribbean Consortium of Engineering Institution.
- Brusilovsky, P. (2001). Adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 11(1-2), 87-110.
- Cantillo, V., Roura, R., y Sánchez, P. (2012). Tendencias Actuales en el uso de dispositivos móviles en educación. *La Educación Digital Magazine*, 147.
- Córdova, K. E. G., y González, J. R. V. (2015). Evaluación del desempeño: acercando la investigación educativa a los docentes. *REVALUE*, 3(2). Recuperado de <http://revalue.mx/revista/index.php/revalue/issue/current>
- Crovi, D., y Lemus, M. C. (2014). Jóvenes estudiantes y cultura digital: una investigación en proceso. *Virtualis*, 9, 36-55. Recuperado de <http://goo.gl/8emHtj>
- Elizalde Lora, L., y Reyes Chávez, R. (2008). Key Elements for the Evaluation of the Teaching Performance. *Revista electrónica de investigación educativa*, 10(spe), 1-13.
- Gewerc, A., Montero, L., y Lama, M. (2014). Colaboración y redes sociales en la enseñanza universitaria [Collaboration and Social Networking in Higher Education]. *Comunicar*, 21(42), 55-63. doi: <https://doi.org/10.3916/C42-2014-05>
- González-Lizárraga, M., Becerra-Traver, M., y Yanez-Díaz, M. (2016). Ciberactivismo: nueva forma de participación para estudiantes universitarios. *Comunicar*, 24(46), 47-54, doi: <http://dx.doi.org/10.3916/C46-2016-05>
- Guerrero, C., Jaume, A., Juiz, C., y Lera, I. (2016). Use of Mobile Devices in the Classroom to Increase Motivation and Participation of Engineering University Students. *IEEE Latin America Transactions*, 14(1), 411-416.
- Gupte, A., Joshi, S., Gadgul, P., y Kadam, A. (2014). Comparative Study of Classification Algorithms used in Sentiment Analysis. *International Journal of Computer Science and Information Technologies*, 5(5), 6261-6264. Recuperado de <http://goo.gl/tiHBT>
- Gutiérrez, G., Padilla, A., Canul-Recih, J., De-Luna, P., y Ponce, J. (2016). Proposal of a Sentiment Analysis Model in Tweets for improvement of the teaching - learning

- process in the classroom using a corpus of subjectivity. *International Journal of Combinatorial Optimization Problems and Informatics*, 7(2), 22-34.
- Jurka, T. (2012). *Sentiment: Tools for Sentiment Analysis*. R package version 0.1, Recuperado de <http://goo.gl/oxASCV>
- Kaur, G., y Singla, A. (2016). Sentimental Analysis of Flipkart reviews using Náïve Bayes and Decision Tree algorithm. *International Journal of Advanced Research in Computer Engineering & Technology*, 5(1), 148-153.
- Liu, B. (2010). Sentiment Analysis and Subjectivity. In N. Indurkhia y F. J. Damerau (Eds.), *Handbook of natural language processing*, 627-666. Chapman and Hall: CRC Press.
- Loureiro, S., Míguez, M., y Otegui, X. (2016). Desempeño docente en la enseñanza universitaria: análisis de las opiniones estudiantiles. *Cuadernos de Investigación Educativa*, 7(1), 55-67. Recuperado de <https://goo.gl/hm1eAq>
- Martínez González, A., Sánchez Mendiola, M., y Martínez Stack, J. (2010). Los cuestionarios de opinión del estudiante sobre el desempeño docente: Una estrategia institucional para la evaluación de la enseñanza en Medicina. *Revista electrónica de investigación educativa*, 12(1), 1-18.
- Mejova, Y. (2009). *Sentiment Analysis: An Overview*. Comprehensive Exam Paper. Recuperado de <https://goo.gl/xsFTVg>
- Moreno, R. D., Cepeda, I. M. L., y Romero, S. P. (2004). El modelo de evaluación, intervención y análisis de procesos como propuesta de diseño instruccional. *Enseñanza e Investigación en Psicología*, 9(2), 271-291.
- Novak, J., y Cowling, M. (2011). *The implementation of social networking as a tool for improving student participation in the classroom*. Hobart: ISANA International Education Association Inc. Recuperado de <http://goo.gl/IW6Igc>
- Ortigosa, A., Martín, J., y Carro, R. (2014). Sentiment analysis in Facebook and its application to e-learning. *Computers in Human Behavior*, 31, 527-541, doi: <http://dx.doi.org/10.1016/j.chb.2013.05.024>
- Prasad, S. (2010). *Micro-blogging Sentiment Analysis Using Bayesian Classification Methods*. CS224N Project Report, Stanford. Recuperado de <http://goo.gl/W2koQT>
- R-Core-Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Recuperado de <http://goo.gl/e4oyiU>
- Riloff, E., y Wiebe, J. (2003). Learning extraction patterns for subjective expressions. *Conference on Empirical Methods in Natural Language Processing*. (pp. 105-112). Recuperado de <https://goo.gl/seoIg>
- Ruiz Carrascosa, J. (2005). La evaluación de la docencia en los planes de mejora de la Universidad, *Educación XXI*, 8, 87-102.
- Salton, G., y McGill, M. J. (1986). *Introduction to modern information retrieval*.
- Spackman, K. A. (1989). Signal detection theory: Valuable tools for evaluating inductive learning. *Proceedings of the Sixth International Workshop on Machine Learning*. San Mateo, CA: Morgan Kaufman, (pp. 160-163).
- Tan, S., Cheng, X., Wang, Y., y Xu, H. (2009). Adapting naive bayes to domain adaptation for sentiment analysis. *Advances in Information Retrieval*, 337-349.
- Tirado Segura, F., Miranda Díaz, A., y Sánchez Moguel, A. (2007). La evaluación como proceso de legitimidad: la opinión de los alumnos. Reporte de una experiencia. *Perfiles educativos*, 29(118), 7-24.
- Valencia, A., González G., y Castañeda M. (2016). Structural Equation Model for Studying the Mobile-Learning Acceptance.

- IEEE Latin America Transactions, 14(4), 1988-1992.*
Zaldivar, A., Tripp, A., Aguilar, J., Tovar, J., y Anguiano, C. (2015). Using Mobile Technologies to Support Learning in Computer Science Students. *IEEE Latin America Transactions, 13(1), 377-382.*

PERFIL ACADÉMICO Y PROFESIONAL DE LAS AUTORAS

Guadalupe Gutiérrez Esparza. Profesora de Tiempo Completo de Ingeniería Mecatrónica en la Universidad Politécnica de Aguascalientes. Posee Doctorado en Ciencias de Computación de la UJAT. Sus líneas de investigación giran en torno al desarrollo de algoritmos de inteligencia artificial aplicados a la minería de textos con enfoque a contextos de la educación y ciencias sociales.

E-mail: guadalupe.gutierrez@upa.edu.mx

María de Lourdes Margain Fuentes. Doctora en Ciencias de la Computación. Miembro del Sistema Nacional de Investigadores. Ha desempeñado funciones directivas a nivel Posgrado y Pregrado en la Universidad Politécnica de Aguascalientes. Evaluador nacional CIEES y CONAIC. Sus líneas de investigación y proyectos de desarrollo tecnológico atienden la ingeniería de software e inteligencia artificial.

E-mail: lourdes.margain@upa.edu.mx

Tania Aglaé Ramírez del Real. Coordinadora de la Maestría en Ciencias en Ingeniería de la Universidad Politécnica de Aguascalientes desde octubre de 2015. Actualmente se encuentra estudiando el Doctorado en Ciencia y Tecnología con orientación en Opto-Electrónica en la Universidad de Guadalajara. Sus áreas de interés abarcan educación en ingeniería, control adaptable e inteligente, automatización y aplicaciones ópticas.

E-mail: tania.ramirez@upa.edu.mx

DIRECCIÓN DE LAS AUTORAS

Universidad Politécnica de Aguascalientes
Calle Paseo San Gerardo No. 207
Fracc. San Gerardo Aguascalientes,
Ags. C.P. 20342. Aguascalientes, México

Juana Canul Reich. Posee un Doctorado en Ciencias de la Computación e Ingeniería de la Universidad del Sur de Florida (USF). Actualmente es profesora de la División Académica de Informática y Sistemas de la UJAT. Miembro del Sistema Nacional de Investigadores. Sus líneas de investigación comprenden la minería de

datos y aprendizaje automático con enfoque en el análisis de datos de microarrays, selección de características y preprocesamiento de datos y análisis de datos.

E-mail: juana.canul@ujat.mx

DIRECCIÓN DE LA AUTORA

Universidad Juárez Autónoma de Tabasco
Av. Universidad, Manuel Sánchez Márquez
Cunduacán, Tab. C.P. 86690. Tabasco, México

Fecha de recepción del artículo: 15/12/2016

Fecha de aceptación del artículo: 02/02/2017

Como citar este artículo:

Gutiérrez, G., Margain, M., Ramírez, T., y Canul, J. (2017). Un modelo basado en el Clasificador Náive Bayes para la evaluación del desempeño docente. *RIED. Revista Iberoamericana de Educación a Distancia*, 20(2), pp. 293-313. doi: <http://dx.doi.org/10.5944/ried.20.2.17717>