

ANÁLISIS DE LA POLARIDAD EN COMENTARIOS DE ESTUDIANTES UNIVERSITARIOS SOBRE EL DESEMPEÑO DE SUS PROFESORES

POLARITY ANALYSIS OF UNIVERSITY STUDENT COMMENTS ABOUT THEIR PROFESSORS PROFESSIONAL PERFORMANCE

Ángeles Belém Priego Sánchez

Universidad Autónoma Metropolitana
abps@azc.uam.mx

David Eduardo Pinto Avendaño

Benemérita Universidad Autónoma de Puebla
dpinto@cs.buap.mx

Mauricio Castro Cardona

Benemérita Universidad Autónoma de Puebla
mayka2000@yahoo.com.mx

Martha Patricia León Arenas

Benemérita Universidad Autónoma de Puebla
patricia.leon@correo.buap.mx

Resumen

El análisis de la polaridad, en cualquier tipo de comentario, es una tarea que está teniendo un gran auge, debido a que actualmente existe un fuerte interés en determinar automáticamente si las opiniones publicadas en medios públicos tienen un carácter positivo o negativo. La minería de opiniones se enfoca en determinar la polaridad de las publicaciones para, generalmente, dar seguimiento a la reputación de una entidad. Este artículo presenta los resultados obtenidos tras analizar la polaridad de comentarios emitidos por estudiantes universitarios en relación a su percepción sobre el desempeño de sus profesores frente a grupo. El enfoque del artículo es estudiar el rendimiento de diversos clasificadores supervisados en la tarea, anteriormente mencionada, con la finalidad de construir a mediano plazo un sistema para la clasificación automática de la polaridad de comentarios escritos por estudiantes universitarios. Los resultados obtenidos muestran que es posible crear un modelo de clasificación basado en máquinas de

vectores de soporte que clasifique correctamente la polaridad de comentarios universitarios con un porcentaje de efectividad cercano al 92% usando un conjunto balanceado de 10,000 datos de entrada.

Palabras Clave: Análisis de la polaridad, minería de opiniones, clasificación supervisada.

Abstract

Analysis of polarity, in any type of document, is a task that is having a great boom. Nowadays, there is a high interest in the construction of automatic methods for determining whether or not opinions published in public media have a positive or negative polarity. Opinion mining focuses on determining the polarity of publications, such as reviews or text comments, in order to analyze the reputation of a given entity. This paper presents the results obtained after analyzing the polarity of university student comments. The aim of this paper is to study the performance of different supervised classifiers in the aforementioned task in order to further construct a computational system for the automatic classification of comments polarity. The results obtained up to now show that it is possible to create a classification model based on support vector machines that correctly classifies the polarity of university student comments with a percentage close to 92%, using a balanced dataset of 10,000 samples.

Keywords: *Analysis of polarity, opinion mining, supervised classification.*

1. Introducción

El análisis de sentimientos es una tarea de clasificación de textos dentro del área del procesamiento del lenguaje natural cuyo objetivo consiste en detectar la polaridad (positiva, negativa o neutra), de una opinión dada por un cierto usuario. El conocer la opinión que una persona tiene hacia un producto o servicio es de gran ayuda para toma de decisiones, ya que permite, entre otras cosas, que posibles consumidores verifiquen calidad del producto o servicio antes de utilizarlo. En este artículo se presentan los resultados obtenidos para identificar automáticamente la polaridad de comentarios emitidos por estudiantes

universitarios en una encuesta correspondiente al desempeño de sus profesores. Para llevar a cabo la identificación de la polaridad de comentarios, se utiliza una técnica basada en aprendizaje automático, por lo que inicialmente se realiza un etiquetamiento manual de los comentarios y posteriormente estos resultados permiten alimentar diferentes algoritmos de aprendizaje con el fin de crear los modelos de clasificación que serán usados para etiquetar automáticamente nuevos comentarios, y determinar así su polaridad (positiva o negativa). Para llevar a cabo la tarea planteada, se diseñaron tres diferentes conjuntos de datos de entrenamiento conformados de una cantidad balanceada de comentarios positivos y negativos (C1000, C5000 y C10000), creados a partir de un conjunto base de 201,146 comentarios proporcionados por el Programa Institucional de Evaluación Académica (PIEVA) de la Benemérita Universidad Autónoma de Puebla (BUAP). Dado que esta investigación se encuentra enmarcada en el ámbito de la minería de opinión, se presenta a continuación el estado del arte relacionado con el trabajo. Dentro de los estudios centrados en la clasificación de opiniones, positivas o negativas, se encuentra el presentado en [Pang, 2002], una de las primeras investigaciones sobre el análisis de sentimientos en la cual se utilizan datos de críticas de películas encontradas en la Web; éstos son empleados en tres algoritmos de clasificación, superando los *baselines* producidos manualmente por un humano. Publicaron su trabajo sobre la clasificación de documentos en base al sentimiento expresado en éstos. Analizaron reseñas sobre películas, encontraron que las técnicas de Aprendizaje Automático mejoran el rendimiento de las líneas base generadas por los expertos humanos. Emplearon tres algoritmos de Aprendizaje Automático: Naïve Bayes (NB), Máxima Entropía (ME) y Máquinas de Soporte Vectorial (SVM).

Uno de los pioneros que introdujeron el término de Análisis de Sentimientos fue el presentado por [Nasukawa & Yi, 2003]. En esta publicación definen esta tarea como encontrar expresiones de sentimientos para un sujeto dado y determinar la polaridad de estos. En las investigaciones anteriores a ésta, se realizaba el análisis de la polaridad general de un documento. Sin embargo, en este enfoque se trata de identificar la opinión de cada sujeto mencionado en el texto.

En [Minqing & Bing, 2004], se expuso una propuesta para minar y resumir reseñas de consumidores. Los objetivos de este trabajo fueron encontrar las características a las cuales se hacían referencia en las críticas, identificar los enunciados que expresaban opiniones y polaridad sobre las mismas y resumir los resultados. Se propuso la creación de una pequeña lista de adjetivos “semilla” etiquetados manualmente dependiendo si expresan sentimiento positivo o negativo. Posteriormente esta lista es aumentada usando WordNet [Miller et al., 1990].

Las opiniones, fabricadas mediante una conversación o comentario, es posible consultarlas a través de foros, blogs o redes sociales, estas últimas siendo la novedad y teniendo el mayor auge actual. Un trabajo basado en una colección de entradas de blogs, es el presentado en [Fernández, 2011] que realiza el análisis de sentimientos y minería de opiniones en dichas entradas, mostrando la relevancia de los sistemas de aprendizaje automático como recurso para la detección de información de opinión. En [Carrillo-de-Albornoz, 2016], se propone una metodología para evaluar resúmenes en el contexto del monitoreo de la reputación en línea, la cual beneficia los informes de reputación.

Actualmente, muchos investigadores del ámbito de la recuperación de información y la lingüística computacional han focalizado sus investigaciones en las redes sociales, especialmente en Twitter, dada la gran cantidad de publicaciones que se generan día a día. De esta manera, se hace necesaria la creación de herramientas capaces de gestionar, analizar y manipular toda esta información disponible, además de categorizarla de acuerdo al contenido generado por el usuario, y -en relación a las opiniones- identificar los matices de opinión vinculados a la posición de los usuarios con respecto a algún tema. Esto se basa en el constante monitoreo de los mensajes producidos en las redes sociales. Hasta el 2014, estos sistemas eran evaluados en el marco de la competencia RepLab [Amigó, 2013] para tuits en inglés y en español. A pesar de incluir el idioma español, dentro de la competencia, la mayoría de los trabajos reportados en la literatura se centran en el idioma inglés. En este sentido, los diferentes métodos aplicados para la clasificación de opiniones de tuits en inglés han sido aplicados para el idioma español [Fernández, 2013]. Estos métodos han considerado desde la utilización de n-gramas de palabras, la

reducción de éstas a su raíz (*stemming*) e incluso su sustitución. Sin embargo, no se obtuvieron los resultados esperados dado que los tuits son muy difíciles de tratar, sobre todo debido a su brevedad y falta de contexto. Tal y como regularmente pasa con lo métodos supervisados, es difícil encontrar conjuntos de datos manualmente anotados para entrenar modelos automáticos de clasificación. Sin embargo, este problema ha disminuido al contar con ambientes de competencias internacionales relacionadas con el análisis de sentimientos. El foro de SemEval (International Workshop on Semantic Evaluation) es sin duda un espacio que ha dado difusión al estudio de la polaridad en Tweets. Desde el año 2013 ha contado con soporte de evaluación de polaridad en Tweets. En particular, en el año 2017, se presentaron 58 equipos con diferentes propuestas, las cuales son resumidas en [Rosenthal et al., 2017]. La más reciente edición de SemEval ha destacado por tener una sección (con tres tareas) dedicadas exclusivamente al análisis de la característica afectiva en Tweets. En [Mohammad et al., 2018] puede verse una descripción completa de los 75 equipos que participaron en la tarea 1 de SemEval denominada "Affect in Tweets". Ellos describen la manera en que se construyeron y etiquetaron manualmente los corpora para los idiomas Inglés, Árabe y Español, así como las técnicas usadas en las 319 ejecuciones presentadas por los equipos. Cada uno de los equipos ha reportado de manera individual su trabajo, sin embargo, referirse a 58 artículos del año 2017 y 75 artículos del año 2018 sería demasiado y es por esta razón que remitimos al lector a los reportes que resumen la participación de todos los equipos.

En [Sidorov, 2012] se presenta un enfoque, para la minería de opinión de tuits en español, basado en el funcionamiento y diferentes configuraciones de algoritmos de aprendizaje automático. A pesar de que los algoritmos empleados presenten buenos resultados para el idioma inglés, en este trabajo se muestra como los diferentes tamaños de *n-gramas*, la longitud del corpus, el número de clases de sentimientos, el corpus balanceado con respecto al corpus no balanceado y los diferentes dominios (configuraciones) afectan la precisión del algoritmo.

La generación de léxicos de palabras, que se encuentren anotadas con su correspondiente polaridad, es otro enfoque en el que se han orientados diferentes

investigadores y que ha colaborado al monitoreo de las opiniones en los comentarios. En el caso del español, ejemplos de estos enfoques son los presentados en [Pérez-Rosas, 2012] y [Brooke, 2009].

Hacer uso de los datos generados por personas, comentarios, es una gran oportunidad para ganar tiempo en las decisiones tomadas, debido a que proporcionan información que puede ser utilizada en diferentes ámbitos. A partir de los datos adquiridos se puede realizar un análisis automático y generar estadísticas sobre la opinión colectiva (positiva o negativa) de un producto, servicio o persona. Dicho análisis es de gran utilidad para los analistas de medios, desde la disminución de tiempos hasta la disminución de costos. Por ejemplo, en los estudios manuales se nota un consumo inmenso de tiempo y costos.

Por ello, en este artículo se presenta un análisis de diferentes modelos de clasificación automática que determina el rendimiento de cada uno de ellos en la tarea de identificación automática de la polaridad de comentarios hechos por estudiantes universitarios.

El resto de este artículo está organizado como sigue. La Sección 2, presenta la descripción de los métodos y datos usados en los experimentos. La Sección 3, presentan los resultados obtenidos. En la sección 4 se presenta la discusión de estos resultados usando datos concentrados globales. Finalmente, en la Sección 5 se exponen las conclusiones y el trabajo futuro.

2. Métodos

En esta sección se presentan los métodos de clasificación automática analizados para la detección automática de la polaridad en comentarios de estudiantes universitarios, así como el conjunto de datos usados en el experimento.

Descripción de los datos

El instrumento que sirve de base para el análisis, realizado en este artículo, fue diseñado por Programa Institucional de Evaluación Académica (PIEVA) de la Benemérita Universidad Autónoma de Puebla (BUAP) y es utilizado para obtener

información de los procesos de enseñanza aprendizaje; con la finalidad de posibilitar las estrategias de intervención que permitan mejorar la función docente. El programa surgió desde el 2002 y hasta la fecha ha ido mejorando en cuanto a su validez y confiabilidad; lo que ha permitido mejoras sustanciales en el desempeño docente. El Instrumento PIEVA, evalúa las siguientes dimensiones:

- **Mediación.** Entendida como la serie de acciones que el docente plantea al estudiante, en función de lo que ya conoce y hace este último, para transformar y ampliar su nivel de conocimientos sobre las temáticas abordadas y su realidad próxima.
- **Estrategias y recursos.** Plan de acción que se lleva a cabo para lograr un determinado fin a largo plazo; se especifican aquí los recursos que el profesor emplea de forma reflexiva y flexible para el logro de aprendizajes significativos.
- **Resultados de aprendizaje.** El trabajo del docente está centrado en el logro de resultados tangibles para el estudiante en cuanto al desarrollo de sus habilidades sus conocimientos y sus actitudes
- **Planeación.** La actividad docente requiere de la planeación, esto es: la articulación de los objetivos, contenidos del curso, metodologías, estrategias educativas y recursos, estableciendo una secuencia de actividades que permita el aprendizaje deseado y el aprovechamiento del tiempo.
- **Evaluación de aprendizajes.** La evaluación de aprendizajes: declarativo, procedimental y actitudinal-valoral, debe contemplar –al menos- los siguientes principios: dar a conocer oportunamente los criterios bajo los cuales se acreditará la materia, ser justa (los mismos para todos), favorecer la retroalimentación oportuna que permita los reajustes cognitivos y de proceso y ser objetiva (centrada en lo que se espera se aprenda de acuerdo al programa).
- **Ejes transversales.** Son el conjunto de características que definen a un modelo académico educativo y que para el caso que se presenta, corresponden al denominado Modelo Universitario Minerva (MUM). Los considerados son: Formación Humana y Social (FHS), Desarrollo de Habilidades del Pensamiento Complejo (DHPC), Desarrollo de habilidades en

el uso de las tecnologías de la información y comunicación (DHTIC), el empleo de una segunda lengua, educación para la investigación y cultura emprendedora.

- **Relacional.** El ejercicio docente requiere un conjunto de habilidades sociales para interactuar respetuosamente con los estudiantes, a saber: motivación, reconocimiento y respeto a la diferencia y al diálogo, incluyendo la comunicación asertiva y la resolución de conflictos.
- **Cumplimiento Institucional.** Compuesto de 4 indicadores: Cumplimiento del horario, asistencia, cobertura del programa y dominio de la materia.

El cuestionario se responde en una escala del 1 al 4 y de acuerdo con las respuestas se calcula un Índice de Satisfacción Ponderada (ISP) que es el que se utiliza junto con otros criterios, para evaluar el desempeño del docente en el aula. Al finalizar el cuestionario existe una sección para comentarios abiertos a decisión del estudiante.

Dado que el objetivo de este trabajo es el análisis de la polaridad de comentarios de estudiantes universitarios, es necesario contar con una etiqueta manual que indique si el comentario es positivo o negativo. Para ello se construyó una tipología de clasificación –manual, y que parte del análisis de contenidos-, que contiene las siguientes categorías:

- **Ámbito.** Si es acerca de lo académico del profesor, de PIEVA o relacional.
- **¿Sobre qué?** Sobre la materia, sobre el profesor sobre otro profesor sobre la institución o sobre PIEVA
- **Tipo.** Positivo, negativo, denuncia, sugerencia y suplantación.
- **Atención.** Director, academia, escuela de formación docente o al mismo maestro.

Un comentario puede ser clasificado, por ejemplo, como del ámbito académico, referirse al profesor y, además, ser de tipo negativo y/ o emitir una sugerencia. Con este tipo de encuestas, realizadas en la BUAP, se puede describir el desempeño del docente. Para lo cual no solo se toma en cuenta el ISP sino, también, el tipo

de comentario - ser positivo, negativo, de denuncia o de suplantación (da clase otro profesor)- que los estudiantes hacen de él.

Como se mencionó, con anterioridad, el principal conjunto de datos, que se utilizará para el análisis de la polaridad, es el que contiene cada uno de los comentarios proporcionados por los estudiantes universitarios tras llevarse a cabo la encuesta. El conjunto de datos, de los comentarios utilizado, consta de 201,146 comentarios totales, los cuales fueron etiquetados manualmente por expertos en alguna de las dimensiones comentadas anteriormente (positivo, negativo, denuncia, suplantación). Para los fines de este artículo, el análisis de la polaridad, se emplearon únicamente los comentarios del tipo positivo y negativo, quedando al final un subconjunto de 136,640 comentarios totales, clasificados en 110,556 positivos y 26,084 negativos. Un ejemplo de los mensajes proporcionados por los alumnos y el tipo de mensaje puede observarse en la tabla 1.

Dado los ejemplos de la tabla 1, se puede observar que existen diversas características al momento de la escritura de los comentarios. Por lo cual, es necesaria una etapa de preprocesamiento que permita homogeneizar el comentario y poder ser utilizado.

Tabla 1 Ejemplo del conjunto de datos.

Comentario	Tipo	
	Positivo	Negativo
<i>“excelente profesor tiene un manejo increíble del tema ya que la clase pasaba tan rapido porque es muy bueno en clase”</i>	X	
<i>“en relacion al trato con los alumnos considero que es elitista e inconstante en su estado de animo, lo que nos perjudica por el trato social que nos presenta”</i>		X

Para esta etapa, se llevó a cabo una conversión de todos los comentarios a minúsculas, se realizó una *tokenización* y sobre todo se balanceó el conjunto de datos. Dado que, como se pudo observar, existen más comentarios positivos que negativos, y, por tanto, el corpus está desbalanceado. Esta diferencia, en la cantidad de datos para cada clase, puede tener un impacto negativo en la

generación de resultados. Por lo tanto, para la obtención de mejores resultados, se obtuvieron 3 subconjuntos de datos más pequeños y balanceados: el primero denominado C1000 consiste en 1,000 comentarios (500 negativos y 500 positivos); el segundo se ha llamado C5000 y consiste de 5,000 comentarios (2,500 positivos y 2,500 negativos); finalmente, el tercero se llama C10000 y contiene 10,000 comentarios en total (5,000 positivos y 5,000 negativos). Uno de los objetivos, al tener diferentes corpora, es observar el rendimiento de los clasificadores supervisados a medida que se incrementan el número de muestras sobre las cuales se lleva a cabo el proceso de entrenamiento y prueba.

Descripción de los clasificadores empleados

Las técnicas de aprendizaje automático, o supervisado, son capaces de aprender el proceso humano para clasificar, entre otras cosas, la polaridad en comentarios de estudiantes universitarios. Este proceso requiere de la extracción de características a partir de un corpus que regularmente es anotado de manera manual (corpus supervisado o de entrenamiento). Se genera, de esta manera, un modelo de clasificación que puede posteriormente ser utilizado para clasificar nuevas muestras de las cuales se desconoce su clase final. En este caso, las muestras de entrenamiento y las nuevas muestras se consideran que pueden ser clasificadas automáticamente como solo una de las dos siguientes clases: positivo o negativo.

En los experimentos llevados a cabo, se consideró utilizar el corpus manualmente etiquetado para conformar el corpus de entrenamiento y el corpus de prueba en porcentajes del 80% y 20%, respectivamente. Se usó una técnica conocida como validación cruzada a v capas (*v-fold cross-validation*), la cual usa el corpus total y lo subdivide, tal y como se mencionó anteriormente, en corpus de entrenamiento y prueba en v ocasiones (se utilizó $v=10$ para los experimentos presentados en este artículo). Los valores obtenidos de las 10 ejecuciones fueron promediados y presentados como resultados finales, resultados que se mostrarán en sección 3. Con el fin de tener una perspectiva del tipo de clasificador que puede tratar mejor el problema, de clasificación de la polaridad, se han seleccionado los siguientes

cuatro algoritmos de aprendizaje (cada uno perteneciente a un tipo de clasificador diferente: Bayes, Lazy, Functions y Trees):

- **Naïve Bayes:** es un clasificador probabilístico basado en el teorema de Bayes y algunas hipótesis simplificadoras adicionales [Manning & Schütze, 1999].
- **K-Star:** Este es el clasificador de los k vecinos más cercanos con una función de distancia generalizada [Manning & Schütze, 1999].
- **SMO:** Se trata de un algoritmo de optimización secuencial mínima para la clasificación de vectores soporte [Manning & Schütze, 1999].
- **J48:** Es un algoritmo usado para generar un árbol de decisión [Manning & Schütze, 1999].

Todos los textos fueron representados por medio de un vector de frecuencias de n-gramas, con valores para $n = 1, 2$ y 3 . Las frecuencias superiores a dos, para los n-gramas, son consideradas para el vector de características. Los resultados obtenidos en los experimentos se presentan a continuación.

3. Resultados

Esta sección presenta los resultados obtenidos después de haber utilizado cada uno de los tres corpora de evaluación (C1000, C5000 y C10000). En la tabla 2 se muestran la cantidad de instancias clasificadas correcta e incorrectamente para el corpus C1000, por cada uno de los cuatro clasificadores supervisados utilizados (Naïve Bayes, K-Star, SMO y J48).

Tabla 2 Resultados de la clasificación usando el corpus C1000.

Clasificador	Tipo	Instancias correctas	Instancias incorrectas
Naïve Bayes	Bayes	76.2%	23.8%
K-Star	Lazy	74.5%	25.5%
SMO	Functions	86.7%	13.3%
J48	Trees	79%	21%

En la tabla 3 se muestran la cantidad de instancias clasificadas correcta e incorrectamente para el corpus C5000, por cada uno de los cuatro clasificadores supervisados utilizados (Naïve Bayes, K-Star, SMO y J48). Nuevamente, es el

modelo de clasificación basado en máquinas de vectores de soporte (SMO) el que resulta vencedor con un porcentaje de clasificación del 91.54%, es decir, 4,577 muestras de un total de 5,000 fueron clasificadas correctamente. La tasa de error de este clasificador es del 8.46%, es decir, 423 errores de un total de 5,000 muestras. En particular, 175 muestras que eran positivas fueron clasificadas erróneamente como negativas, y 248 muestras que eran negativas fueron clasificadas incorrectamente como positivas.

Tabla 3 Resultados de la clasificación usando el corpus C5000.

Clasificador	Tipo	Instancias correctas	Instancias incorrectas
Naïve Bayes	Bayes	78.04%	21.96%
K-Star	Lazy	77.84%	22.16%
SMO	Functions	91.54%	8.46%
J48	Trees	85.38%	14.62%

La tabla 4 muestra cantidad de instancias clasificadas correcta e incorrectamente para el corpus C10000, por cada uno de los cuatro clasificadores supervisados utilizados (Naïve Bayes, K-Star, SMO y J48). De nuevo, el modelo de clasificación basado en máquinas de vectores de soporte (SMO) resulta vencedor con un porcentaje de clasificación del 91.95%, es decir, 9,195 muestras de un total de 10,000 fueron clasificadas correctamente. La tasa de error de este clasificador es del 8.05%, es decir, 805 errores de un total de 10,000 muestras. En particular, 337 muestras positivas fueron clasificadas erróneamente como negativas, y 468 muestras que eran negativas fueron clasificadas incorrectamente como positivas.

Tabla 4 Resultados de la clasificación usando el corpus C10000.

Clasificador	Tipo	Instancias correctas	Instancias incorrectas
Naïve Bayes	Bayes	77.26%	22.74%
K-Star	Lazy	80.37%	19.63%
SMO	Functions	91.95%	8.05%
J48	Trees	87.3%	12.7%

4. Discusión

En la figura 1 se puede apreciar el rendimiento de todos y cada uno de los clasificadores evaluados sobre tres diferentes corpora (C1000, C5000 y C10000).

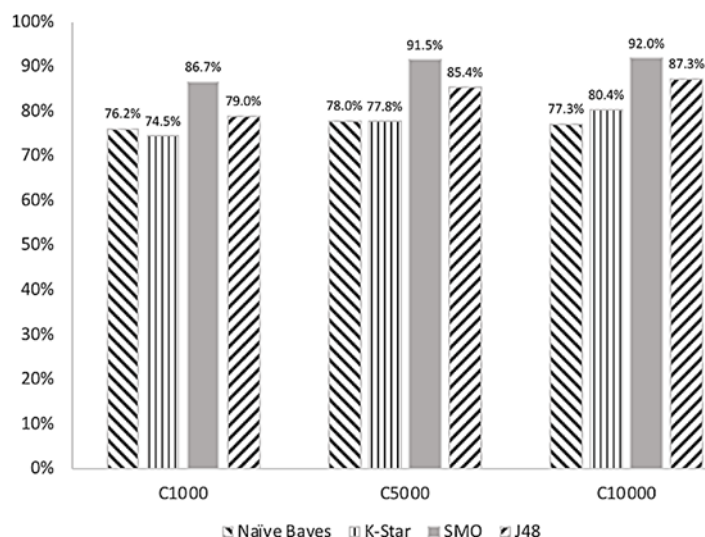


Figura 1 Porcentaje de instancias clasificadas correctamente.

El clasificador K-Star, del tipo Lazy, es el que mayoritariamente presenta el menor rendimiento, principalmente cuando el conjunto de datos es pequeño (1,000 y 5,000), aunque cuando el conjunto de datos es mayor (10,000) el rendimiento alcanzó el 80.4%.

El clasificador de Naïve Bayes exhibe un comportamiento también pobre en comparación con SMO, ya que sus porcentajes de clasificación se encuentran entre el 76% y el 78%.

El clasificador J48, basado en árboles de decisión, es el segundo que mejor se comporta al alcanzar un porcentaje máximo de clasificación del 87.3% cuando el conjunto de datos es de 10,000 muestras. En general, se observa que SMO es el mejor clasificador supervisado, alcanzando un porcentaje cercano al 92%, lo cual se considera muy bueno en el estado del arte. Habría que analizar dicho comportamiento cuando el conjunto de datos es mayor, por ejemplo, duplicando las muestras positivas y negativas. Sin embargo, este análisis será considerado como trabajo a futuro.

En la figura 2 se presenta el porcentaje de instancias clasificadas incorrectamente por cada uno de los clasificadores supervisados (Naïve Bayes, K-Star, SMO y J48) en los tres diferentes corpora evaluados (C1000, C5000 y C10000). Es relevante mencionar que existe una diferencia significativa en la tasa de error obtenida por

el clasificador SMO con respecto a los otros tres clasificadores supervisados (Naïve Bayes, K-Star y J48), la cual llega a ser de hasta 14.6 puntos cuando se usa el corpus C10000.

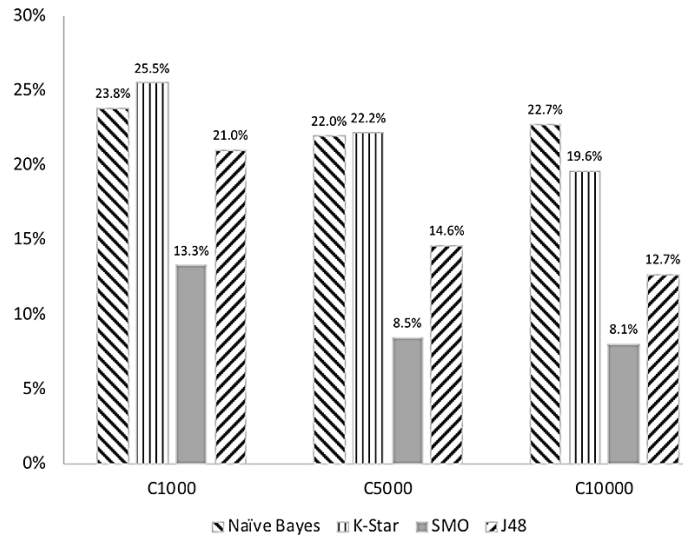


Figura 2 Porcentaje de instancias clasificadas incorrectamente.

Desde esta perspectiva, los resultados son muy buenos dada la cantidad de datos usados. Es evidente que el mejor clasificador ha resultado ser SMO, una implementación de las máquinas de vectores de soporte.

5. Conclusiones

En este artículo se presentó un análisis de la polaridad de comentarios de estudiantes universitarios. Se analizaron cuatro diferentes métodos de clasificación supervisada para determinar el que mejor comportamiento exhibe sobre la tarea planteada. Se usaron tres diferentes corpora para determinar, también, el impacto en el tamaño de las muestras usadas para el proceso de entrenamiento del modelo de clasificación.

Dados los resultados obtenidos en los experimentos, se concluye que el clasificador basado en máquinas de vectores de soporte es el que mejor comportamiento tiene sobre la clasificación de la polaridad de comentarios de estudiantes universitarios.

Las principales aportaciones de este trabajo son:

- Corpora balanceado para llevar a cabo experimentos relacionados con el cálculo de la polaridad de comentarios de estudiantes universitarios.
- Análisis de cuatros clasificadores supervisados para la tarea de clasificación de comentarios positivos y negativos.

Es importante destacar que este artículo hace una contribución importante en la generación de recursos de análisis de textos para el español de México, pues es generalizado el conocimiento de que existe una carencia de este tipo de recursos para este idioma.

Como trabajo futuro, se considera importante experimentar con un mayor conjunto de datos y con otros clasificadores tales como campos aleatorios condicionales (CRF), por citar un ejemplo. Así también, y dados los resultados obtenidos, se cree que es posible generar un módulo computacional que permita llevar a cabo el proceso de clasificación de polaridad (positivo/negativo) de comentarios de estudiantes universitarios.

6. Bibliografía y Referencias

- [1] Amigó, E., Carrillo de Albornoz, J., Chugur, I., Corujo, A., Gonzalo, J., Meij, E., de Rijke, M., & Spina, D. Overview of RepLab 2013: Evaluating Online Reputation Monitoring Systems. P. Forner, H. Müller, R. Paredes, P. Rosso, B. Stein (eds.) CLEF. Lecture Notes in Computer Science. Vol. 8138. 2013, pp 333–352.
- [2] Pang, B., Lee L., & Vaithyanathan, S.: Thumbs up? Sentiment Classification using Machine Learning Techniques. "Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing-Volume 10". Association for Computational Linguistics, 2002, pp 79-86.
- [3] Brooke, J., Tofiloski, M., & Taboada, M.: Cross-Linguistic Sentiment Analysis: From English to Spanish. RANLP. 2009, pp 50-54.
- [4] Carrillo-de-Albornoz, J., Amigó, E., Plaza, L., & Gonzalo, J.: Tweet Stream Summarization for Online Reputation Management. Ferro N. et al. (eds)

- Advances in Information Retrieval. ECIR 2016. Lecture Notes in Computer Science, vol. 9626. Springer, Cham. 2016, pp 378-389.
- [5] Fernández, A., Nuñez, L., Morere, P., & Santos, A.: Sentiment Analysis and Topic Detection of Spanish Tweets: A comparative Study of NLP Techniques. "Procesamiento del Lenguaje Natural, Revista nº 50", marzo de 2013, pp 45-52.
- [6] Fernández, J., Boldrini, Gómez, E. J. M., & Martínez-Barco, P.: Análisis de sentimientos y minería de opiniones: el corpus EmotiBlog. Procesamiento del Lenguaje Natural, Revista nº 47", septiembre de 2011, pp 179-187.
- [7] Hu, M., & Liu, B.: Mining and summarizing customer reviews. Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '04, ACM, New York, NY, USA, 2004, pp 168-177.
- [8] Manning, C., & Schütze, H.: Foundations of statistical natural language processing. MIT Press, Cambridge, USA, 1999.
- [9] Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K.: Wordnet: An on-line lexical database. International Journal of Lexicography 3, 1990, pp. 235-244
- [10] Mohammad, S.M., Bravo-Marquez, F., Salameh M., & Kiritchenko, S.: Semeval-2018 Task 1: Affect in Tweets. "Proceedings of the International Workshop on Semantic Evaluation (SemEval-2018)", New Orleans, LA, USA, June 2018.
- [11] Nasukawa, T., & Yi, J.: Sentiment analysis: Capturing favorability using natural language processing. "Proceedings of the 2Nd International Conference on Knowledge Capture". K-CAP '03, ACM, New York, NY, USA, 2003, pp 70-77.
- [12] Pérez-Rosas, V., Banea, C., & Mihalcea, R.: Learning Sentiment Lexicons in Spanish. LREC. 2012, pp 3077-3081.
- [13] Rosenthal, S., Farra, N., Nakov P.: SemEval-2017 Task 4: Sentiment Analysis in Twitter. Proceedings of the 11th International Workshop on Semantic Evaluation. Vancouver, Canada. August 2017.
- [14] Sidorov, G., Miranda-Jiménez, S., Viveros-Jiménez, F., Gelbukh, A., Castro-Sánchez, N., Velásquez, F., Díaz-Rangel, I., Suárez-Guerra, S., Treviño, A., & Gordon, J.: Empirical Study of Machine Learning Based Approach for Opinion Mining in Tweets. Lecture Notes in Artificial Intelligence LNAI, vol 7629. 2012, pp 1-14.