

# Combinación de clasificadores para el análisis de sentimientos

Montserrat Ramirez García<sup>1</sup>, Maya Carrillo Ruiz<sup>1</sup> y Abraham Sánchez López<sup>1</sup>

Benemérita Universidad Autónoma de Puebla,  
Facultad de Ciencias de la Computación, Puebla, México

{mramirez88, cmaya, asanchez}@cs.buap.mx

**Resumen.** El presente trabajo, propone una arquitectura fusionando los clasificadores: SVM, árboles de decisión y Naive Bayes, mediante mayoría de votos, ventanas y cascada, para explorar el desempeño de la tarea de análisis de sentimientos. Se utilizó un corpus en español de 2625 opiniones, previamente preprocesado. Para representarlo, se emplearon bigramas, bolsa de palabras con pesado tf-idf, etiquetado POS y una representación basada en la teoría de la valoración. Los resultados obtenidos muestran una mejora en medida F del 18.13 %, con respecto a los resultados de los clasificadores base.

**Palabras clave:** combinación de clasificadores, análisis de opinión, ensamble de clasificadores.

## 1. Introducción

La combinación de múltiples clasificadores es considerado un reto importante para lograr la robustez y precisión en diversas tareas, como el análisis de opinión o sentimientos. A pesar de los avances tecnológicos, el análisis de sentimientos sigue siendo un tema de investigación abierto, especialmente en el idioma español. En este trabajo se exploran diversas formas de combinar las decisiones de múltiples clasificadores como una forma viable de mejorar el rendimiento de la tarea de análisis de sentimientos. Al respecto cabe mencionar el trabajo desarrollado por Shoushan Li [1] et al. quienes utilizan máquinas de soporte vectorial y atributos como unigramas, adjetivos, adjetivos+adverbios, nombres, entre otros y mejorar la precisión obtenida por los clasificadores individuales al combinar los tres mejores de los seis utilizados. El presente artículo está organizado de la siguiente manera: en la sección 2 se describe la tarea de análisis de sentimientos, en la sección 3 se describe la tarea de clasificación, así como cada uno de los clasificadores utilizados, posteriormente en la sección 4 se define el concepto de ensamble de clasificadores, y los métodos de combinación empleados. En la sección 5 se describe la arquitectura propuesta, en la sección 6 se muestran los experimentos y resultados obtenidos, en la sección 7 se presentan las conclusiones y finalmente en la sección 8 se presenta el trabajo futuro.

## 2. Análisis de sentimientos

Las opiniones son fundamentales para casi todas las actividades humanas, porque son importantes factores de influencia en nuestros comportamientos. El análisis de sentimientos (AS), también llamado minería de opinión es un campo de estudio que analiza las opiniones, sentimientos, evaluaciones, actitudes de las personas hacia entidades como productos, servicios, organizaciones, individuos, cuestiones, eventos, tópicos y sus atributos.

El problema se define de la siguiente manera:

Dado un conjunto de documentos de texto de evaluación  $D$  que contienen opiniones (o sentimientos) acerca de objetos, se pretende extraer atributos y componentes de objetos que han sido comentados en cada documento  $d$  en  $D$  y determinar si los comentarios son positivos, negativos o neutros [2].

Aquí, opinión es una cuádrupla:

$$(g, s, h, t) \tag{1}$$

Donde  $g$  es la opinión,  $s$  es el sentimiento de la opinión,  $h$  es la persona que expresa la opinión, y  $t$  es el tiempo o la fecha en que se expresa la opinión.

Los indicadores más importantes de sentimientos, son las palabras que expresan sentimiento, llamadas palabras de opinión (opinion words). Estas son palabras que comúnmente son usadas para expresar sentimientos positivos o negativos. Por ejemplo, bueno, maravilloso y estupendo son palabras que expresan sentimiento positivo, en cambio malo, peor y terrible son ejemplos de palabras que expresan sentimiento negativo, a dichas palabras se les conoce comúnmente como lexicón de opiniones (sentiment lexicon o opinion lexicon). A pesar de que las palabras y frases con sentimiento son muy importantes, para el análisis de sentimientos no son suficientes para obtener éxito, la tarea es mucho más compleja, es decir que el lexicón de opiniones es necesario pero no suficiente para el AS. A continuación se describen algunas situaciones que hacen de AS un problema complejo [7].

- Una palabra que expresa un sentimiento negativo o positivo puede tener orientaciones opuestas, según el contexto de la oración.
- Una oración que contiene una palabra considerada como expresión de sentimiento, puede no expresar un sentimiento.
- Oraciones Sarcásticas.
- Opiniones spam

### 2.1. Representaciones textuales

Para analizar las opiniones es necesario representarlas de manera que puedan ser procesadas en una computadora, en este trabajo se utilizaron cuatro representaciones textuales, tres de ellas comúnmente empleadas: bolsa de palabras con pesado tf-idf, n-gramas, etiquetas POS y una representación basada en la teoría de la valoración.

**N-gramas** Es una representación tradicional en recuperación de la información, que consiste de palabras individuales (unigramas), o conjuntos de palabras (n-gramas), con sus frecuencias asociadas. En algunos casos podemos representar mejor un concepto mediante la unión de n palabras que se encuentran adyacentes al término principal, lo que se le conoce como n-gramas. La importancia de esta representación radica en que la posición de las palabras es considerada, puesto que el significado de una palabra, no tiene sentido sin las palabras adyacentes que le acompañan en cualquier texto, por lo que la posición de una palabra afecta potencialmente en el sentido del significado de la oración, es decir el sentimiento o la subjetividad dentro de una unidad textual. Para el trabajo realizado se utiliza n-gramas de tamaño  $n = 2$ , es decir, bigramas.

**Partes de la oración (POS)** Una técnica de representación muy utilizada se basa en las reglas lingüísticas, donde las palabras y frases son categorizadas como sustantivos, verbos, adjetivos y adverbios. De acuerdo con Turney, son características gramaticales que tienen la capacidad de expresar subjetividad [3]. Existen investigaciones enfocadas principalmente en adjetivos y adverbios, como en el trabajo reportado por Farah Benamara et al [4], en donde expone que las expresiones de una opinión dependen principalmente de algunas palabras, por ejemplo, la palabra *bueno* es utilizada comúnmente para una opinión positiva, y la palabra *malo*, para algo negativo, dichas palabras son identificadas como adjetivos en términos lingüísticos.

En general los adjetivos son importantes indicadores en una opinión, son considerados características especiales, sin embargo no significa que otras partes de la oración no contribuyan a la expresión de sentimientos. Existen trabajos en donde los sustantivos, verbos, adverbios y sustantivos subjetivos también han tenido buenos resultados [14].

**TF-IDF (term frequency-inverse document frequency)** Es un esquema de ponderación de términos comúnmente utilizado para representar documentos de texto como vectores, que se ajusta al modelo denominado bolsa de palabras, donde cada documento es representado como serie de palabras sin orden. Se trata de una medida estadística de cuan importante es una palabra para un documento y para un corpus. Dicho esquema, se utiliza frecuentemente en tareas de ordenamiento o reordenamiento de los resultados de búsqueda, generación de resúmenes de texto, agrupación y clasificación de documentos, identificación del autor de algún texto, recomendación de documentos, etc.

**Cálculo del TF** Un término  $t_i$  que aparece muchas veces en un documento  $d_j$  es mas importante que otro que aparece pocas.

$$tf_{ij} = \frac{(n_{ij})}{\sum_{k=1}^N n_{kj}} = \frac{(n_{ij})}{|d_j|} \quad (2)$$

Donde  $n_{ij}$  es el número de ocurrencias del término  $t_i$  en el documento  $d_j$  y  $\sum_{i=1}^N n_{kj}$  es la suma del número de ocurrencias de todos los términos en el documento  $d_j$ , es decir el tamaño del documento  $d_j$ .

### Cálculo del IDF

Un término  $t_j$  que aparece en pocos documentos, discrimina mejor que uno que aparece en muchos.

$$idf_j = \log\left(\frac{N}{n_j}\right) \quad (3)$$

Donde  $N$  es el número total de documentos, y  $n_j$  es el número de documentos que contiene el término  $t_j$ .

### Representación final del documento

Cada elemento queda representado como un vector de características  $d_j$ :

$$d_j = (d_{j1}, \dots, d_{jn}) \quad (4)$$

$$\text{donde, } d_{ij} = t_{ij} * idf_{ij}$$

Es decir finalmente se seleccionan  $n$  términos con los valores más altos en todos los documentos.

**Teoría de la valoración utilizando reglas sintácticas** La teoría de la valoración propuesta por Peter R.R White [5], se ocupa de los recursos lingüísticos por medio de los cuales las personas expresan alguna opinión. Particularmente del lenguaje (expresiones lingüísticas), la valoración, la actitud y la emoción del conjunto de recursos que explícitamente posicionan de manera interpersonal las propuestas y proposiciones textuales. Es decir, trabaja con los significados de las palabras que hacen variar o modificar los términos del compromiso del hablante en sus emisiones, por lo que modifican lo que está en juego en la relación interpersonal.

Dicha representación fue implementada por Morales de Jesús V. M. en su trabajo de tesis [6], que haciendo uso de un diccionario de actitud, el cual contiene elementos taxonómicos de la teoría de la valoración (*juicio, apreciación y afecto*), y utilizando sintagmas adverbiales, busca obtener un significado más preciso de las expresiones de valoración presentes en los textos. El objetivo es contabilizar los valores de positivo, negativo, juicio, apreciación y afecto, que están presentes en una opinión cualquiera, como si se tratase de una bolsa de palabras ponderada, sin embargo las reglas sintácticas, empleadas para identificar los sintagmas adverbiales, juegan un papel primordial en este proceso, ya que dependiendo del tipo de regla, los valores asignados a los elementos de actitud pueden aumentar, disminuir, o intercambiarse, afectando de esa manera los valores finales asignados al sentimiento de cada opinión.

## 3. Clasificación

Los algoritmos de clasificación son métodos que dado un conjunto de ejemplos de entrenamiento infieren un modelo de las categorías en las que se agrupan los

datos, de tal forma que se pueda asignar a nuevos ejemplos una o más categorías de manera automática mediante analogías a los patrones de dicho modelo.

En todo proceso de clasificación supervisada, se cuenta con dos conjuntos de ejemplos etiquetados, uno de entrenamiento y otro de pruebas. Primeramente se utiliza el conjunto de entrenamiento a fin de construir el modelo de clasificación y se verifica haber alcanzado el resultado adecuado según la métrica definida, entonces el proceso termina, sino el proceso de entrenamiento se repite, hasta obtener el resultado deseado.

Para clasificar documentos, primero se representan, empleando representaciones, como las descritas en la sección 2.1. Una vez representados, podrán introducirse al clasificador seleccionado. Recientemente se han desarrollado técnicas para construir conjuntos de clasificadores cuyas decisiones son combinadas (de forma pesada o no) para clasificar nuevos ejemplos. Lo que se ha encontrado es que en general son mejores clasificadores que los clasificadores individuales (o base) que se usaron en su construcción.

La clasificación, puede definirse como la tarea de predecir una variable discreta “ $y$ ” usando un conjunto de características  $x_1, x_2, \dots, x_n$  como variables independientes. Para realizar el entrenamiento del clasificador se necesita una función hipótesis  $h$  de una colección de ejemplos de entrenamiento. Dicha colección tiene la forma  $(X, Y)$  y usualmente se refiere a un conjunto de datos. Cada entrada del conjunto de datos es una tupla  $(x, y)$ , donde  $x$  es el conjunto de características y  $y$  es la clase o etiqueta la cual es una variable discreta con  $c$  posibles categorías. Cuando los resultados posibles son restringidos a valores binarios,  $y_i \in \{+1, -1\}, \forall i \in \{1, \dots, N\}$  [10].

En este trabajo se utilizan tres clasificadores base: Máquina de Soporte Vectorial, Naive Bayes y Árboles de Decisión. Cada uno de estos algoritmos se describe brevemente a continuación:

### 3.1. Naive bayes

Es un clasificador probabilístico que aplica el Teorema de Bayes para estimar la probabilidad posterior  $P(y | x)$  de la clase  $y$  dada la variable  $x$

$$P(y|x) = \frac{P(y|x)P(y)}{P(x)} \quad (5)$$

Naive Bayes se centra en las probabilidades  $P(x|y)$  que se refieren a la verosimilitud y representan la probabilidad de observar el valor  $x$ , dado el valor de clase  $y$ . Debido a esto Naive Bayes es considerado un *clasificador generativo*.

### 3.2. Máquina de soporte vectorial

La máquina de Soporte Vectorial SVM es un clasificador binario discriminante, dirigido a encontrar el hiperplano óptimo ( $w^T * x + b$ ) que separa los dos posibles valores de la variable etiquetada  $y \in \{+1, -1\}$ , de acuerdo al espacio de características representado por  $x$ . El hiperplano óptimo es aquel que maximiza

el margen entre las instancias positivas y negativas en el conjunto de datos de entrenamiento formado por  $N$  observaciones. La tarea de aprendizaje de una SVM se formaliza con el siguiente problema de optimización:

$$\min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \quad (6)$$

$$\text{sujeto a } y_i(w^T x_i + b) \geq 1 - \xi_i \quad \forall i \in \{1, \dots, N\}$$

$$\xi_i \geq 0, \quad \forall i \in \{1, \dots, N\}$$

El objetivo del problema se enfoca en dos aspectos, el primero, obtener el máximo margen en el hiperplano y minimizar el error  $\sum_{i=1}^N \xi_i$ . El parámetro  $C$  se refiere al parámetro suave de regularización de margen y controla la sensibilidad de la SVM para los posibles valores atípicos.

### 3.3. Árboles de decisión

Un árbol de decisión describe un conjunto de reglas organizadas de forma jerárquica, que implementan una estructura de decisión. Se compone de hojas y nodos. Una hoja registra una respuesta (*clase*) y un nodo especifica algunas condiciones de las pruebas que se llevarán a cabo en un valor único, rasgo de una instancia, con una rama y sub-árbol para cada posible resultado de la prueba. Para un determinado vector, se toma una decisión partiendo de la raíz de un árbol, y se recorre el árbol en función del resultado de una prueba de estado en cada nodo, hasta que se encuentra una hoja [11]. El proceso de construcción de un árbol de decisión es una partición recursiva de un conjunto de entrenamiento.

## 4. Ensamble de clasificadores

La idea de un ensamble de clasificadores, es combinar un conjunto de clasificadores para resolver una tarea en conjunto, en donde el objetivo principal es combinar las salidas de los clasificadores base, para generar una salida en donde sean considerados todos los clasificadores y dicha salida sea mejor que la obtenida por cualquier clasificador base, bajo una cierta evaluación [9].

Un ensamble de clasificadores es un grupo de clasificadores quienes individualmente toman decisiones que son fusionadas de alguna manera, para finalmente obtener una decisión por consenso. Los métodos de ensamble son muy efectivos, debido principalmente a que varios tipos de clasificadores tienen sesgos inductivos, y provocan que la diversidad de los clasificadores utilizados reduzca el error de la varianza, sin incrementar el error bias [8].

La combinación de clasificadores y por lo tanto la creación de ensamble de clasificadores fue propuesto para mejorar los resultados obtenidos por los clasificadores base. La llave para producir un ensamble exitoso, es elegir los métodos de clasificación apropiados y seleccionar los clasificadores base indicados para el problema planteado.

Dado el potencial uso del ensamble de clasificadores, existen algunos factores que deben ser diferenciados entre los métodos de ensamble. Los principales factores se listan a continuación:

1. Relación inter-clasificadores. ¿Cómo cada clasificador afecta a otros clasificadores?
2. Método de combinación. La estrategia de combinar los clasificadores generados por un algoritmo de inducción. El combinador simple determina la salida exclusivamente a partir de las salidas de los inductores individuales.
3. Generador de diversidad. Con el objetivo de realizar un ensamble eficiente, debe existir diversidad entre los clasificadores involucrados. La diversidad puede ser obtenida a través de diferentes presentaciones de entrada de datos, como en bagging, variaciones en el diseño de aprendizaje, aplicando una sanción a las salidas para fomentar la diversidad.

Por otra parte, existen diferentes estructuras para combinar los clasificadores, a continuación se describen las utilizadas en el presente trabajo.

#### 4.1. Cascada

Es una arquitectura para combinar clasificadores, que puede presentar  $n$  niveles, sin embargo normalmente presenta dos niveles, en donde el nivel 1 es entrenado con el conjunto de datos original, el nivel 2 con un conjunto de datos aumentado, el cual contiene las características del conjunto de datos original junto con la salida del clasificador del nivel 1. La salida del clasificador del nivel 1 es un vector que contiene la distribución de probabilidad condicional  $(p_1, \dots, p_c)$ , donde  $c$  es el número de clases del conjunto de datos original, y  $p_i$  es la estimación de probabilidad calculada por el clasificador del nivel 1, de que la instancia pertenezca a la clase  $i$ .

El entrenamiento del clasificador del nivel 2 es influenciado por el clasificador del nivel anterior, debido a que considera su salida obtenida, derivando un esquema global sobreentrenado. Sin embargo, en cascada se reduce este problema debido a dos razones: en cada nivel se utiliza un clasificador de diferente naturaleza al otro y además el clasificador del nivel 2 no se entrena únicamente con la salida del clasificador de nivel 1, sino que además tiene en cuenta las características originales.

#### 4.2. Mayoría de votos

Es un método simple de combinación de clasificadores base, en el cual todos los clasificadores incluidos proveen un voto a alguna de las clases, el método realiza la sumatoria de dichos votos y la clase que recibe más votos es seleccionada como la decisión final. Dicho método es representado por la siguiente ecuación:

$$x \rightarrow w \text{ if } w = \arg \max_{w \in \theta} \sum_{i=1}^T 1(C_i(x) = w) \quad (7)$$

$x$  es una instancia,  $\theta$  es el conjunto de etiquetas de clase,  $w$  es la clase asignada para la instancia  $x$  y  $C_1, \dots, C_T$  son los clasificadores base.

### 4.3. Ventanas

El método de *Ventanas* es una técnica general, que tiene por objetivo mejorar la eficiencia de los métodos de aprendizaje o clasificadores utilizados, mediante la identificación de un subconjunto adecuado de instancias de entrenamiento. Dicho método se lleva a cabo mediante el uso de un procedimiento de submuestreo.

El método funciona de la siguiente manera: Se selecciona un subconjunto aleatorio de instancias para el entrenamiento de un clasificador (*una ventana*), el resto de instancias son utilizadas para los datos de prueba, si la precisión obtenida del clasificador es insuficiente, las instancias de prueba clasificadas erróneamente se eliminan de las instancias de prueba y se añaden al conjunto de instancias para el entrenamiento en la siguiente iteración. El proceso continúa hasta que se obtiene una precisión suficiente.

Es importante mencionar que *ventanas* no combina clasificadores, su tarea radica en mejorar el resultado de un clasificador.

Una vez que se han explicado los métodos de clasificación y arquitecturas utilizadas, es importante conocer las métricas que permitirán evaluar el resultado obtenido por los mismos.

### 4.4. Métricas de evaluación

Para realizar la evaluación de los métodos de clasificación aplicados sobre un conjunto de datos, existen métricas, a continuación se describen la utilizadas, donde  $TP$  son las instancias clasificadas correctamente como positivas,  $FP$ , son las instancias clasificadas erróneamente como positivas y de la misma manera para las instancias negativas,  $FN$ , son las instancias clasificadas erróneamente como negativas y  $TN$  son las clasificadas correctamente como negativas. Ahora teniendo las salidas antes descritas los siguientes criterios de evaluación pueden ser utilizados.

**Precisión.** Es la fracción de observaciones clasificadas correctamente como positivas, sobre todas las predicciones clasificadas como positivas.

$$Precision = \frac{TP}{TP + FP} \quad (8)$$

**Recuerdo.** Es la fracción de observaciones clasificadas correctamente como positivas, sobre todas las observaciones positivas.

$$Recuerdo = \frac{TP}{TP + FN} \quad (9)$$

**Medida F.** Es el significado armónico entre precisión y recuerdo



$$MedidaF = \frac{(1 + \beta^2)(2 * Precision * Recuerdo)}{(\beta^2 * Precision) + Recuerdo} \quad (10)$$

Las medidas de evaluación son promediadas por todas las submuestras, asegurando que todas las observaciones fueron usadas para entrenamiento y prueba.

## 5. Arquitectura propuesta

La arquitectura propuesta para el corpus en español, consiste de tres niveles los cuales se describen a continuación.

1. En el primer nivel, se aplica el método de mayoría de votos en dos fases.
  - a) En la primera fase, se realiza mayoría de votos con los clasificadores base: SVM, Naive Bayes y arboles de decisión, para cada representación del corpus.
  - b) En la segunda fase, se aplica mayoría de votos a las mejores salidas de cada representación.
2. El segundo nivel se incluye el método de Cascada. El cual recibe como entrada la salida obtenida por mayoría de votos mejores, obtenido en 1 b). Posteriormente incorpora la entrada al conjunto de datos originales del problema, y realiza la clasificación empleando el clasificador que produzca los mejores resultados en cuanto a medida F de los considerados en el paso 1 a) y 1 b)
3. En el tercer y último nivel, se utiliza el método de ventanas. Dado que los resultados obtenidos por cascada tienen una mejora, el método de ventanas toma como entrada la salida proporcionada por Cascada, y selecciona automáticamente el método de clasificación base mejor en medida F, respecto a representaciones y clasificadores, y define un valor  $N$  para el número máximo de iteraciones. Teniendo estos parámetros, realiza la clasificación  $N$  veces, en cada iteración el algoritmo selecciona las instancias clasificadas de manera errónea y las agrega al conjunto de datos de entrenamiento, intercambiando instancias, hasta que el valor de  $N$  se cumpla. Los resultados obtenidos son los resultados finales de la arquitectura.

## 6. Experimentos y resultados

En esta sección se describen las condiciones de los experimentos realizados y los resultados obtenidos, no sin antes describir el corpus utilizado.

### 6.1. Corpus utilizado

Corpus en español de películas de cine, creado por Fermín L. Cruz, et al, como se describe en [12], con 3878 críticas que contiene una puntuación asignada del 1 al 5 donde 1 es la más negativa y 5 es la más positiva, del cual se tomaron 2625 críticas (1351 positivas, 1274 negativas) no incluyendo las críticas neutras, es decir con puntuación 3.

## **6.2. Aplicaciones desarrolladas**

1. Se realizó una aplicación en Microsoft Visual Studio 2012, para implementar el módulo de representación de textos.
2. Se construyó también una aplicación en Matlab 2014b, para la implementación del sistema de clasificación.

En la tabla 1, se muestra la longitud del vocabulario del corpus empleado, después de realizar el preprocesamiento de los textos.

**Tabla 1.** Cardinalidad del vocabulario.

Corpus	Vocabulario completo	Vocabulario Truncado
Español	57713	19876 (34.43 %)

## **6.3. Preprocesamiento de los datos**

Una vez elegido el corpus y antes de realizar el Análisis de Sentimientos, primero se realizó un preprocesamiento de los datos, ya que el corpus fue construido a partir de opiniones introducidas por usuarios comunes de la web y no por críticos especializados. Se eliminaron: palabras vacías, símbolos no alfanuméricos, números y signos de puntuación.

## **6.4. Condiciones de ejecución**

Las representaciones utilizadas, fueron bolsa de palabras, bigramas, etiquetas POS considerando adjetivos y adverbios, y la representación de la teoría de la valoración.

Los resultados reportados son el promedio de 10 ejecuciones, aplicando validación cruzada a 10 pliegues, el mecanismo de selección para los diferentes ejemplos, fue aleatorio y sin remplazo, considerando el 50 % de instancias con clase positiva y el otro 50 % de clase negativa, tanto para el entrenamiento como para la prueba, es decir con corpus balanceados.

Los experimentos realizados fueron variando el porcentaje de datos de entrenamiento y de prueba, con 80 % - 20 % y 60 %- 40 % respectivamente. Empleando los clasificadores mencionados previamente: SVM, arboles de decisión y Naive Bayes.

Para el clasificador SVM, se utilizó un kernel lineal.

Tabla 2. Resultados de la arquitectura propuesta

Corpus	Representación	Clasificador	Precisión	Recuerdo	Medida F	
		Mayoría de votos Mejores	0.7323	0.7028	0.7172	
	Bigramas	SVM(Nivel 1)	0.7157	0.7761	0.7447	
		Cascada(Nivel 2)	<b>0.8346</b>	0.8360	<b>0.8353</b>	
		Ventanas(Nivel 3)	0.8091	<b>0.8385</b>	0.8235	
60 %-40 %	Bolsa de palabras	Naive Bayes(Nivel 1)	0.6057	0.5468	0.5748	
		Cascada(Nivel 2)	0.7647	<b>0.7860</b>	<b>0.7752</b>	
		Ventanas(Nivel 3)	<b>0.9685</b>	0.6178	0.7544	
	POS	Árboles(Nivel 1)	0.5758	0.5431	0.5590	
		Cascada(Nivel 2)	0.6352	<b>0.6581</b>	<b>0.6464</b>	
		Ventanas(Nivel 3)	<b>0.8946</b>	0.4773	0.6225	
	Valoración	SVM(Nivel 1)	0.6932	0.5596	0.6193	
		Cascada(Nivel 2)	0.7575	<b>0.7506</b>	0.7541	
		Ventanas(Nivel 3)	<b>0.8632</b>	0.7047	<b>0.7760</b>	
			Mayoría de votos Mejores	0.9159	0.8920	0.9038
		Bigramas	SVM(Nivel 1)	0.8241	0.8213	0.8227
			Cascada(Nivel 2)	0.9252	0.9340	0.9296
	Ventanas(Nivel 3)		<b>0.9788</b>	<b>0.9652</b>	<b>0.9719</b>	
80 %-20 %	Bolsa de palabras	Naive Bayes(Nivel 1)	0.6555	0.5361	0.5898	
		Cascada(Nivel 2)	0.9240	<b>1</b>	<b>0.9592</b>	
		Ventanas(Nivel 3)	<b>0.9526</b>	0.9198	0.9359	
	POS	Naive Bayes(Nivel 1)	0.5799	0.6735	0.6232	
		Cascada(Nivel 2)	0.9204	0.6198	0.5750	
		Ventanas(Nivel 3)	<b>0.9221</b>	<b>0.9457</b>	<b>0.9338</b>	
	Valoración	SVM(Nivel 1)	0.7500	0.5979	0.6654	
		Cascada(Nivel 2)	<b>0.9828</b>	<b>0.9385</b>	<b>0.9602</b>	
		Ventanas(Nivel 3)	<b>0.9828</b>	<b>0.9385</b>	<b>0.9602</b>	

## **6.5. Experimentos**

En la tabla 2 se muestran los resultados obtenidos, para cada nivel de la arquitectura propuesta. La salida del primer nivel, son los resultados obtenidos por el clasificador con mejor medida F, mismos que entran al segundo nivel. Por tal motivo el clasificador base de entrada al segundo nivel varía.

Mayoría de votos mejores, se refiere a la aplicación del método de mayoría de votos, a los mejores resultados obtenidos por SVM, Naive Bayes y arboles de decisión empleando las representaciones: tf-idf, bigramas, POS y teoría de la valoración.

Como puede observarse en la tabla 2 en la partición 60-40 en tres casos se alcanza la mejor medida F en el segundo nivel de la arquitectura, siendo 0.8353 la más alta obtenida con bigramas, seguida por teoría de la valoración con 0.7760, digno de recalcarse pues es una representación que emplea vectores de dimensión 8.

En la partición 80-20, la mejor medida F igual a 0.9719, se obtiene también con bigramas en el tercer nivel de la arquitectura, lo que representa una mejora de 18.13 % con respecto a los resultados del mejor clasificador base SVM con medida F de 0.8227. La clasificación empleando bolsa de palabras y teoría de la valoración también superan al mejor clasificador base desde el segundo nivel de la arquitectura y POS en el tercer nivel.

## **7. Conclusiones**

Existen trabajos limitados de análisis de sentimientos para textos en español, en los que es notorio que la medida F alcanzada es mucho más baja que los métodos reportados para el idioma inglés.

Se han realizado varios experimentos con distintos métodos de clasificación y distintas formas de representación de los datos, los resultados obtenidos han sido muy diversos. Se ha podido distinguir que las características utilizadas han sido un factor determinante para obtener resultados satisfactorios, la selección de características es un aspecto muy importante, puesto que el éxito de la clasificación depende de tomar las características o los atributos que mejor representen a los documentos, ya que de esto dependerá expresar la polaridad correcta de los documentos y obtener resultados satisfactorios.

Los resultados obtenidos con la arquitectura propuesta, son superiores en precisión, recuerdo y medida F, con respecto a los obtenidos por los clasificadores base.

Bigramas resultó ser la representación de textos que permitió obtener mejores resultados, en la arquitectura.

SVM es el clasificador base que mostró mejor desempeño en la arquitectura propuesta.

La arquitectura propuesta obtuvo una mejora del 12.52 %, con respecto a un trabajo publicado en 2011, que utiliza el mismo corpus [13].

## 8. Trabajo futuro

Se tiene considerado para trabajo futuro, realizar un análisis cualitativo de las instancias clasificadas erróneamente, con el objetivo de identificar similitudes y características de las instancias, para estudiarlas y posteriormente, proponer algún método que contribuya a clasificar correctamente las instancias mal clasificadas.

También se tiene planeado aplicar la arquitectura propuesta, en otras áreas de conocimiento, con el fin de observar el comportamiento de la misma.

## Referencias

1. Shoushan L., Chengqing Z., Xia W.: Sentiment classification through combining classifiers with multiple feature sets. In: Natural Language Processing and Knowledge Engineering, NLP-KE 2007. International Conference IEEE, pp. 135–140 (2007)
2. Bo Pang, Lillian Lee: Opinion Mining and Sentiment Analysis. Found. Trends Inf. Retr., Vol 2. Now Publishers Inc., Hanover, MA, USA (2008)
3. Peter D. T.: Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL '02), Association for Computational Linguistics, Stroudsburg, PA, USA, pp. 417–424 (2002)
4. Benamara F., Cesarano C., Picariello A., Reforgiato D., V.S. Subrahmanian: Sentiment analysis: Adjectives and adverbs are better than adjectives alone. In: Proceedings of the International Conference on Weblogs and Social Media (ICWSM) (2007)
5. Peter R. R. White, J. R. Martin: The Language of Evaluation: Appraisal in English. Palgrave Macmillan, London/New York (2005)
6. Morales de Jesús, V. M.: Utilización de expresiones de actitud para el Análisis de Sentimientos. Tesis de Licenciatura, Benemérita Universidad Autónoma de Puebla, Puebla, México (2014)
7. Liu, B.: Sentiment Analysis and Opinion Mining. Vol. 5, Morgan & Claypool Publishers (2012)
8. Kai Ming T., Ian H. W.: Stacked Generalization: when does it work? (Working paper 97/03), Hamilton, New Zealand: University of Waikato, Department of Computer Science, pp. 866–871 (1997)
9. Kagan T., Joydeep G.: Linear and Order Statistics Combiners for Pattern Classification. In: Combining Artificial Neural Networks, Ed. Amanda Sharkey, Springer Verlag, pp. 127–161 (1999)
10. Witten, I. H., Mark A.: Data Mining: Practical Machine Learning Tools and Techniques. (3rd ed.) Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (2011)
11. Duda, R.O., Hart, P.E., Stork D.G.: Pattern Classification. New York: John Wiley & Son (2001)
12. Fermín L. C., José A. T., Fernando E., F. Javier O.: Clasificación de documentos basada en la opinión: experimentos con un corpus de críticas de cine en español. Procesamiento del lenguaje natural, vol. 41, pp. 73–80 (2008)
13. Eugenio M. C., Martín V.: Opinion classification techniques applied to a Spanish corpus. In: Natural Language Processing and Information Systems, Vol. 6716, Springer Berlin/Heidelberg, pp. 169–176 (2011)

*Montserrat Ramírez García, Maya Carrillo Ruiz y Abraham Sánchez López*

14. Bo P., Lillian L.: Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, Vol. 2, Computer Science Department, Cornell University, Ithaca, NY, U.S.A., No. 1-2, pp. 1–135 (2008)