



**Universidad del Bío-Bío**  
**Facultad de Ciencias Empresariales**  
Departamento de Sistemas Información

# **Comparación de rendimiento de técnicas de aprendizaje automático para análisis de afecto sobre textos en español**

Tesis presentada por Jorge Andrés Elgueta Morales  
para optar al título de Magister en Ciencias de la Computación  
Dirigida por Dr. Christian Vidal Castro y Dra. Alejandra Segura Navarrete

Concepción, Chile – 2017

## **RESUMEN**

Analizar la subjetividad en textos disponibles puede entregar diversos beneficios para las organizaciones y personas. En particular, el Análisis de Afecto intenta descubrir el afecto que le imprime el escritor a sus textos. El análisis automatizado de textos se convierte en un problema de clasificación que puede ser resuelto con técnicas de aprendizaje automático. Sin embargo, la mayor parte de estos análisis han sido realizados para corpus de textos escritos en inglés. Este trabajo comparó la aplicación de técnicas de aprendizaje automático y de NLP sobre textos en idioma español, realizando un proceso de análisis de afecto mediante la aplicación de 3 clasificadores basados en técnicas de aprendizaje automático y estudiando la incidencia de técnicas de NLP, como stopwords y etiquetado POS. Para ello, se configuró un experimento sobre un corpus en español elaborado con titulares de periódicos online de Chile. Los mejores resultados de rendimiento del clasificador se obtienen con la técnica Support Vector Machine sobre un corpus.

Palabras clave – Análisis de afecto, Aprendizaje automático, Rendimiento de clasificadores supervisados

## **ABSTRACT**

Analyze texts available subjectivity can deliver many benefits for organizations and individuals. In particular, the Affect analysis tries to discover the affection that adds the writer to his texts. The automated text analysis becomes a classification problem can be solved with machine learning techniques. However, most of these analyzes have been conducted to corpus of texts written in English. This study compared the application of machine learning techniques and NLP on texts in Spanish language by applying an affect analysis process with 3 classifiers based on machine learning techniques and studying the incidence of NLP techniques, as stopwords and POS tagging. To do this, an experiment on a corpus made headlines in Spanish online newspapers in Chile was configured. The best results are obtained classifier performance with Support Vector Machine technique on a corpus.

Key words - Affect Analysis, Machine learning, Performance of supervised classifiers

# ÍNDICE DE CONTENIDOS

<b>1.- INTRODUCCIÓN .....</b>	<b>7</b>
<b>2.- HIPÓTESIS, OBJETIVOS Y LIMITACIONES DE LA INVESTIGACIÓN .....</b>	<b>9</b>
2.1.- Hipótesis.....	9
2.2.- Objetivo General.....	9
2.3.- Objetivos Específicos .....	9
2.4.- Limitaciones de la investigación.....	9
<b>3.- ESTADO DEL ARTE.....</b>	<b>10</b>
3.1.- Conceptos preliminares .....	10
3.1.1.- Procesamiento del Lenguaje Natural.....	10
3.1.2.- Análisis de subjetividad en Textos .....	10
3.1.3.- Enfoques para análisis de subjetividad.....	12
3.2 Proceso de Análisis de Afecto .....	13
3.2.1 Categorías Afectiva .....	13
3.2.2 Preparación de los datos: limpieza del corpus.....	13
3.2.3 Selección de Features: Etiquetado POS .....	14
3.2.4 Algoritmos de clasificación Automática.....	15
3.2.5 Métodos de evaluación de clasificadores supervisados .....	16
3.2.6 Métricas.....	17
3.3 Herramientas utilizadas en análisis de texto .....	19
3.3.1 Gate .....	20
3.3.2 Stanford coreNLP.....	20
3.3.3 WEKA .....	21
3.4 Trabajos Relacionados.....	22
<b>4.- METODOLOGÍA .....</b>	<b>23</b>
4.1 Selección de la categoría Afectiva.....	24
4.2 Corpus de Trabajo .....	25
4.3 Experimento.....	28
4.3.1 Programa en JAVA .....	30
4.3.2 Clasificación Automática en WEKA.....	33

<b>5.- RESULTADOS PRELIMINARES Y DISCUSIÓN .....</b>	<b>35</b>
<b>6.- CONCLUSIONES Y TRABAJOS FUTUROS .....</b>	<b>38</b>
<b>REFERENCIAS .....</b>	<b>41</b>
<b>ANEXOS .....</b>	<b>43</b>

## ÍNDICE DE FIGURAS

<b>Figura 1:</b> Ejemplo 10-Fold Cross-Validation .....	17
<b>Figura 2:</b> Ejemplo de estructura de archivo ARFF - Datos relacionados a "Titulares de periódico" ....	21
<b>Figura 3:</b> Metodología de trabajo .....	24
<b>Figura 4:</b> Nube de palabras del corpus final .....	27
<b>Figura 5:</b> Balance del Corpus .....	27
<b>Figura 6:</b> Proceso general de experimentación .....	28
<b>Figura 7:</b> Método para eliminar comillas y espacios dobles o superior .....	30
<b>Figura 8:</b> Método para reemplazo de caracteres especiales .....	31
<b>Figura 9:</b> Método para eliminar signos de puntuación.....	31
<b>Figura 10:</b> Método para eliminar signos de exclamación e interrogación.....	31
<b>Figura 11:</b> Método para eliminar StopWords .....	32
<b>Figura 12:</b> Ejemplo de funcionamiento del Stanford POS Tagger.....	32
<b>Figura 13:</b> Ejemplo de resumen de clasificación provisto por WEKA .....	33
<b>Figura 14:</b> Proceso de clasificación automática para las 4 versiones del corpus.....	34
<b>Figura 15:</b> Anexo Resultados Clasificación WEKA - Corpus 1 Clasificador DT.....	54
<b>Figura 16:</b> Anexo Resultados Clasificación WEKA - Corpus 1 Clasificador NB .....	54
<b>Figura 17:</b> Anexo Resultados Clasificación WEKA - Corpus 1 Clasificador SVM .....	55
<b>Figura 18:</b> Anexo Resultados Clasificación WEKA - Corpus 2 Clasificador DT.....	55
<b>Figura 19:</b> Anexo Resultados Clasificación WEKA - Corpus 2 Clasificador NB .....	56
<b>Figura 20:</b> Anexo Resultados Clasificación WEKA - Corpus 2 Clasificador SVM .....	56
<b>Figura 21:</b> Anexo Resultados Clasificación WEKA - Corpus 3 Clasificador DT.....	57
<b>Figura 22:</b> Anexo Resultados Clasificación WEKA - Corpus 3 Clasificador NB .....	57
<b>Figura 23:</b> Anexo Resultados Clasificación WEKA - Corpus 3 Clasificador SVM .....	58
<b>Figura 24:</b> Anexo Resultados Clasificación WEKA - Corpus 4 Clasificador DT.....	58
<b>Figura 25:</b> Anexo Resultados Clasificación WEKA - Corpus 4 Clasificador NB .....	59
<b>Figura 26:</b> Anexo Resultados Clasificación WEKA - Corpus 4 Clasificador SVM .....	59

## ÍNDICE DE TABLAS

<b>Tabla 1:</b> Evaluación de los clasificadores - evaluación de textos .....	18
<b>Tabla 2:</b> Evaluación de los clasificadores – Ejemplo de clasificación .....	19
<b>Tabla 3:</b> Evaluación de los clasificadores - Ejemplo de métricas .....	19
<b>Tabla 4:</b> Módulos disponibles para los lenguajes soportados CoreNLP – fuente: [22] .....	21
<b>Tabla 5:</b> Ejemplo de titulares y etiquetado previo de los evaluadores .....	25
<b>Tabla 6:</b> Ejemplo de titulares y decisión de mantener o eliminar titular .....	26
<b>Tabla 7:</b> Extracto del corpus resultante.....	26
<b>Tabla 8:</b> Versiones del corpus para el experimento .....	29
<b>Tabla 9:</b> Ejemplo de las versiones del corpus .....	29
<b>Tabla 10:</b> Rendimiento de los clasificadores .....	35
<b>Tabla 11:</b> Resultados obtenidos para la métrica Accuracy.....	36
<b>Tabla 12:</b> Resultados obtenidos para la métrica Precision .....	36
<b>Tabla 13:</b> Resultados obtenidos para la métrica Recall.....	36
<b>Tabla 14:</b> Resultados obtenidos para la métrica F-Measure.....	36
<b>Tabla 15:</b> Anexo Matriz Confusión Kappa Ev1 vs Ev2.....	43
<b>Tabla 16:</b> Anexo Cálculo de Kappa entre Ev1 y Ev2 .....	43
<b>Tabla 17:</b> Anexo Matriz Confusión Kappa Ev1 vs Ev3.....	43
<b>Tabla 18:</b> Anexo Cálculo de Kappa entre Ev1 y Ev3 .....	43
<b>Tabla 19:</b> Anexo Matriz Confusión Kappa Ev2 vs Ev3.....	44
<b>Tabla 20:</b> Anexo Cálculo de Kappa entre Ev2 y Ev3 .....	44
<b>Tabla 21:</b> Anexo Resumen de etiquetas EAGLES.....	45
<b>Tabla 22:</b> Anexo Detalle de adjetivos etiquetas EAGLES.....	46
<b>Tabla 23:</b> Anexo Detalle de adverbios etiquetas EAGLES.....	46
<b>Tabla 24:</b> Anexo Detalle de determinantes etiquetas EAGLES .....	47
<b>Tabla 25:</b> Anexo Detalle de nombre etiquetas EAGLES .....	47
<b>Tabla 26:</b> Anexo Detalle de verbos etiquetas EAGLES .....	48
<b>Tabla 27:</b> Anexo Detalle de pronombres etiquetas EAGLES .....	49
<b>Tabla 28:</b> Anexo Detalle de conjunciones etiquetas EAGLES .....	49
<b>Tabla 29:</b> Anexo Detalle de interjecciones etiquetas EAGLES .....	50
<b>Tabla 30:</b> Anexo Detalle de preposiciones etiquetas EAGLES .....	50
<b>Tabla 31:</b> Anexo Detalle de puntuación etiquetas EAGLES .....	50
<b>Tabla 32:</b> Anexo Detalle de numeración etiquetas EAGLES .....	50
<b>Tabla 33:</b> Anexo Detalle de fecha y hora etiquetas EAGLES .....	51

## 1.- INTRODUCCIÓN

Los sentimientos o emociones son un subtipo de actitudes definidas como relativamente duraderas, creencias afectivas, preferencias y predisposiciones hacia objetos o personas [1]. Cuando se escribe una opinión es posible extraer los sentimientos o afectos que el autor impregnó en el texto escrito. Desde el punto de vista computacional, el procesamiento de lenguaje natural (NLP) es una de las técnicas utilizada para identificar estos sentimientos o emociones. A esta área la llamaremos análisis de subjetividad y se puede dividir en Análisis de Sentimiento (SA) y Análisis de Afecto (AA).

El análisis de subjetividad intenta detectar sentimientos o emociones, determinar quién las posee (holder), cuál es el objeto que genera la emoción (target) y cuál es el tipo de sentimiento asociado (me gusta, me encanta, lo valoro, lo odio) o su polaridad (positiva, negativa, neutra) [2]. El SA se encarga de identificar opiniones, emociones y valoraciones, utilizando herramientas computacionales para formalizar y polarizar este contenido [3]. Esto quiere decir que las sentencias contenidas en un texto podrán tener una orientación tanto positiva como negativa (o neutra en algunos casos). Por otro lado, AA incluye un conjunto de técnicas del procesamiento de lenguaje natural para estimar el aspecto afectivo de un texto [4]. Para ello, el AA clasifica las emociones identificadas en taxonomías propuestas por expertos del área de la psicología, las cuales incluyen emociones como alegría, confianza, miedo, tristeza, entre otras.

Para implementar las técnicas SA y AA existen dos enfoques. El primero, utiliza recursos léxicos como diccionarios o lexicones, que contienen palabras etiquetadas con su polaridad (positiva, negativa o neutra) o con la clase afectiva que representa [5]. El segundo enfoque utiliza un corpus de textos etiquetados que permite la construcción de un clasificador que se entrena para esta labor. En este enfoque, suelen utilizarse técnicas provenientes del aprendizaje automático y la estadística [6].

A raíz del enfoque del aprendizaje automático, el proceso de AA define una serie de pasos a realizar para cumplir con su objetivo de clasificación. Las técnicas de aprendizaje automático utilizadas en la realización de AA y que sirven para el entrenamiento del clasificador son principalmente del tipo supervisadas. Esto se logra mediante algoritmos que aprenden de un conjunto previamente etiquetado con la ayuda de atributos considerados importantes para la clasificación [7]. Este enfoque ha sido utilizado principalmente en textos en inglés.

Por otra parte, muchos periódicos tienen sus ediciones en forma digital y permiten desplegar las noticias en una forma rápida y económica. En forma intuitiva, existe la creencia de que los titulares informan principalmente noticias asociadas a afectos negativos como tristeza y terror. De esta forma, los titulares se convierten en un importante objeto de estudio desde el punto de vista del AA.

El análisis de afectos aplicado a textos extraídos de periódicos puede ofrecer una percepción global de la aceptación o rechazo de acontecer nacional, productos, servicios, opiniones, entre otras cosas. La información obtenida de un análisis eficiente o suficientemente confiable puede ser usada para la toma de decisiones importantes tanto por individuos como por organizaciones. La gran mayoría de trabajos dedicados a realizar estos análisis se centran en el análisis de sentimientos y son realizados principalmente en inglés, por lo que parece interesante y necesario conocer el comportamiento de las técnicas de aprendizaje automático y de tratamiento de textos en español, específicamente en la clasificación en alguna categoría de afecto para obtener un mayor grado de especificidad en las emociones que poseen estos textos.

Se requiere saber qué técnicas obtienen mejor rendimiento sobre textos en español. Para ello, es necesario conocer cómo influyen en los resultados de la clasificación los textos procesados con técnicas como stopwords y etiquetado POS. Este artículo intenta obtener información acerca de la aplicación de técnicas NLP y aprendizaje automático en textos en español, mediante la comparación de resultados de aplicar 3 clasificadores independientes sobre un corpus con y sin etiquetas POS.

El resto del documento se organiza de la siguiente manera: En la sección 2 se presentan la hipótesis, objetivos y alcances de la investigación. Luego, en la sección 3 se presenta el trabajo relacionado base para enfrentar la investigación. En la sección 4 se determina la metodología de trabajo y se explica el proceso de experimentación que se llevó a cabo, para que en la sección 5 se expongan los resultados preliminares y discusión. Finalmente, en la sección 6 se exponen las conclusiones y trabajos futuros.



## **2.- HIPÓTESIS, OBJETIVOS Y LIMITACIONES DE LA INVESTIGACIÓN**

En esta sección, se presenta la hipótesis de la tesis, objetivo general y objetivos específicos, además de los alcances de la investigación.

### **2.1.- Hipótesis**

La hipótesis de esta tesis es: La aplicación de diferentes técnicas de aprendizaje automático obtiene diferentes niveles de rendimiento cuando son aplicadas en un mismo proceso de Análisis de Afectos sobre un mismo corpus.

### **2.2.- Objetivo General**

El objetivo general de esta tesis es comparar técnicas de aprendizaje automático utilizadas en análisis de afecto en textos, aplicadas en un caso práctico.

### **2.3.- Objetivos Específicos**

Los objetivos específicos de esta tesis son:

- Estudiar las técnicas de aprendizaje automático más utilizadas en los procesos de análisis de afecto.
- Aplicar un proceso de análisis de afecto sobre un corpus, tomando en cuenta las técnicas de aprendizaje automático más utilizadas estudiando la incidencia de técnicas de NLP, como stopwords y etiquetado POS.
- Realizar un análisis comparativo de las técnicas de aprendizaje automático de acuerdo a métricas establecidas de rendimiento.

### **2.4.- Limitaciones de la investigación**

Los límites de trabajo que tiene esta tesis son los siguientes:

- En este trabajo no se propondrá una nueva clasificación de las emociones, y se utilizará la clasificación de 8 categorías propuesta por Plutchik.
- En este trabajo no se propondrá un nuevo clasificador y solo se analizarán los existentes en la literatura y que son más utilizados.
- En este trabajo no se propondrá una nueva métrica para analizar el rendimiento de los clasificadores y se trabajará con las más importantes existentes según la literatura.
- En este trabajo no se aplicarán técnicas de tratamiento de negación e ironía. Sólo se utilizarán las técnicas de limpieza más adecuadas para el corpus.

### **3.- ESTADO DEL ARTE**

Este capítulo comprende la parte teórica que será base para el desarrollo de esta investigación. Inicia con el estado del arte que abarca desde algunas definiciones necesarias, pasando por el procesamiento de lenguaje natural y técnicas de procesamiento hasta llegar a la clasificación automática y métricas de evaluación de clasificadores. Se pretende incluir contenidos que son relevantes para el desarrollo de esta investigación.

#### **3.1.- Conceptos preliminares**

En esta sección, se introducen algunos conceptos relevantes que serán utilizados para caracterizar el proceso de AA.

- Documento: Se llama documento a cada unidad de texto que conforma la colección de datos. En sistemas de AA, un documento es un texto no estructurado compuesto por un conjunto de sentencias o secuencia de palabras que expresan opiniones y emociones.
- Corpus: El corpus de datos está compuesto por el conjunto de documentos que se utilizan como entrada para entrenar el sistema de AA (conjunto de datos de entrenamiento) y por el conjunto de documentos que serán clasificados utilizando el sistema de AA (conjunto de datos de prueba).
- Features: En esta investigación se utilizará el término feature para referirnos a características o atributos de los documentos que nos permitan encontrar patrones para predecir las características que clasifican a las emociones de las opiniones expresadas en un texto.

##### **3.1.1.- Procesamiento del Lenguaje Natural**

El procesamiento de lenguaje natural (NLP) es un área de investigación y de aplicación que explora cómo las computadoras pueden ser utilizadas para comprender y manipular en texto las expresiones del ser humano [8]. Los fundamentos del NLP se encuentran en una serie de disciplinas, como lo son la computación y ciencias de la información, la lingüística, matemáticas, inteligencia artificial, psicología, entre otras. La aplicación del NLP está enfocada a diversos campos de estudio, tales como la traducción automática, el procesamiento del lenguaje natural en textos, interfaces de usuario, multilingüe y la información de lenguajes de recuperación (CLIR), reconocimiento de voz, entre otros. Se han desarrollado muchos métodos de detección de emociones y su posterior clasificación. Estos métodos pueden procesar y analizar la expresión facial [9][10], la postura corporal [11], el discurso verbal [12][13] y el contenido de los textos.

##### **3.1.2.- Análisis de subjetividad en Textos**

Parte del lenguaje natural de los seres humanos incluye la transmisión de sentimientos, los cuales son un subtipo de actitudes definidas como: relativamente duraderas, creencias afectivas, preferencias y predisposiciones hacia objetos o personas [1]. Dentro del área del NLP en textos, uno de los enfoques llamado análisis de subjetividad, está orientado a identificar sentimientos o emociones y clasificarlos

dentro de categorías, con el fin de obtener información relevante sobre la orientación afectiva del autor. Este enfoque abarca 2 técnicas que son las más utilizadas de acuerdo a la literatura y que son en análisis de sentimientos y el análisis de afectos.

### **3.1.2.1.- Análisis de Sentimientos**

El Análisis de Sentimientos (AS) se encarga de identificar opiniones, sentimientos y valoraciones, utilizando herramientas computacionales para formalizar y polarizar este contenido [3]. Esto quiere decir que las sentencias contenidas en un texto podrán tener una orientación tanto positiva como negativa (o neutra en algunos casos).

Ejemplo: Se tienen las siguientes expresiones y su correspondiente sentimiento basado en polaridad.

- Neutro: La serie “Los 80” tiene 67 capítulos
- Positivo: La serie “Los 80” es muy buena.
- Negativo: La serie “Los 80” es muy larga, me aburre.

Existen distintas tareas que pueden realizarse en sistemas de AS: la tarea más simple es la clasificación binaria de la actitud de un texto en positiva, negativa o neutra; mientras que la tarea más avanzada es la identificación de aspectos dentro de un texto y su correspondiente sentimiento.

### **3.1.2.2.- Análisis de Afectos**

El Análisis de Afecto (AA) busca estimar el aspecto afectivo de un texto [4]. Para ello, el AA clasifica las emociones identificadas en taxonomías propuestas por expertos del área de la psicología, las cuales incluyen emociones como alegría, confianza, miedo, tristeza, entre otras.

Ejemplo: Se tienen las siguientes expresiones y su correspondiente afecto.

- Alegría: “Me he sacado buena calificación en la prueba. Estoy feliz que mis esfuerzos rindan frutos”.
- Confianza: “Que bueno que la policía haya llegado. Ahora podré dormir tranquilo”.
- Interés: “La sonda que recorrió Marte encontró agua. Existe posibilidad de encontrar vida en el planeta Rojo”.
- Disgusto: “Es hora de terminar con esto. No voy a ninguna parte. No estoy haciendo nada”.
- Miedo: “Él podría tener un cáncer de pulmón. Es solo un rumor, pero tiene sentido. No quiero que nada malo le ocurra”.
- Tristeza: “Una chica de mi antigua escuela fue atropellada y murió. Es una lamentable noticia”.

Si se estuviese utilizando análisis de sentimiento, las 2 primeras expresiones corresponderían a una expresión positiva, la tercera expresión correspondería a una expresión neutra y las 3 últimas corresponderían a una expresión negativa.

En el AA, a diferencia del AS, la identificación de la emoción se realiza con un nivel mayor de profundidad, por lo que se busca ser más específico en identificar la intención expresiva de la persona.

### **3.1.3.- Enfoques para análisis de subjetividad**

A la hora de realizar procesos de AS o AA, existen 2 enfoques que permiten llevar a cabo las tareas de identificación de sentimientos o emociones.

#### **3.1.3.1 Enfoque semántico o basado en lexicones**

El enfoque semántico se caracteriza por emplear un lexicón, que es un conjunto de palabras previamente polarizadas [3], es decir, donde cada palabra está etiquetada de acuerdo a su orientación semántica (orientación de una opinión según los adjetivos o adverbios que contiene). Existen lexicones dedicados y diferenciados tanto para realizar tareas de AS como de AA, y estos a su vez están clasificados en 2 categorías:

- Lexicón basado en diccionario: Contiene un listado de palabras con una polaridad fija.
- Lexicón basado en corpus: Contiene un listado de palabras con una polaridad variable. Esta variabilidad está dada por características semánticas distintas que puede adquirir cierta palabra (corpus semántico), o por ciertos pesos asignados a las palabras según su grado de participación en una oración (corpus estadístico).

#### **3.1.3.2 Enfoque Automático**

Para entender el enfoque automático, es necesario definir lo que es el aprendizaje automático: Se define Machine Learning (Aprendizaje automático en español) como el campo de estudio que le da a las computadoras la habilidad de aprender sin haber sido explícitamente programadas [14].

Teniendo en cuenta la definición, el enfoque automático se caracteriza por utilizar algoritmos de aprendizaje automático para construir un clasificador que permitirá asignar el sentimiento o emoción. Esto implica que a medida que se va adquiriendo conocimiento de los patrones existentes en el texto, estos serán utilizados para clasificar nuevos documentos. En este enfoque, suelen utilizarse técnicas provenientes del aprendizaje automático y la estadística [6] y es necesario entender que el programa que realizará la clasificación debe aprender la experiencia de realizar una determinada tarea repetidas veces midiendo los resultados, a lo que se le llama entrenamiento.

Existen diversas formas de aprendizaje, dentro de las que se encuentra:

- Aprendizaje No Supervisado: Para los algoritmos basados en aprendizaje no supervisado, el entrenamiento se realiza a partir de una gran colección de experiencias y atributos de ellas pero sobre las cuales no se tienen datos de los resultados correctos. En este tipo de aprendizaje se busca encontrar similitudes o patrones que permitan realizar cierto agrupamiento.
- Aprendizaje Supervisado: Para los algoritmos basados en aprendizaje supervisado, el entrenamiento se realiza a partir de una colección de experiencias, atributos y resultados correctos en base a los cuales el algoritmo de clasificación predice nuevos resultados. Este tipo de aprendizaje utiliza clasificadores de distinto tipo, siendo los más utilizados los clasificadores lineales, clasificadores probabilísticos, clasificadores basados en árboles de decisiones y clasificadores basados en reglas.

### 3.2 Proceso de Análisis de Afecto

Si bien, el proceso para llevar a cabo una tarea de análisis de afecto no está definido como tal en la literatura, se puede apreciar que en la gran mayoría se realizan procedimientos comunes los cuales se establecen en el siguiente orden:

- a) Selección de una categoría afectiva
- b) Selección de un corpus de trabajo
- c) Limpieza del corpus de trabajo
- d) Selección de features
- e) Aplicación del clasificador automático: Entrenamiento y clasificación
- f) Análisis de resultados
- g) Aplicación de métricas de rendimiento de la clasificación

Cada una de estas tareas a realizar tiene un fundamento teórico el cual se estudia a continuación.

#### 3.2.1 Categorías Afectiva

El AA busca clasificar las emociones en categorías definidas por taxonomías propuestas por expertos del área de la psicología. Estas taxonomías establecen distintos niveles de detalle con respecto a la cantidad de categorías definidas y dentro de un proceso de AA es la persona quien realiza el proceso quien definirá qué categorización se utilizará. Dentro de las categorías propuestas en la literatura, destacan 3 categorizaciones más utilizadas en la actualidad.

Picard [15] menciona que existen 2 a 20 emociones básicas, y los 4 tipos más comunes son: miedo, ira, tristeza y alegría. Plutchik [16] distingue 8 emociones básicas: miedo, ira, tristeza, alegría, aversión, confianza, anticipación y sorpresa. Ekman [17] propone 6 categorías: ira, disgusto, miedo, alegría, tristeza y sorpresa.

La decisión de trabajar con alguna de las 3 categorizaciones está dada por el contexto de trabajo y la necesidad de detallar en menor o mayor medida las emociones contenidas en el texto.

#### 3.2.2 Preparación de los datos: limpieza del corpus

Con el objetivo de mejorar los resultados por el clasificador, y dependiendo de la tarea de procesamiento que se desee realizar, del origen de los datos y del idioma en que esté escrito el texto, pueden requerirse algunas transformaciones en el corpus de entrada antes de su procesamiento.

Algunas de las técnicas existentes para la preparación del corpus son:

- Tokenización: Separación de sentencias y palabras de un documento a partir de tokens, o caracteres especiales, que indican el término de una palabra o sentencia y el comienzo de la que sigue.

- Normalización: Consiste en unificar los términos que representan la misma información y pueden ser escritos en diferentes formatos. Por ejemplo, “restaurante, restorán, ristorante, restaurante, restó”
- Stemming: Permite obtener la raíz de la palabra eliminando sus terminaciones, con el fin de unificar aquellos términos que aportan la misma información al clasificador. Por ejemplo, los términos “recomendable, recomendamos, recomendar, recomendación” son reemplazados por su raíz “recommend”. Esta raíz de la palabra no necesariamente será un término válido del vocabulario.
- Lematización: Otra forma de unificar aquellos términos que aportan la misma información al clasificador es reemplazar cada palabra por su lema. El lema si es un término válido del vocabulario (a diferencia de la raíz de la palabra) y representa todas las formas flexionadas de la palabra, es decir, se eliminan todas las flexiones (género, conjugaciones, grado, número, etc).  
Ejemplo:  
-Lema (pésimo) = malo  
-Lema (empieza) = empezar  
-Lema (primeras) = primero
- Filtrado de StopWords: Es el nombre que reciben las palabras sin significado como artículos, pronombres, preposiciones, etc. y que no entregan información relevante al clasificador. Algunos ejemplos de StopWords en español son: un, desde, cierto, quien, solo, dentro, a, etc.
- Otros pre-procesamientos: Otras transformaciones que pueden representar una mejora en la efectividad del clasificador incluyen: eliminar términos o secuencias de palabras que no aportan información como signos de puntuación, fechas, caracteres especiales, emoticones, entre otros. Además puede ser de utilidad corregir errores ortográficos.

Hay veces que la utilización de estas técnicas de limpieza no representan una mejora en los resultados de la clasificación, llevando incluso a empeorar los resultados. Es por ello que hay que tener cuidado al momento de seleccionar las técnicas a utilizar.

### 3.2.3 Selección de Features: Etiquetado POS

Un feature, como se menciona en el apartado 3.1, es una característica del texto que resulta relevante para una tarea de procesamiento en particular. Existen features que resultan de gran utilidad en la mayoría de las tareas de clasificación y otros que resultan de utilidad seleccionando para la tarea específica a realizar. Por ejemplo: si la tarea consiste en identificar nombres propios, un feature trivial sería identificar las palabras que comienzan con mayúscula.

Uno de los features más utilizados de manera general en las tareas de procesamiento de textos es el etiquetado POS (del inglés, Part Of Speech Tagging). Esta tarea consiste a asignar a cada término del documento una etiqueta que identifique si la palabra actúa como sustantivo, verbo, adjetivo, artículo, adverbio, etc., dependiendo del contexto.

En clasificación de textos, puede ser de utilidad agregar a cada término un sufijo con la etiqueta asignada realizando etiquetado POS, con el fin de diferenciar aquellos términos idénticos pero que expresan significados distintos dependiendo del contexto del documento.

Por ejemplo:

-D1: “En este restaurant el vino es muy malo”.

-D2: “Todo salió perfecto, vino el chef y lo felicitamos por su plato”.

En este caso, puede interesarnos diferenciar la palabra “vino” cuando esta actúa como sustantivo, o se encuentra actuando como verbo. Si el ejemplo considera agregar la etiqueta POS correspondiente, un ejemplo para el término vino sería reemplazado por “vino\_SUST” y “vino\_VERB” según corresponda.

### 3.2.4 Algoritmos de clasificación Automática

Recordemos que se mencionó que el enfoque automático utiliza algoritmos de aprendizaje que pueden ser supervisados y no supervisados. Los más utilizados son los supervisados, debido a la forma de trabajo, que permite verificar los resultados de clasificación y así poder estudiar el rendimiento del algoritmo, que es parte del objetivo de esta investigación.

Los algoritmos supervisados de clasificación automática pueden ser divididos en clasificadores Generativos o Discriminativos (también llamados condicionales). Los clasificadores generativos aprenden el modo a partir de la probabilidad conjunta  $P(C, D)$  (Donde  $D$  es el documento y  $C$  la clase) y realizan la predicción maximizando la probabilidad condicional  $P(C|D)$  que se calcula utilizando el teorema de Bayes. A diferencia de estos, los clasificadores discriminativos modelan directamente la probabilidad condicional  $P(C|D)$  para realizar la predicción, es decir, a partir de la estructura oculta de los datos sin tener en cuenta la forma en que se generan. La desventaja de los algoritmos generativos es que resuelven un problema más general como paso intermedio, en cambio, los algoritmos discriminativos resuelven el problema de clasificación directamente [18].

Según la literatura estudiada, el algoritmo generativo más utilizado es el algoritmo Naive Bayes, y dentro de los modelos discriminativos más relevantes se encuentra el algoritmo de Árboles de Decisión y el algoritmo Support Vector Machine (SVM).

#### 3.2.4.1 Clasificador de Naive Bayes

En un método de clasificación supervisado y generativo que se basa en el teorema de Bayes y en la premisa de la independencia de los atributos de una clase. Esta premisa es conocida como “Naive Assumption” y se llama “Naive” (o ingenua), considerando que en la práctica los atributos raramente son independientes. Considerando el teorema de Bayes, se expresa como se muestra en la ecuación 1:

$$P(C_i|D) \propto P(C_i) \prod_{k=1}^n P(f_k|C_i) \quad (1)$$

Donde  $f_k$  son los atributos del documento,  $C_i$  es la clase y  $P(f_k|C_i)$  es la probabilidad de ocurrencia de la clase dada. La clase seleccionada por el clasificador será la que maximice la probabilidad anterior. Las implementaciones del algoritmo de Naive Bayes difieren principalmente en la aproximación de  $P(f_k|C_i)$  y las técnicas de smoothing utilizadas para el tratamiento de probabilidades bajas o nulas. Se conoce como smoothing a la técnica para suavizar o unificar la distribución de probabilidad, ajustando las bajas o nulas hacia arriba y las altas hacia abajo.

### 3.2.4.2 Clasificador de Árboles de Decisión

Es un método de clasificación supervisado y discriminativo donde el entrenamiento consiste en la construcción de un árbol de múltiples caminos en el que para cada nodo se busca el atributo que provee mayor ganancia de información para la clase. El árbol crece hasta su tamaño máximo y luego es acotado para mejorar su capacidad de generalización para los datos que no ocurren en el conjunto de datos de entrenamiento. A partir de este modelo se infieren reglas de decisión sobre las cuales se basa la clasificación.

Existen diversos algoritmos que trabajan con la construcción de árboles de decisión. El algoritmo más utilizado es el C4.5, donde la construcción del árbol se realiza en base a la ganancia de información que provee cada atributo. Este algoritmo convierte el árbol de decisión en un conjunto de reglas “si-entonces” y evalúa la exactitud de cada regla para definir el orden en que deben aplicarse. Para acotar el árbol, se realiza una poda que elimina aquellas precondiciones que al removerlas producen una mejora en la exactitud de la regla. La implementación más utilizada de este algoritmo es la que se conoce como J48, desarrollada en java [19].

### 3.2.4.3 Clasificador Support Vector Machine

Es un método de clasificación supervisado y discriminativo donde el entrenamiento consiste en encontrar un hiperplano que separe los vectores de atributos que representan los documentos del conjunto de datos en dos grupos, siendo esta separación la más grande posible. Aquellos vectores que definen los márgenes de la máxima separación entre las clases se conocen como support vectors. La ecuación 2 se utiliza para la predicción de la clase utilizando este modelo.

$$f(x) = \text{sign}(\sum_i \alpha_i x_i \cdot x + b) \quad (2)$$

Siendo,  $x$  el vector de atributos del documento a clasificar,  $i$  cada uno de los pesos que ponderan los vectores de atributos identificados como support features,  $x_i$  cada uno de los support features y  $b$  el término independiente. Un valor de -1 indicará que el documento pertenece a una clase y un valor de +1 a la otra, lo que representa de qué lado del hiperplano se encuentra  $x$ .

### 3.2.5 Métodos de evaluación de clasificadores supervisados

En los métodos supervisados, la clasificación depende del entrenamiento que se realice sobre el clasificador, por lo que su efectividad dependerá en gran medida de qué datos del corpus se utilicen para entrenamiento (training set) y cuáles para evaluarlo (test set). El criterio de selección de estos conjuntos de datos y la cantidad de veces que se evalúe el clasificador utilizando distintos criterios, definirá la confiabilidad de la evaluación. Los métodos más utilizados son:

- Hold-out (Set de entrenamiento): Es el método más sencillo de evaluación y consiste en dividir el conjunto de datos en dos partes: entrenamiento y prueba (generalmente se utilizan dos tercios para entrenamiento y un tercio para prueba). Utilizando este método se obtienen resultados para el conjunto de datos de prueba elegido pero se ignora el comportamiento del clasificador para



otros conjuntos de prueba, por lo que podría ocurrir que el clasificador funcione bien para los documentos de prueba elegidos pero no en otros casos.

- K-fold cross-validation (Validación cruzada): Este método es una mejora del método Hold-out. En este método el conjunto de datos se divide en k partes, cada una denominada fold, y se repite el método Hold-out k veces utilizando un fold como conjunto de prueba y el resto (k-1 fold) como conjunto de entrenamiento. Luego, las métricas del clasificador se calcularán como el promedio de las mediciones obtenidas en cada ejecución. Un ejemplo del funcionamiento de este método queda expresado en la figura:

1	TEST	TRAIN		
2	TRAIN	TEST	TRAIN	
3	TRAIN	TEST	TRAIN	
4	TRAIN	TEST	TRAIN	
5	TRAIN	TEST	TRAIN	
6	TRAIN	TEST	TRAIN	
7	TRAIN	TEST	TRAIN	
8	TRAIN	TEST	TRAIN	
9	TRAIN	TEST	TRAIN	
10	TRAIN	TEST	TRAIN	

Figura 1: Ejemplo 10-Fold Cross-Validation

El valor de k puede seleccionarse como k=3, 5, 10, etc. La desventaja del método es que al ejecutar la clasificación, un k mayor hace más costoso en tiempo/recursos computacionalmente hablando.

### 3.2.6 Métricas

Para poder evaluar el rendimiento de los clasificadores utilizados en el proceso de AA, se usan métricas generalmente utilizadas en recuperación de información, las cuales son adaptadas en función de los casos correctamente e incorrectamente clasificados. En general, las métricas consideran los siguientes casos:

- True Positives ( $t_p$ ): elementos a los que el clasificador asignó la clase relevante y esta era correcta.
- False Positives ( $f_p$ ): elementos a los que el clasificador asignó la clase relevante y esta no era correcta.
- False Negatives ( $f_n$ ): elementos a los que el clasificador asignó la clase no-relevante y esta no era correcta.
- True Negatives ( $t_n$ ): elementos a los que el clasificador asignó la clase no-relevante y esta era correcta.

Esto quiere decir:

	Relevantes	No Relevantes
Seleccionados	$t_p$	$f_p$
No Seleccionados	$f_n$	$t_n$

Tabla 1: Evaluación de los clasificadores - evaluación de textos

Estos valores mencionados anteriormente están dados según el documento estudiado, y a las coincidencias que entregue el clasificador sobre los estados afectivos definidos en el documento. Es así como es posible realizar el cálculo del rendimiento de acuerdo a las siguientes métricas:

- Accuracy: Representa la porción de documentos que son clasificados correctamente sobre el total de casos (Ecuación 3).

$$Accuracy_{TOTAL} = \frac{t_p + t_n}{t_p + t_n + f_p + f_n} \quad (3)$$

- Precision: En ella se representa la porción de documentos que son clasificados correctamente para la clase A sobre el total de casos clasificados como clase A (Ecuación 4).

$$Precision_A = \frac{t_p}{t_p + f_p} \quad (4)$$

- Recall: Esta métrica representa la porción de documentos de clase A que son clasificados correctamente (Ecuación 5).

$$Recall_A = \frac{t_p}{t_p + f_n} \quad (5)$$

- F-Measure: También llamada Medida-F, combina las medidas de precisión y recall a partir de la media armónica ponderada de estos dos valores (Ecuación 6).

$$F_1 = \frac{2PR}{P+R} \quad (6)$$

Para entender las métricas descritas anteriormente, se presenta el siguiente ejemplo: Considerar el siguiente conjunto de datos: (clases disponibles: Alegría, Confianza, Miedo y Tristeza).

Documento	Clase	Predicción del Clasificador
D1	Alegría	Alegría
D2	Alegría	Alegría
D3	Alegría	Miedo
D4	Alegría	Tristeza
D5	Miedo	Miedo
D6	Tristeza	Alegría
D7	Miedo	Confianza
D8	Miedo	Miedo
D9	Tristeza	Tristeza
D10	Confianza	Miedo
D11	Alegría	Alegría

D12	Alegría	Confianza
D13	Confianza	Miedo
D14	Miedo	Miedo
D15	Miedo	Confianza
D16	Tristeza	Confianza
D17	Alegría	Tristeza
D18	Alegría	Alegría
D19	Confianza	Tristeza
D20	Miedo	Alegría

Tabla 2: Evaluación de los clasificadores – Ejemplo de clasificación

De acuerdo a los datos entregados por el clasificador, se pueden obtener los siguientes casos:

	Alegría	Confianza	Miedo	Tristeza
<b>Clasificados como Alegría</b>	4	0	1	1
<b>Clasificados como Confianza</b>	1	0	2	1
<b>Clasificados como Miedo</b>	1	2	3	0
<b>Clasificados como Tristeza</b>	2	1	0	1

Tabla 3: Evaluación de los clasificadores - Ejemplo de métricas

Para los datos anteriores, obtendremos las siguientes métricas:

- $Accuracy_{TOTAL} = 8/20 = 0.20$
- $Precision_{ALEGRIA} = 4/6 = 0.66$
- $Recall_{ALEGRIA} = 4/8 = 0.50$
- $F1_{ALEGRIA} = 0.57$

Podemos interpretar los valores obtenidos anteriormente como:

- El clasificador acierta el 20% de las veces ( $Accuracy_{TOTAL}$ ).
- Los documentos de alegría se predicen con una probabilidad de acierto del 66% ( $Precision_{ALEGRIA}$ ).
- Del total de documentos de alegría, el 50% será clasificado como tal ( $Recall_{ALEGRIA}$ ).
- La efectividad del clasificador para documentos de alegría, teniendo en cuenta la precisión y el recall, es del 57% ( $F1_{ALEGRIA}$ ).

### 3.3 Herramientas utilizadas en análisis de texto

En esta sección se describen y analizan algunas herramientas para el Procesamiento de Lenguaje Natural y para la extracción de información, que es una de las principales funciones que realiza el PLN. De estas herramientas, se seleccionarán las que mejor se adapten al objetivo de la investigación. Para ello, se explica en qué consisten estas herramientas y las ventajas de su uso. Una explicación en detalle del uso de cada una de ellas se presenta en la sección 4.

### 3.3.1 Gate

Esta es una herramienta para el desarrollo y despliegue de componentes que procesan el lenguaje natural, destacando en el análisis de textos de cualquier tamaño [20]. Además es un software de código abierto desarrollado por la Universidad de Sheffield (Reino Unido), en respuesta a la necesidad de un proceso de estandarización para el desarrollo de herramientas de análisis lingüístico. El proyecto comenzó en 1995 y todavía está activo con muchas actualizaciones y mejoras [21].

Este framework cuenta con un conjunto de módulos llamados ANNIE (A Nearly- New Information Extraction System), y son los siguientes:

- Tokenizador
- División de oraciones
- Etiquetador gramatical
- Reconocedor de nombres
- Etiquetador de correferencia

Los módulos de ANNIE vienen por defecto configurados para trabajar con textos en idioma inglés, pero es posible instalar un plugin de idiomas para poder implementar algunos de ellos en el idioma español.

### 3.3.2 Stanford coreNLP

CoreNLP es un conjunto de herramientas para el procesamiento del lenguaje natural que se basa en modelos estadísticos, desarrollado por el Grupo de Procesamiento de Lenguaje Natural de la Universidad de Stanford [22]. Estas herramientas proporciona un alto nivel de PLN, están desarrolladas en Java, pero también existe para las siguientes plataformas: Apache Thrift, C#/F#/ .NET, Ruby, Python, Perl, Scala, Clojure, ZeroMq Server y Javascript. Estos lenguajes pueden variar dependiendo del módulo desarrollado.

En este kit de herramientas de análisis del lenguaje natural se puede ingresar un texto sin procesar y de esta forma llegar a presentar la estructura base de las palabras, partes de la oración, nombres de personas u organizaciones, etc. Una de las finalidades que tiene es facilitar la aplicación de herramientas de análisis lingüístico a una parte de un texto.

La herramienta integra algunas aplicaciones de PLN las cuales son las siguientes:

- Etiquetador (POS)
- Reconocedor de la entidad (NER)
- El analizador
- El sistema de resolución de la correferencia
- El análisis de sentimientos
- El patrón de aprendizaje bootstrapper

Esta herramienta ofrece módulos que trabajan con textos en idioma inglés, pero existen algunas versiones que ofrecen estos módulos con cierta compatibilidad, la cual es posible ver en la tabla 4:

Anotador	Árabe	Mandarín	Inglés	Francés	Alemán	Español
Tokenizador	✓	✓	✓	✓	✓	✓
División de sentencias	✓	✓	✓	✓	✓	✓
Truecase			✓			
POS	✓	✓	✓	✓	✓	✓
Lema			✓			
Gender			✓			
NER		✓	✓		✓	✓
Regexner	✓	✓	✓	✓	✓	✓
Parsing	✓	✓	✓	✓	✓	✓
Análisis de dependencias		✓	✓			
Análisis de sentimientos			✓			
Coref			✓			

Tabla 4: Módulos disponibles para los lenguajes soportados CoreNLP – fuente: [22]

Dentro de las aplicaciones más utilizadas provistas por esta herramienta se encuentra el POS Tagger (Etiquetador POS). Lo que hace esta aplicación es identificar la morfología de una palabra según su aparición dentro de una sentencia. Para ello, utilizan una nomenclatura propuesta por el grupo EAGLES para la anotación morfosintáctica de los lexicones para los idiomas europeos. La estructura de las etiquetas está definida según lo presentado en el anexo.

### 3.3.3 WEKA

Weka es una colección de algoritmos de aprendizaje automático para tareas de minería de datos desarrollada por un grupo de investigadores del área del PLM de la Universidad de Waikato [19]. Los algoritmos disponibles en Weka pueden ser aplicados directamente a un conjunto de datos o llamados desde su propio código Java. Weka contiene herramientas para pre-procesamiento de datos, clasificación, regresión, clustering, reglas de asociación y visualización.

La herramienta permite ingresar un set de datos, la cual debe estar en formato ARFF (formato propio de Weka) para luego poder realizar las tareas requeridas. El formato ARFF solicita definir en su estructura los atributos y el tipo de dato de los elementos por los que está compuesto el set de datos. Un ejemplo de la estructura de un archivo ARFF es la que se presenta en la ilustración 1.

```
@Relation Corpus

@Attribute Titular STRING
@Attribute Affect {alegria,confianza,miedo,sorpresa,tristeza,aversion,ira,anticipacion,ninguna}

@Data

'Banco de Francia preve un crecimiento del 02% en el segundo trimestre del ano',confianza
'Joven guardia de seguridad muere tras ser acribillado en Pudahuel',ira
'Rusia no negociara rebajas de gas hasta que Ucrania no pague deudas',aversion
'Alemania coloca bonos por US$ 2734 millones a seis meses',ninguna
'Cervecera espanola Mahou San Miguel firma acuerdo con empresa filipina',alegria
'Bolsa de Tokio cierra a la baja por inestabilidad en Ucrania y subida del yen',anticipacion
'Expertos seguridad de Google alertan de vulnerabilidad de los medios',miedo
'Encuesta del Banco Central preve mantencion de la Tasa de Politica Monetaria en 400%',ninguna
'El mundo del rock de Manchester tambien celebro el titulo del City',alegria
'Dolar abre sesion a la baja presionado por la importante alza del cobre en Londres',anticipacion
```

Figura 2: Ejemplo de estructura de archivo ARFF - Datos relacionados a "Titulares de periódico"

En relación al pre-procesamiento, Weka provee de distintas herramientas como lo son tokenizadores, stemmers, filtrado de StopWords, entre otros, todos ellos desarrollados para el idioma inglés sin tener compatibilidad en otros idiomas. Aun así, es posible cargar al sistema un set de datos previamente procesados con otras herramientas, el cual debe estar en el formato ARFF mencionado anteriormente.

En relación a la clasificación, Weka incorpora diversos algoritmos de clasificación automática, dentro de los que se incluyen los mencionados en apartados anteriores. Estos algoritmos son utilizados sobre un set de datos y realizan la clasificación mediante métodos de set de entrenamiento o de validación cruzada (explicados en la sección 3.2.5).

### 3.4 Trabajos Relacionados

Dentro del objetivo que se busca alcanzar con el desarrollo de esta investigación, en la literatura existen comparaciones al rendimiento de técnicas de clasificación automática, pero estas han sido realizadas para procesos de análisis de sentimiento (AS), es decir, clasificando según la polaridad, y no por clases de afectos (AA). Adicionalmente, las comparaciones reportadas son en su mayoría aplicadas sobre corpus en idioma inglés y sólo algunas en otros idiomas como el español, ruso, chino o hindú [7][23][24][25][26][27].

Por ejemplo, en [23] se presenta el uso a lo largo del tiempo de las técnicas SVM, NB, DT, entre otras, en diversos trabajos realizados. En [7] se presenta un trabajo de AS sobre un corpus de opiniones de restaurants en idioma español, en el cual se logra obtener de los resultados de la clasificación, valores de rendimiento promedio para NB de 91,4%, para SVM de 91,0% y para DT de 85,8%. En [24], se compara el rendimiento de SVM, NB y ME, pero incorporando features a la clasificación como lo son la frecuencia de unigramas y etiquetado POS. Las técnicas se utilizaron sobre un set de 1200 tweets, con opiniones y comentarios acerca de productos electrónicos como celulares y computadoras donde el rendimiento reportado fue similar para todos los clasificadores utilizados, llegando a valores cercanos al 90% para SVM y NB. En [25] se realiza un estudio sobre un corpus de valoraciones de clientes de un banco, para ver cómo la lematización afecta la precisión de los clasificadores SVM y NB, llegando a un rendimiento de 88% para SVM y 87% para NB.

Los trabajos realizados en procesos de análisis de afecto (AA), tanto en idioma inglés como en otros idiomas reportados, en su mayoría aplican el enfoque basado en diccionario (lexicones) [28][29][30].

Por ejemplo, en [28] se aplicaron técnicas basadas en diccionario para clasificar un texto en 7 categorías, donde el clasificador basado en un Multinomial Naive Bayes obtiene un rendimiento del 71,35% de precisión. En [29], utilizan un diccionario basado en WordNet para predecir textos y clasificarlos en 6 categorías, utilizando diferentes técnicas, entre ellas, un clasificador Naive Bayes. El rendimiento de estas técnicas varió entre un 29% y un 55%, obteniendo NB un 31,2%. En [30] desarrollan un modelo para detectar emociones sobre un corpus en chino que trabaja con 7 categorías, mediante la combinación de métodos basados en lexicones y SVM. El rendimiento presentado en este trabajo varía entre el 28% y 41%.

#### 4.- METODOLOGÍA

La metodología de trabajo definida para esta investigación está propuesta con el fin de cumplir con los objetivos establecidos en la sección 2 y comprobar o refutar la hipótesis propuesta.

Las actividades realizadas en este trabajo comienzan con la construcción y etiquetado del corpus. Para ello, es necesario identificar el dominio de trabajo, el cual serán titulares de periódicos online de Chile. Junto con ello, es necesario identificar la categoría afectiva con la cual se clasificarán las emociones.

Con el objetivo de mejorar los resultados obtenidos por los clasificadores a estudiar, se aplicaron algunas transformaciones al corpus relacionadas con la limpieza. Las transformaciones utilizadas fueron: normalización, tokenización, corrección de errores ortográficos, eliminación de StopWords, eliminación de caracteres o palabras repetidas, entre otras. Además, es necesario considerar ciertos atributos (features) que permitirán encontrar patrones dentro de los textos para así ayudar a predecir la clasificación de las opiniones expresadas en un texto. Para esto último, se utilizó la técnica de etiquetado POS.

A partir de los procesamiento mencionados aplicados al corpus, se construyeron 4 versiones distintas sobre las cuales se aplicaron posteriormente las técnicas de clasificación. Lo anterior, con el objeto de analizar la incidencia en el rendimiento de un clasificador de la aplicación de StopWords y del etiquetado POS.

Luego se aplicaron tres técnicas de clasificación sobre cada una de los 4 corpus generados. Posteriormente los resultados obtenidos por los clasificadores, son analizados mediante el cálculo de las métricas Accuracy, Precision, Recall y F-measure.

Finalmente, se entrega un análisis general de la evaluación de las métricas de manera individual y en conjunto, discutiendo el rendimiento de los clasificadores en las distintas versiones del corpus. La Figura 3 muestra un resumen con las actividades descritas anteriormente.

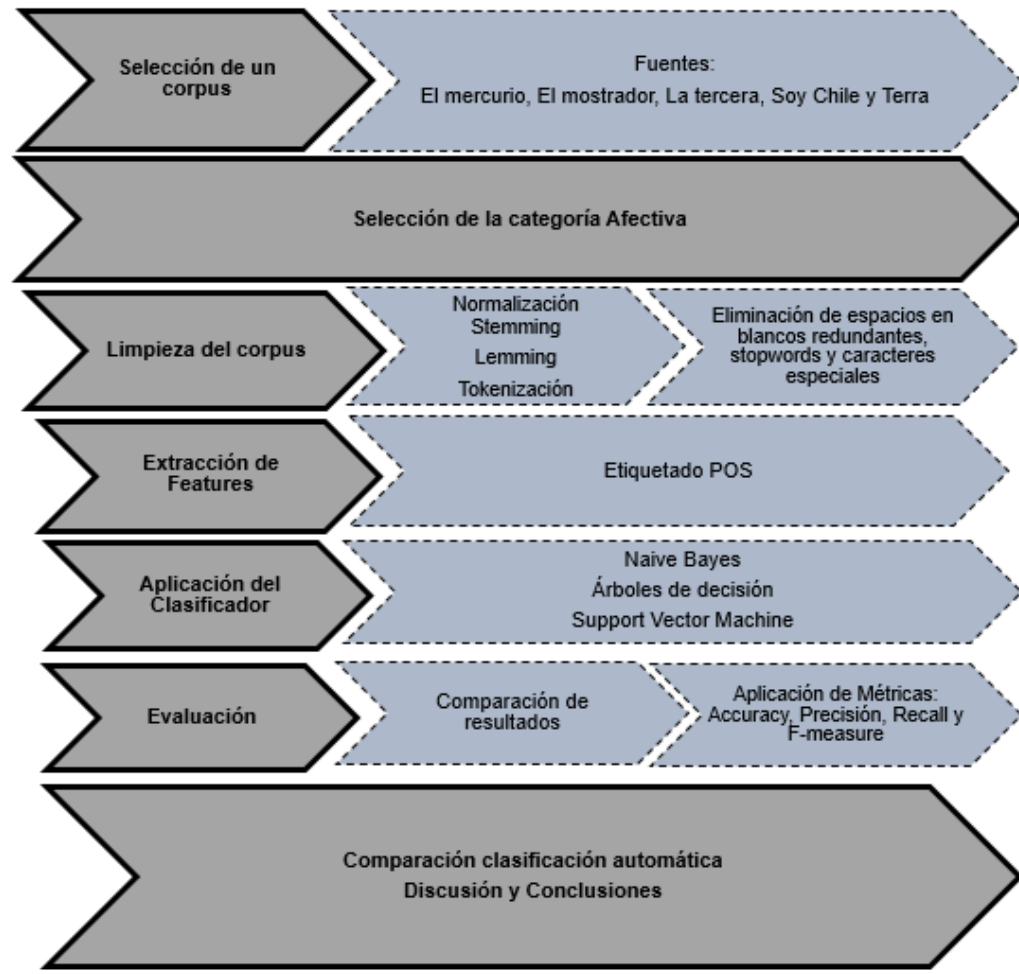


Figura 3: Metodología de trabajo

Los siguientes apartados del capítulo muestran un poco más en profundidad algunos aspectos interesantes de analizar, con el fin de validar la metodología propuesta.

#### 4.1 Selección de la categoría Afectiva

Para poder realizar un proceso de AA, es necesario definir un modelo que proporcione las categorías afectivas que se utilizarán en la clasificación. De los modelos existentes, descritos en la sección 3.2.1, se decidió utilizar la categorización de 8 emociones propuesta por Plutchik, considerando las razones expuestas en [31]: (a) como algunas de las otras opciones de clasificación de emociones básicas, esta opción también está bien fundada en la investigación psicológica, fisiológica, y empírica; (b) a diferencia de otras opciones, por ejemplo la propuesta por Ekman, no se compone de emociones en su mayoría negativas y provee un abanico de opciones más equitativo entre emociones positivas y negativas; y por último (c) es un superconjunto de las emociones propuestas por algunos otros, por ejemplo, las categorías propuestas por Plutchik forman un superconjunto que contiene a las seis emociones propuestas por Ekman.



## 4.2 Corpus de Trabajo

Se decidió establecer como dominio de trabajo textos de titulares de periódicos chilenos en su versión online. Las fuentes consideradas son las siguientes: El Mercurio, El Mostrador, La Tercera, Soy Chile y Terra.cl. Estos titulares están escritos en idioma español y fueron publicados entre marzo del 2014 y abril del 2015.

Una consideración previa se relaciona con la naturaleza de los titulares de periódicos. Los titulares relatan hechos, por lo que puede que algunos titulares no expresen algún tipo de emoción. Por este motivo es que se añadió a la categorización de 8 emociones de Plutchik la categoría “ninguna”, resultando en 9 categorías posibles para clasificar los titulares (miedo, ira, tristeza, alegría, aversión, confianza, anticipación, sorpresa y ninguna).

Para construir el corpus, fue necesario extraer titulares de las fuentes mencionadas, dando a lugar un total de 2953 titulares. Para que el corpus pueda ser entrenado y puesto a prueba por los clasificadores, se hace necesario etiquetar cada titular con la emoción correspondiente que pretende transmitir y que está sea lo más objetiva posible.

Para ello, el corpus fue previamente etiquetado por 3 evaluadores humanos de entre 18 y 41 años, quienes clasificaron los 2953 titulares en algunas de las siguientes categorías: miedo, ira, tristeza, alegría, aversión, confianza, anticipación, sorpresa y ninguna. La tabla 5 presenta un extracto del corpus en donde se muestran titulares de periódicos y las etiquetas asignadas por los evaluadores.

Titular	Fuente	Rater 1	Rater 2	Rater 3
Arturo Vidal manda alentador mensaje para la "Roja" desde Italia: "Vamos que llegamos"	El Mercurio	Alegría	Alegría	Alegría
Felipe Berríos y Reforma Educacional: “Mientras haya clasismo en Chile, cualquier cosa que se haga saldrá mal”	El Mostrador	Confianza	Confianza	Ira
Indignación en Turquía por asesor de Erdogan que pateó a manifestante en mina	La Tercera	Aversión	Ira	Ira
El optimismo a toda prueba de Lindorfo	Soy Chile	Anticipación	Ninguna	Alegría
Jamaica aumenta la seguridad aérea ante la visita del presidente de EE.UU.	Terra.cl	Confianza	Confianza	Confianza

Tabla 5: Ejemplo de titulares y etiquetado previo de los evaluadores

Se analizó el nivel de acuerdo de los evaluadores mediante el coeficiente Kappa [29] que tiene en cuenta el hecho de que los evaluadores pueden estar en acuerdo, desacuerdo o simplemente clasificar por mera casualidad. Para ello, se parte de la premisa de que se requieren clasificar N elementos en C categorías mutuamente excluyentes. Queda expresado en la ecuación 6:

$$K = \frac{Pr(a) - Pr(e)}{1 - Pr(e)} \quad (6)$$

Donde Pr (a) es el acuerdo observado relativo entre los observadores, y Pr (e) es la probabilidad hipotética de acuerdo por azar, utilizando los datos observados para calcular las probabilidades de que cada observador clasifique aleatoriamente cada categoría. El coeficiente de Kappa mide el nivel de acuerdo en un rango de valores entre un valor mínimo igual a 0 y un máximo de 1 que indica acuerdo absoluto.

Para optimizar este coeficiente y obtener un nivel de acuerdo aceptable, en cada uno de los titulares, al menos 2 de los evaluadores debían coincidir en la misma emoción. En los casos en que se obtuvo 3 etiquetas distintas para un titular, este fue eliminado del corpus.

Titular	Fuente	Rater 1	Rater 2	Rater 3	Se conserva
Arturo Vidal manda alentador mensaje para la "Roja" desde Italia: "Vamos que llegamos"	El Mercurio	Alegría	Alegría	Alegría	✓
Felipe Berríos y Reforma Educacional: "Mientras haya clasismo en Chile, cualquier cosa que se haga saldrá mal"	El Mostrador	Confianza	Confianza	Ira	✓
Indignación en Turquía por asesor de Erdogan que pateó a manifestante en mina	La Tercera	Aversión	Ira	Ira	✓
El optimismo a toda prueba de Lindorfo	Soy Chile	Anticipación	Ninguna	Alegría	X
Jamaica aumenta la seguridad aérea ante la visita del presidente de EE.UU.	Terra.cl	Confianza	Confianza	Confianza	✓

Tabla 6: Ejemplo de titulares y decisión de mantener o eliminar titular

Tras realizar el proceso mencionado anteriormente, el corpus quedó compuesto por 2031 titulares, con los cuales se obtuvo un valor de Kappa= 0,39. El cálculo de este valor fue obtenido con la herramienta "Graph Pad Software" y el detalle de las tablas de confusión y el cálculo se encuentra disponible en el anexo 1.

Aunque este valor puede parecer bajo, se considera aceptable tomando en consideración que cada titular podía ser etiquetado en 9 categorías afectivas, lo que hace difícil un alto nivel de acuerdo entre los evaluadores. Así, a cada titular se le asignó la clase afectiva predominante. La tabla 7 muestra para el ejemplo antes mencionado el resultado final de este proceso.

Titular	Emoción
Arturo Vidal manda alentador mensaje para la "Roja" desde Italia: "Vamos que llegamos"	Alegría
Felipe Berríos y Reforma Educacional: "Mientras haya clasismo en Chile, cualquier cosa que se haga saldrá mal"	Confianza
Indignación en Turquía por asesor de Erdogan que pateó a manifestante en mina	Ira
Jamaica aumenta la seguridad aérea ante la visita del presidente de EE.UU.	Confianza

Tabla 7: Extracto del corpus resultante

De esta forma, las estadísticas del corpus (Sin considerar la etiqueta de emoción de cada titular) son las siguientes:

- Cantidad de titulares: 2031 titulares.
- Cantidad de palabras totales: 25.334 palabras.
- Cantidad de caracteres utilizados: 157.429 caracteres.

Las palabras más utilizadas que componen el corpus se muestran mediante una nube de palabras, presentada en la Figura 4:



Figura 4: Nube de palabras del corpus final

La distribución de fuentes de los titulares que componen el corpus final se distribuye de la siguiente forma:

- El mercurio: 840 titulares.
- El mostrador: 197 titulares.
- La tercera: 202 titulares.
- Soy Chile: 157 titulares.
- Terra.cl: 635 titulares.

El balance del corpus de acuerdo a la cantidad de titulares que están clasificados en las distintas categorías afectivas queda expresado según el gráfico de la figura 5.

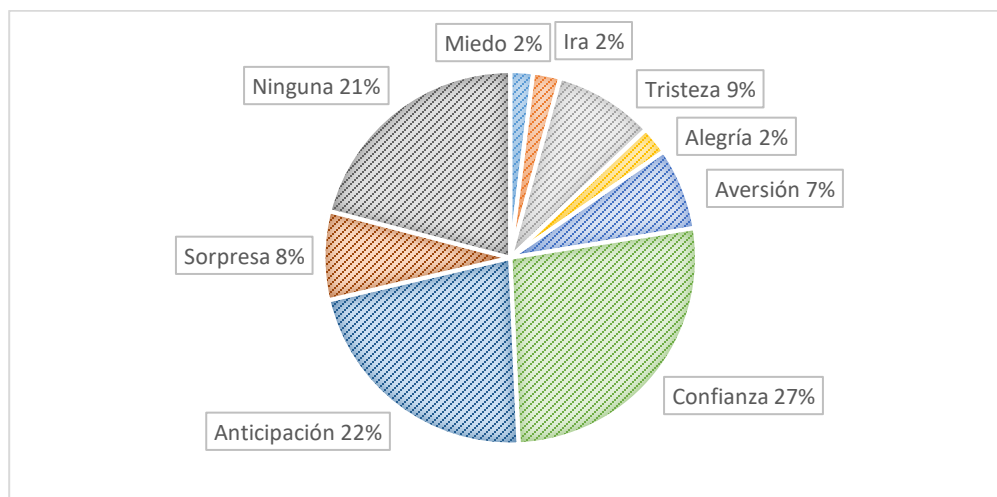


Figura 5: Balance del Corpus

El corpus de trabajo queda disponible en texto plano o en formato Excel, este último, con el fin de mantener un orden de los documentos y acceder directamente a los campos requeridos. <sup>1</sup>

<sup>1</sup> Corpus original disponible en: [http://dsi.face.ubiobio.cl/somos/newscorpus/newscorpus\\_original.xls](http://dsi.face.ubiobio.cl/somos/newscorpus/newscorpus_original.xls)

La hoja está compuesta por 2 columnas, donde la primera corresponde al titular y la segunda corresponde al afecto asociado. Cada una de las filas va a contener los documentos a estudiar.

### 4.3 Experimento

El objetivo de este experimento es comparar el rendimiento de los 3 algoritmos de clasificación más utilizados en la literatura: Árboles de decisión, Naive Bayes y Support Vector Machine. Además, se desea estudiar la incidencia de la aplicación de técnicas de StopWords y etiquetado POS en el rendimiento del clasificador.

Para ello, se aplicará un proceso de análisis de afecto sobre el corpus elaborado en la sección 4.2. Este proceso considera una etapa de limpieza del corpus y selección de features según corresponda, para luego entrenar el clasificador y realizar la clasificación correspondiente utilizando los algoritmos mencionados anteriormente. Finalmente, es posible estudiar y comparar el rendimiento de estos por medio de las métricas obtenidas.

Para realizar este proceso, se desarrolla un programa en lenguaje de programación Java que permite realizar la lectura del corpus en Excel, realizar los procesos de limpieza y aplicar los features necesarios; esto de acuerdo a las herramientas disponibles mencionadas en el apartado 3.3. Ya que el proceso de clasificación automática y obtención de métricas puede ser realizado con la herramienta WEKA, es necesario que el programa genere finalmente un archivo ARFF con el corpus ya procesado.

Finalmente, los resultados obtenidos con la herramienta WEKA, serán procesados mediante tablas y gráficos y analizados posteriormente.

El proceso general a realizar en la etapa de experimentación queda ilustrado en la figura 6.

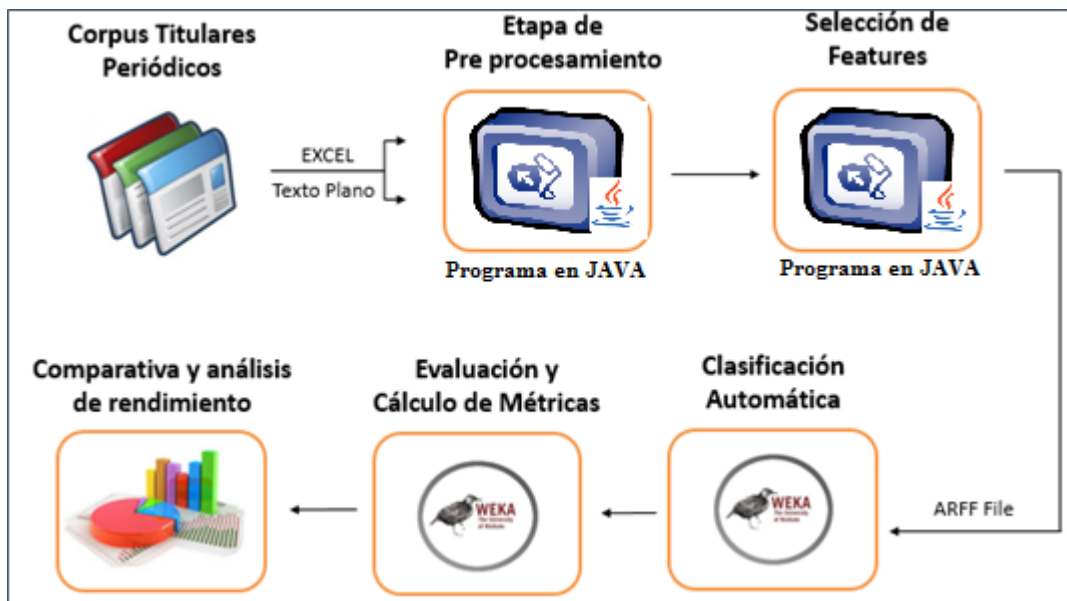


Figura 6: Proceso general de experimentación

Debido a que se desea estudiar la incidencia de técnicas de StopWords y etiquetado POS, en el proceso de experimentación ya definido, cuando el proceso de análisis de afecto desea entrenar y realizar la clasificación correspondiente, es necesario contar con 4 versiones pre-procesadas del corpus original. Estas quedan descritas en la tabla 8.

Corpus	Limpieza	Filtro stopwords	Uso de etiquetado POS
Corpus 1	✓		
Corpus 2	✓	✓	
Corpus 3	✓		✓
Corpus 4	✓	✓	✓

Tabla 8: Versiones del corpus para el experimento

- Corpus 1: Conformado por el corpus original al cual se le aplicó un proceso de limpieza, que incluyó: normalización, tokenización, eliminación de palabras y caracteres repetidos, eliminación de signos de puntuación, exclamación e interrogación, por ejemplo “”, ’, /, \, ;, ¡, ¿. entre otros. Además, se reemplazaron letras “ñ” por n y se eliminaron las tildes. Nota: La eliminación de tildes y el reemplazo de la letra “ñ” se realizó debido a la incompatibilidad de WEKA para el tratamiento de estos caracteres.
- Corpus 2: A partir del Corpus 1 se aplicó un proceso de filtrado de stopwords, el cual consiste en eliminar todas las palabras sin significado como artículos, pronombres, preposiciones, etc., y que no son relevantes para el clasificador.
- Corpus 3: A partir del Corpus 1 se aplicó un proceso de etiquetado POS.
- Corpus 4: A partir del Corpus 1 aplicó un proceso de filtrado de stopwords más etiquetado POS.

Cada uno de estos corpus es necesario tenerlos en formato ARFF, por lo que para obtener cada una de las 4 versiones necesarias, se determinó trabajar con 4 versiones del programa en JAVA. Todas las versiones realizan la lectura del mismo archivo que contiene al corpus original en formato Excel, pero varían en la aplicación de los distintos filtros de limpieza y selección de features para así generar el archivo ARFF necesario para cada versión del corpus.

Así, un ejemplo para mostrar la diferencia entre las distintas versiones del corpus es lo que se presenta en la tabla:

Titular Original	¿Los ‘hackers’ siempre son los malos del paseo?
Versión 1	Los hackers siempre son los malos del paseo
Versión 2	hackers siempre malos paseo
Versión 3	Los_da0000 hackers_nc0p000 siempre_rg son_vsip000 los_da0000 malos_aq0000 del_sp000 paseo_nc0s000
Versión 4	hackers_nc0p000 siempre_rg malos_aq0000 paseo_nc0s000

Tabla 9: Ejemplo de las versiones del corpus

Las 4 versiones del corpus quedan disponibles en formato ARFF<sup>2</sup> para ser leído directamente por WEKA. Para entender cómo el programa genera cada una de las versiones, es que se explica parte del funcionamiento del programa en JAVA en el siguiente apartado.

2

Versión 1: <http://dsi.face.ubiobio.cl/somos/newscorpus/corpusarffversions/SinPosSinStopwords.arff>

Versión 2: <http://dsi.face.ubiobio.cl/somos/newscorpus/corpusarffversions/SinPosConStopwords.arff>

Versión 3: <http://dsi.face.ubiobio.cl/somos/newscorpus/corpusarffversions/ConPosSinStopwords.arff>

Versión 4: <http://dsi.face.ubiobio.cl/somos/newscorpus/corpusarffversions/ConPosConStopwords.arff>

### 4.3.1 Programa en JAVA

Esta sección pretende explicar a grandes rasgos el funcionamiento del programa que permite generar las 4 versiones necesarias del corpus, mencionadas en el apartado anterior.

Para poder utilizar el programa, es necesario instalar las siguientes librerías:

- Apache POI: Para realizar la lectura del archivo en formato Excel.
- Stanford NLP Tagger: Para aplicar el etiquetado POS para idioma español.

Para realizar la lectura del archivo que contiene al corpus original en formato Excel, el programa permite buscar el archivo en el equipo gracias a una interfaz gráfica implementada con un JFileChooser. Inmediatamente, el programa permite buscar una ruta de destino en la cual se va a almacenar el archivo de salida en formato ARFF.

Posteriormente, el programa escribe en el archivo de salida la cabecera del archivo de acuerdo al formato de salida establecido para los archivos en formato ARFF y de acuerdo a los atributos que se van a almacenar.

Luego, utilizando los métodos provistos por Apache POI, se realiza la lectura del Excel para cada uno de los documentos, recordando que está compuesto por 2 columnas, donde la primera corresponde al titular y la segunda corresponde al afecto asociado. Ubicándose en una fila, se lee la columna correspondiente al titular y se aplican los filtros de limpieza y de selección de features de acuerdo a la versión del corpus que se desea crear para la clasificación. Una vez aplicados los filtros, se añaden al archivo de salida en el apartado de set de datos. Se continúa con la columna de afecto, en la cual sólo se aplica el filtro de limpieza. Se repite este proceso para cada una de las filas disponibles en el corpus. Finalmente, se procede a cerrar el archivo de salida y finaliza el programa.

Para aplicar los filtros de limpieza mencionados anteriormente, se desarrollaron los métodos que se describen a continuación:

- filterComillasEspacios: Permite eliminar todo tipo de comillas, además de eliminar el espacio doble o superior. Este filtro es utilizado para generar todas las versiones.

```
private static String filterComillasEspacios(String stringCellValue){
    String output = stringCellValue.replaceAll("\\\"", "")
        .replaceAll("\\'", "")
        .replaceAll("\"", "")
        .replaceAll("'", "")
        .replaceAll(" ", "")
        .replaceAll(" ", "");
    while(output.contains(" ")) {
        output = output.replaceAll(" ", "");
    }
    return output;
}
```

Figura 7: Método para eliminar comillas y espacios dobles o superior

- filterTildes: Permite reemplazar letras con tildes, diéresis, ñ y caracteres especiales de letras por su homólogo sin caracteres especiales. Este filtro es utilizado para generar todas las versiones.

```
private static String filterTildes(String stringCellValue) {
    String output = stringCellValue.replaceAll("á", "a")
        .replaceAll("é", "e")
        .replaceAll("í", "i")
        .replaceAll("ó", "o")
        .replaceAll("ú", "u")
        .replaceAll("Á", "A")
        .replaceAll("É", "E")
        .replaceAll("Í", "I")
        .replaceAll("Ó", "O")
        .replaceAll("Ú", "U")
        .replaceAll("ü", "u")
        .replaceAll("Ü", "U")
        .replaceAll("ñ", "n")
        .replaceAll("Ñ", "N")
        .replaceAll("ã", "a")
        .replaceAll("Ã", "A");
    return output;
}
```

Figura 8: Método para reemplazo de caracteres especiales

- filterSignos: Permite eliminar los signos de puntuación. Este filtro es utilizado para generar todas las versiones.

```
private static String filterSignos(String stringCellValue) {
    return stringCellValue.replaceAll("\\.", "").replaceAll(":", "")
        .replaceAll("...", "").replaceAll(",", "");
}
```

Figura 9: Método para eliminar signos de puntuación

- filterExclInt: permite eliminar los signos de exclamación e interrogación. Este filtro es utilizado para generar todas las versiones.

```
private static String filterExclInt(String stringCellValue) {
    return stringCellValue.replaceAll("¿", "").replaceAll("\\?", "")
        .replaceAll("¡", "").replaceAll("!", "");
}
```

Figura 10: Método para eliminar signos de exclamación e interrogación

- filterStopWords: Permite eliminar los StopWords leídos desde archivo, los cuales se encuentran posteriormente almacenados en un ArrayList. Este filtro es utilizado solamente para generar las versiones 2 y 4. En las versiones 1 y 3 no es utilizado.



```
private static String filterStopWords(String stringValue, ArrayList<String> stopWordsVector){
    ArrayList<String> stopLower = new ArrayList<>();
    String output = " "+stringValue+ " ";
    for(String stopword: stopWordsVector){
        stopLower.add(stopword.toLowerCase());
        output = output.replaceAll(" "+stopword+" ", " ");
    }
    for(String stopword2: stopLower){
        output = output.replaceAll(" "+stopword2+" ", " ");
    }
    return output.substring(1, output.length()-1);
}
```

Figura 11: Método para eliminar StopWords

La idea de realizar la limpieza por medio de métodos independientes responde a la necesidad de poder aplicar estos filtros según la configuración de limpieza que se requiera, esto dado por el contexto, los documentos y la necesidad de estudio del investigador pensando en trabajos futuros.

Por otro lado, para poder aplicar el feature correspondiente al etiquetado POS, es necesario configurar en el programa la herramienta provista por el Stanford POS Tagger, utilizando los modelos para el idioma español. Para ello es necesario realizar lo siguiente:

1. En la dirección o ruta, donde se ubica el nuevo proyecto se crea una carpeta con el nombre de taggers (puede tener otro nombre).
2. Descomprimir archivos provistos por Stanford (<http://nlp.stanford.edu/software/tagger.shtml>) y ubicarlos en la carpeta creada en el paso anterior, estos archivos tienen la extensión tagger y props.
3. Importar la librería para dejar disponible el modelo en el programa.
4. Concluida la configuración, codificar el método principal en el que constará un constructor de la clase MaxentTagger, al cual se le tiene que enviar como parámetro un archivo entrenado, que en este caso corresponde al tagger en español.
5. Por último se etiqueta la cadena a procesar.

Para poder entender de mejor manera el funcionamiento del Stanford POS Tagger, en la figura 12 se presenta un ejemplo en el cual se etiqueta una cadena de prueba:

```
import edu.stanford.nlp.tagger.maxent.MaxentTagger;

public class Tagger {
    public static void main(String[] args){
        //Iniciar el tagger
        MaxentTagger tagger = new MaxentTagger("taggers/spanish.tagger");
        // Ejemplo de una cadena
        String ejemplo = "Esta es una oración de prueba";
        // Cadena
        String taggeo = tagger.tagString(ejemplo);
        // Presentamos el resultado
        System.out.println(taggeo);
    }
}
```

Figura 12: Ejemplo de funcionamiento del Stanford POS Tagger



Al momento de ejecutar este código el resultado es el siguiente:

Esta\_pd000000 es\_vsip000 una\_di0000 oración\_nc0s000 de\_sp000 prueba\_nc0s000

Cada una de las palabras que componen el string de prueba queda con su etiqueta asociada. El significado de cada una de ellas está definido por la nomenclatura propuesta por el grupo EAGLES. Esta nomenclatura puede ser revisada en el anexo.

### 4.3.2 Clasificación Automática en WEKA

Teniendo las 4 versiones a estudiar en formato ARFF, se trabaja con cada una de ellas de manera independiente en la herramienta WEKA. Para ello, se debe trabajar con la aplicación “explorer” disponible en la ventana principal del programa. Una vez abierta la aplicación, se muestra una ventana con las distintas opciones disponibles. Las que interesan para el desarrollo de la investigación son la pestaña “Preprocess” y “Classify”.

En la pestaña “preprocess” se carga el archivo de la versión del corpus a analizar y se configuran los filtros necesarios para el funcionamiento, como lo es transformar el atributo correspondiente al titular desde un string a un vector de palabras (filtro: StringToWordVector), que permitirá estudiar el titular no como un dato único, sino que como un conjunto de datos compuestos por las palabras que lo componen.

Una vez cargada y configurada la versión del corpus a estudiar, en la pestaña “classify” se selecciona el algoritmo de clasificación que se utilizará. Dado que se aplicaron filtros en la configuración del corpus, es necesario seleccionar el algoritmo “filtered classifier”, el cual permite seleccionar el algoritmo de clasificación, explicándole a este que se aplicó cierto filtro como lo es el “StringToWordVector”.

Para realizar la evaluación de la clasificación, se trabaja con el método K-fold cross-validation (Validación cruzada) implementado en WEKA, debido a las mejoras que entrega el procedimiento en el cálculo de resultados. Luego se selecciona el atributo por el cual se desea realizar la clasificación, que para este caso será el de “affect” y finalmente se lanza la aplicación.

Una vez finalizada, WEKA entrega los datos correspondientes a los resultados con 2 estructuras: una donde se entrega un resumen de las clasificaciones correctas e incorrectas y valores de error absoluto; la otra correspondiente a una matriz de confusión en la cual se tabula cada asignación de afecto realizada de acuerdo a lo que se esperaba y lo que se etiquetó finalmente. Un ejemplo de ello es lo que se aprecia en la figura 13.

Correctly Classified Instances	583	28.7051 %							
Incorrectly Classified Instances	1448	71.2949 %							
Kappa statistic	0.1018								
Mean absolute error	0.1651								
Root mean squared error	0.3368								
Relative absolute error	91.0531 %								
Root relative squared error	111.8997 %								
Total Number of Instances	2031								
=== Detailed Accuracy By Class ===									
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,041	0,019	0,051	0,041	0,045	0,025	0,568	0,035	alegría
	0,406	0,342	0,303	0,406	0,347	0,059	0,542	0,295	confianza
	0,000	0,013	0,000	0,000	0,000	-0,016	0,563	0,024	miedo
	0,077	0,047	0,120	0,077	0,094	0,037	0,558	0,092	sorpresa
	0,366	0,033	0,512	0,366	0,427	0,389	0,697	0,378	tristeza
	0,104	0,035	0,185	0,104	0,133	0,091	0,530	0,109	aversion
	0,021	0,006	0,077	0,021	0,033	0,028	0,535	0,025	ira
	0,283	0,198	0,288	0,283	0,286	0,086	0,559	0,261	anticipación
	0,331	0,209	0,296	0,331	0,312	0,117	0,589	0,273	ninguna
Weighted Avg.	0,287	0,189	0,276	0,287	0,277	0,103	0,570	0,243	

Figura 13: Ejemplo de resumen de clasificación provisto por WEKA

Una vez obtenidos los resultados correspondientes, se procede a realizar el análisis. Los valores de las métricas pueden obtenerse directamente desde la matriz de confusión, asociando los valores promedios para Accuracy (TP Rate), Precision, Recall y F-Measure. El detalle de los resultados obtenidos en el experimento se presenta en el punto 5 de esta investigación.

En resumen, considerando la etapa de pre procesamiento realizada con el programa en JAVA en la cual se preparan estas 4 versiones del corpus, además de la etapa clasificación automática realizada en WEKA, el proceso para cada uno de los algoritmos de clasificación automática se realiza de acuerdo a lo presentado en la figura:

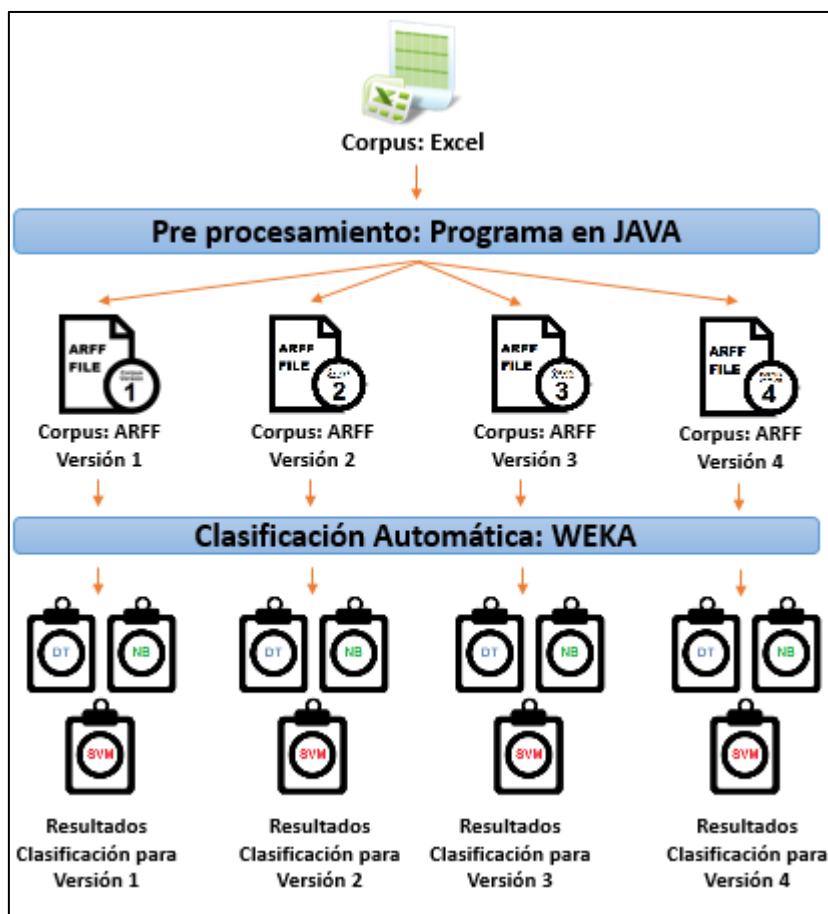


Figura 14: Proceso de clasificación automática para las 4 versiones del corpus

Las 4 versiones del programa en JAVA para generar los archivos ARFF correspondientes quedan disponibles<sup>3</sup>.

3

Version1: <http://dsi.face.ubiobio.cl/somos/newscorpus/applications/SinPosSinStopwords.java>  
 Version2: <http://dsi.face.ubiobio.cl/somos/newscorpus/applications/SinPosConStopwords.java>  
 Version3: <http://dsi.face.ubiobio.cl/somos/newscorpus/applications/ConPosSinStopwords.java>  
 Version4: <http://dsi.face.ubiobio.cl/somos/newscorpus/applications/ConPosConStopwords.java>

## 5.- RESULTADOS PRELIMINARES Y DISCUSIÓN

En esta sección se presentan los resultados de la aplicación de 3 técnicas de aprendizaje automático (Árboles de decisión, Naive Bayes y Support Vector Machine) sobre las 4 versiones del corpus utilizado en el experimento, procesando en cada uno de ellos un total de 2031 titulares. El detalle de la aplicación de los clasificadores sobre las distintas versiones del corpus se presenta en el anexo 4.

En primer lugar, la tabla 10 muestra los resultados obtenidos en términos de los titulares clasificados correcta e incorrectamente.

<b>Rendimiento del clasificador: Porcentaje de clasificación</b>						
<b>Corpus</b>	<b>Decisión Tree</b>		<b>Naive Bayes</b>		<b>SVM</b>	
	<b>Correctos</b>	<b>Incorrectos</b>	<b>Correctos</b>	<b>Incorrectos</b>	<b>Correctos</b>	<b>Incorrectos</b>
C1- Original	28%	71%	32%	67%	<b>39%</b>	60%
C2- Filtro stopword	31%	68%	33%	66%	<b>38%</b>	61%
C3- POS	29%	70%	31%	68%	<b>39%</b>	60%
C4- Filtro stopword +POS	29%	70%	34%	65%	<b>38%</b>	61%

Tabla 10: Rendimiento de los clasificadores

Respecto del rendimiento del clasificador, se aprecia que SVM se comporta mejor en todos los corpus de prueba, con resultados correctos entre 38% y 39%, seguido por NB y DT respectivamente.

Se puede observar que el rendimiento de los clasificadores DT y NB aplicados sobre un corpus procesado con stopwords (C2 y C4) aumentó levemente respecto al rendimiento obtenido cuando se aplicó al corpus original (C1). Este aumento se observó entre 1% - 2% aunque no alcanza el mejor rendimiento obtenido por SVM en el experimento. Cabe mencionar que para SVM la eliminación de stopwords significó una leve disminución en su rendimiento (-1%).

Respecto de la aplicación de etiquetado POS al Corpus C3 y C4, el clasificador basado en DT muestra una mejora del 1% respecto del Corpus original (C1). Por su parte el clasificador basado en NB disminuyó en un 1% cuando se aplicó sobre C3 respecto a su aplicación sobre C1. Este mismo clasificador mostró un aumento de 2% cuando se aplicó sobre el corpus C4. Para el caso de SVM, el rendimiento no muestra variación al ser aplicado al Corpus C3, pero si disminuye al ser aplicado sobre el corpus C4 en un 1%.

Respecto del rendimiento general de los clasificadores en este experimento y su relación con la cantidad de categorías afectivas, versus trabajos que realizan AS presentados en la sección 3.4, claramente se observa una gran disminución en los porcentajes. Esto se debe a la dificultad que tiene el clasificador para asignar correctamente la categoría afectiva, debido al aumento de categorías afectivas en el AA (9 categorías en este experimento) respecto al AS (3 categorías).

Con respecto al rendimiento general de los clasificadores en este experimento respecto a trabajos que realizan AA utilizando el enfoque basado en diccionarios presentados en la sección 3.4, también se observa una disminución en los porcentajes, pero un poco más leve. Esto también está relacionado a la cantidad de categorías afectivas con las que se trabaja, pero también a las características propias del corpus como idioma, el dominio de trabajo, cantidad de documentos, etc.

Para complementar este análisis, las Tablas 11, 12, 13 y 14 muestran los resultados obtenidos por los clasificadores para las métricas Accuracy, Precisión, Recall y F-Measure.

<b>Métrica: Accuracy</b>			
<b>Corpus</b>	<b>Decision Tree</b>	<b>Naive Bayes</b>	<b>SVM</b>
C1- Original	0.28	0.32	<b>0.39</b>
C2- Filtro stopwords	0.31	0.33	<b>0.38</b>
C3- POS	0.29	0.31	<b>0.39</b>
C4- Filtro stopwords +POS	0.29	0.34	<b>0.38</b>

Tabla 11: Resultados obtenidos para la métrica Accuracy

<b>Métrica: Precision</b>			
<b>Corpus</b>	<b>Decision Tree</b>	<b>Naive Bayes</b>	<b>SVM</b>
C1- Original	0.276	0.327	<b>0.400</b>
C2- Filtro stopwords	0.356	0.358	<b>0.406</b>
C3- POS	0.284	0.322	<b>0.410</b>
C4- Filtro stopwords +POS	0.319	0.369	<b>0.410</b>

Tabla 12: Resultados obtenidos para la métrica Precision

<b>Métrica: Recall</b>			
<b>Corpus</b>	<b>Decision Tree</b>	<b>Naive Bayes</b>	<b>SVM</b>
C1- Original	0.287	0.325	<b>0.395</b>
C2- Filtro stopwords	0.314	0.338	<b>0.387</b>
C3- POS	0.294	0.318	<b>0.399</b>
C4- Filtro stopwords +POS	0.297	0.346	<b>0.389</b>

Tabla 13: Resultados obtenidos para la métrica Recall

<b>Métrica: F-Measure</b>			
<b>Corpus</b>	<b>Decision Tree</b>	<b>Naive Bayes</b>	<b>SVM</b>
C1- Original	0.277	0.324	<b>0.377</b>
C2- Filtro stopwords	0.260	0.322	<b>0.366</b>
C3- POS	0.282	0.318	<b>0.381</b>
C4- Filtro stopwords +POS	0.238	0.329	<b>0.366</b>

Tabla 14: Resultados obtenidos para la métrica F-Measure

Analizando los resultados obtenidos de las 4 métricas, para todos los corpus del experimento se observa que los mejores resultados son obtenidos utilizando el clasificador SVM, seguido por NB y DT respectivamente. Puede apreciarse que SVM obtiene los más altos valores para las cuatro métricas cuando se aplica sobre un corpus con etiquetas POS. Salvo para la métrica F-Measure, los peores valores se obtienen para las restantes 3 métricas cuando se aplica DT sobre el corpus original.

En resumen, luego de analizar los resultados de este experimento podemos distinguir los siguientes aspectos:

- El mejor rendimiento se obtiene al aplicar un clasificador basado en SVM.
- Los mejores resultados al aplicar DT se obtiene cuando el corpus ha sido procesado con stopwords.
- Los mejores resultados al aplicar NB se obtienen sobre un corpus procesado con Stop Word y etiquetado POS conjuntamente.

No es posible asegurar que para textos en español, la aplicación de procesos de stopwords y/o etiquetado POS aumente en todos los casos el rendimiento de un clasificador, puesto que se debe considerar la naturaleza del algoritmo de clasificación.

Distintos factores pueden afectar el rendimiento de la clasificación. Uno de estos elementos se refiere al balanceo del corpus, es decir, que el entrenamiento del clasificador se realice con casos equitativamente representativos de las distintas clases. En este caso no se pudo contar con un corpus balanceado debido al costo del proceso de etiquetado que esto implica.

Otro elemento que puede afectar los resultados es el uso de reconocedores de entidades lo que podría aportar a distinguir palabras que podrían no aportar al contexto emocional del titular, pero no se consideró su utilización debido a las características de diccionario en su funcionamiento.

## 6.- CONCLUSIONES Y TRABAJOS FUTUROS

En esta sección se dan a conocer los aportes de esta investigación, además de contrastar el trabajo realizado con los objetivos propuestos en la sección 2. Se exponen las conclusiones finales y se presentan los trabajos futuros que es posible desarrollar a partir de esta investigación.

### 6.1 Aportes

El principal aporte de este trabajo es determinar cuál de las técnicas de aprendizaje automático es la mejor alternativa a utilizar al momento de realizar un proceso de análisis de afecto sobre un corpus en español. Realizar este proceso sobre un corpus en español es importante, ya que no se reportan trabajos comparativos utilizando estas técnicas en procesos de análisis de afecto sobre textos en este idioma, lo que permitiría sentar un precedente a investigadores en futuros trabajos que deseen realizar estos procesos.

Se entrega un análisis actualizado del estado del arte sobre esta materia. Desde el punto de vista práctico, se presenta un análisis de herramientas necesarias en el desarrollo de estos trabajos, adaptando los procedimientos existentes para su utilización y elaborando un programa de código abierto que permite realizar parte del proceso.

La realización de este trabajo trajo consigo la elaboración de un corpus de titulares de periódicos en idioma español junto a su correspondiente emoción de acuerdo a la categorización de Plutchik. Esto fue necesario ya que no existían corpus en español que trabajaran con dicha categorización. El corpus queda disponible para su utilización por la comunidad.

Finalmente, gracias a este trabajo es posible generar un punto de discusión para realizar trabajos futuros de investigación, pudiendo replicar parte de los experimentos utilizando otras técnicas ya existentes o incorporando nuevas técnicas y herramientas que permitan mejorar los resultados.

### 6.2 Contraste de lo realizado con los objetivos planteados

Este trabajo permitió comparar técnicas de aprendizaje automático aplicados al análisis de afectos sobre textos escritos en español. A continuación se presentan los objetivos específicos y se detallan sus cumplimientos:

- **OE1: Estudiar las técnicas de aprendizaje automático más utilizadas en los procesos de análisis de afecto**

El trabajo realizado presenta una completa investigación desde un análisis teórico de los conceptos utilizados en análisis de afecto hasta el análisis de herramientas necesarias para aplicar estas técnicas.

- **OE2: Aplicar un proceso de análisis de afecto sobre un corpus, tomando en cuenta las técnicas de aprendizaje automático más utilizadas estudiando la incidencia de técnicas de NLP, como stopwords y etiquetado POS**

El trabajo realizado utiliza el aspecto teórico y las herramientas mencionadas para definir una metodología que permite realizar el proceso de análisis de afecto, pudiendo comparar las técnicas de aprendizaje automático. Finalmente, se aplica la metodología propuesta por medio de una etapa experimental que entrega resultados objetivos para analizar.

- **OE3: Realizar un análisis comparativo de las técnicas de aprendizaje automático de acuerdo a métricas establecidas de rendimiento**

El trabajo realizado entrega tablas de contraste de resultados y análisis de estos, a raíz de las métricas establecidas, siendo importantes para cumplir con los objetivos y elaborar las conclusiones finales.

### 6.3 Conclusiones

Los resultados del experimento realizado, muestran que los mejores resultados de un proceso de Análisis de Afectos utilizando técnicas de aprendizaje automático sobre titulares de periódicos en español se obtienen aplicando un clasificador basado en SVM. NB resultó ser la segunda mejor técnica mientras que DT obtuvo los más bajos resultados. Este comportamiento de los clasificadores mantiene la tendencia con respecto al rendimiento que han tenido al aplicarlos en procesos de Análisis de Sentimientos sobre corpus en inglés, según lo presentado en el apartado 3.4.

Una de las características que afecta al realizar la clasificación es el balanceo del corpus. Esto está directamente relacionado a su proceso de elaboración, debido a lo disperso que puede ser el etiquetado de titulares considerando las características del dominio y además, la cantidad de evaluadores que participan. Esto puede mejorar realizando un etiquetado con mayor cantidad de evaluadores, pero teniendo en cuenta el costo de tiempo y recursos de procesamiento de datos, además de la influencia en el valor de Kappa dada las características de este índice.

Por otra parte, debido a que el corpus utilizado corresponde a un dominio específico y de tamaño moderado, no es posible concluir que el uso de stopword o etiquetado POS genera los mejores resultados en todos los casos, y su incidencia debe analizarse para casos específicos de análisis de un clasificador. Esto indica que distintas configuraciones de herramientas de pre procesamiento entregan distintos resultados de clasificación, pudiendo depender del algoritmo de clasificación utilizado y del corpus de trabajo.

Si bien, los valores obtenidos con respecto al mejor rendimiento del clasificador (38-39%) y sus métricas asociadas (0,36 a 0,41) pueden parecer bajos respecto a trabajos de Análisis de Sentimientos, esto puede explicarse a que en Análisis de Afectos existe una mayor cantidad de categorías que se utilizan para clasificar. Los niveles de rendimiento pueden resultar tan bajos como el nivel de acuerdo obtenido por evaluadores humanos para una cantidad considerable de categorías.

## **6.4 Trabajos Futuros**

Para trabajos futuros, se podría considerar una forma de mejorar el nivel de detalle del clasificador, aplicando un proceso más detallado de procesamiento lingüístico, como por ejemplo la inclusión de un reconocedor de entidades. Por otro lado, se podría considerar realizar un estudio sobre la configuración de herramientas de pre-procesamiento que entrega el mejor rendimiento para cada uno de los clasificadores. Finalmente, podría considerarse la realización de una aplicación, basada en reglas que se puedan obtener de la clasificación automática realizada en esta investigación, con el fin de poder identificar automáticamente la emoción asociada a un titular ingresado, considerando que ese titular se etiquetará con un 38-39% de efectividad.



## REFERENCIAS

- [1] K. R. Scherer, "Emotion as a Multicomponent Process," *Review of Personality and Social Psychology*, vol. 5, pp. 37–63, 1984.
- [2] B. Liu, "Sentiment Analysis and Subjectivity," *Handb. Nat. Lang. Process.*, no. 1, pp. 1–38, 2010.
- [3] T. Wilson, J. Wiebe, and P. Hoffman, "Recognizing contextual polarity in phrase level sentiment analysis," *Acl*, vol. 7, no. 5, pp. 12–21, 2005.
- [4] P. Subasic and A. Huettner, "Affect analysis of text using fuzzy semantic typing," *IEEE Trans. Fuzzy Syst.*, vol. 9, no. 4, pp. 483–496, 2001.
- [5] J. C. De Albornoz, L. Plaza, and P. Gervás, "A Hybrid Approach to Emotional Sentence Polarity and Intensity Classification," *Comput. Linguist.*, no. July, pp. 153–161, 2010.
- [6] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," *Proc. Conf. Empir. Methods Nat. Lang. Process.*, pp. 79–86, 2002.
- [7] L. Dubiau and J. Ale, "Análisis de sentimientos sobre un corpus en español: experimentación con un caso de estudio," *14th Argentine Symp. ...*, 2013.
- [8] G. G. Chowdhury, "Natural language processing," *Annu. Rev. Inf. Sci. Technol.*, vol. 37, no. 1, pp. 51–89, 2005.
- [9] J. Klein, Y. Moon, and R. W. Picard, "This computer responds to user frustration:," *Interact. Comput.*, vol. 14, no. 2, pp. 119–140, 2002.
- [10] A. Metallinou, A. Katsamanis, M. Wollmer, F. Eyben, B. Schuller, and S. Narayanan, "Context-sensitive learning for enhanced audiovisual emotion classification (Extended abstract)," in *2015 International Conference on Affective Computing and Intelligent Interaction, ACII 2015*, 2015, pp. 463–469.
- [11] A. A. Kapur, N. Virji-babul, G. Tzanetakis, and P. Driessen, "Gesture-Based Affective Computing on Motion Capture Data," *Affect. Comput. Intell. Interact.*, pp. 1–7, 2005.
- [12] M. C. Lin, M. L. Hummert, and J. Harwood, "Representation of age identities in on-line discourse," *J. Aging Stud.*, vol. 18, no. 3, pp. 261–274, 2004.
- [13] Lu Ye, Rui-Feng Xu, and Jun Xu, "Emotion prediction of news articles from reader's perspective based on multi-label classification," in *2012 International Conference on Machine Learning and Cybernetics*, 2012, pp. 2019–2024.
- [14] A. L. Samuel, "Some Studies in Machine Learning Using the Game of Checkers," *IBM J. Res. Dev.*, vol. 3, no. 3, pp. 210–229, 1959.
- [15] R. W. Picard, "Affective Computing," *MIT Press*, no. 321, pp. 1–16, 1995.
- [16] R. Plutchik, "Emotions and Life: Perspectives for Psychology, Biology, and Evolution," *American Journal of Psychiatry*, vol. 162, pp. 409–409, 2005.
- [17] P. Ekman, "Cross-cultural studies of facial expression," *Darwin facial Expr. A century Res. Rev.*, pp. 169–222, 1973.
- [18] a Ng and M. Jordan, "On generative vs. discriminative classifiers: A comparison of logistic regression and naive bayes," *Proc. Adv. Neural Inf. Process.*, vol. 28, no. 3, pp. 169–187, 2002.

- [19] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. Witten, “The WEKA data mining software: An update,” *SIGKDD Explor.*, vol. 11, no. 1, pp. 10–18, 2009.
- [20] H. Cunningham, D. Maynard, K. Bontcheva, and V. Tablan, “GATE - A Framework and Graphical Development Environment for Robust NLP Tools and Applications,” *Proc. 40th Annu. Meet. Assoc. Comput. Linguist. - ACL '02*, no. July, p. 168, 2001.
- [21] H. Cunningham, D. Maynard, and K. Bontcheva, *Text Processing with GATE (Version 6)*. 2011.
- [22] C. D. Manning, J. Bauer, J. Finkel, S. J. Bethard, M. Surdeanu, and D. McClosky, “The Stanford CoreNLP Natural Language Processing Toolkit,” *Proc. 52nd Annu. Meet. Assoc. Comput. Linguist. Syst. Demonstr.*, pp. 55–60, 2014.
- [23] K. Khan, B. B. Baharudin, A. Khan, and Fazal-E-Malik, “Mining opinion from text documents: A survey,” in *2009 3rd IEEE International Conference on Digital Ecosystems and Technologies, DEST '09*, 2009, pp. 217–222.
- [24] M. S. Neethu and R. Rajasree, “Sentiment analysis in twitter using machine learning techniques,” *2013 Fourth Int. Conf. Comput. Commun. Netw. Technol.*, pp. 1–5, 2013.
- [25] N. Yussupova, D. Bogdanova, and M. Boyko, “Applying of Sentiment Analysis for Texts in Russian Based on Machine Learning Approach,” *IMMM 2012 Second Int. Conf. Adv. Inf. Min. Manag.*, no. c, pp. 8–14, 2012.
- [26] E. Jang, B. Park, and S. Kim, “Emotion Classification Based on Bio-Signals Emotion Recognition Using Machine Learning Algorithms,” *Int. Conf. Inf. Sci. Electron. Electr. Eng.*, pp. 1373–1376, 2014.
- [27] J. Zhao, K. Liu, and G. Wang, “Adding redundant features for CRFs-based sentence sentiment classification,” *Comput. Linguist.*, no. October, pp. 117–126, 2008.
- [28] B. Thomas, K. A. Dhanya, and P. Vinod, “Synthesized feature space for multiclass emotion classification,” in *1st International Conference on Networks and Soft Computing, ICNSC 2014 - Proceedings*, 2014, pp. 188–192.
- [29] C. Strapparava and R. Mihalcea, “Semeval-2007 task 14: Affective text,” *Proc. of SemEval-2007*, no. June, pp. 70–74, 2007.
- [30] S. Wen and X. Wan, “Emotion Classification in Microblog Texts Using Class Sequential Rules,” *Proc. Twenty-Eighth AAAI Conf. Artif. Intell.*, pp. 187–193, 2014.
- [31] S. M. Mohammad and P. D. Turney, “Crowdsourcing a Word – Emotion Association Lexicon,” *Comput. Intell.*, vol. 59, no. 0, pp. 1–24, 2011.

## ANEXOS

### Anexo 1: Matriz de confusión para cálculo de Kappa.

- Cálculo para Evaluador 1 vs Evaluador 2:

		RATER 2								
RATER 1		alegría	confianza	miedo	sorpres	tristeza	aversión	ira	anticipaci	ninguna
	alegría	3	6	0	3	0	0	0	2	4
	confianza	29	202	5	49	17	23	17	93	100
	miedo	0	1	1	3	7	2	0	1	0
	sorpres	3	11	1	59	4	12	2	10	25
	tristeza	0	3	6	17	74	9	3	14	10
	aversión	0	8	9	19	6	57	21	7	19
	ira	1	2	5	3	10	5	20	3	1
	anticipaci	8	57	23	44	29	30	12	386	160
	ninguna	10	14	0	23	4	7	3	22	99

Tabla 15: Anexo Matriz Confusión Kappa Ev1 vs Ev2

KAPPA obtenido de GRAPH PAD SOFTWARE	
Acuerdos observados	901 (46,02% de las observaciones)
Acuerdos esperados	366,2 (18,7% de las observaciones)
KAPPA	0,336
SE de KAPPA	0,013
Intervalo de confianza del 95%	desde 0,31 a 0,362
Fuerza de acuerdo	Lejano
KAPPA PONDERADO	0,363

Tabla 16: Anexo Cálculo de Kappa entre Ev1 y Ev2

- Cálculo para Evaluador 1 vs Evaluador 3:

		RATER 3								
RATER 1		alegría	confianza	miedo	sorpres	tristeza	aversión	ira	anticipaci	ninguna
	alegría	4	6	0	0	0	0	0	3	4
	confianza	10	380	2	19	16	3	10	7	90
	miedo	0	1	6	2	5	0	0	1	0
	sorpres	0	13	2	59	7	6	2	3	36
	tristeza	0	0	3	6	114	3	1	2	7
	aversión	1	3	12	13	15	79	13	3	7
	ira	1	3	6	5	12	7	13	1	2
	anticipaci	16	157	43	61	41	45	11	96	274
	ninguna	2	21	1	7	2	3	0	4	143

Tabla 17: Anexo Matriz Confusión Kappa Ev1 vs Ev3

Acuerdos observados	894 (45,71% de las observaciones)
Acuerdos esperados	297,7 (15,22% de las observaciones)
KAPPA	0,36
SE de KAPPA	0,012
Intervalo de confianza del 95%	desde 0,336 a 0,383
Fuerza de acuerdo	Lejano
KAPPA PONDERADO	0,413

Tabla 18: Anexo Cálculo de Kappa entre Ev1 y Ev3

- Cálculo para Evaluador 2 vs Evaluador 3:

		RATER 3								
		alegría	confianza	miedo	sorpresa	tristeza	aversión	ira	anticipación	ninguna
RATER 2	alegría	7	28	0	3	0	0	1	6	8
	confianza	9	203	2	17	13	7	4	7	42
	miedo	0	5	33	0	4	2	3	3	0
	sorpresa	3	49	2	86	18	17	2	10	35
	tristeza	0	11	3	4	123	2	0	4	4
	aversión	1	22	7	17	17	54	6	11	10
	ira	1	12	4	4	4	16	25	8	4
	anticipación	13	193	23	34	28	30	9	60	144
	ninguna	0	59	1	8	5	18	0	12	320

Tabla 19: Anexo Matriz Confusión Kappa Ev2 vs Ev3

Acuerdos observados	911 (46,48% de las observaciones)
Acuerdos esperados	297,2 (15,16% de las observaciones)
KAPPA	0,363
SE de KAPPA	0,013
Intervalo de confianza del 95%	desde 0,344 a 0,394
Fuerza de acuerdo	Lejano
KAPPA PONDERADO	0,394

Tabla 20: Anexo Cálculo de Kappa entre Ev2 y Ev3

Para obtener el valor final de Kappa, es necesario promediar el valor de Kappa ponderado obtenido de los cálculos entre todos los evaluadores. Para este caso, sería:

$$Kappa = \frac{0,363+0,413+0,394}{3} = 0,39 \quad (7)$$

## Anexo 2: Estructura de las etiquetas POS elaboradas por el grupo EAGLES

- Esta tabla es un resumen de la nomenclatura de etiquetas EAGLES

Etiquetas			
Código	Descripción	Código	Descripción
<b>Adjetivos</b>		<b>Intersecciones</b>	
AO	Adjetivo Ordinal	I	Intersección
AQ	Adjetivo Calificativo	<b>Preposiciones</b>	
<b>Adverbios</b>		SP	Preposición, Adposición
RG	Adverbio General	<b>Puntuación</b>	
RN	Adverbio Negativo	Faa	Inicia exclamación (!)
<b>Determinantes</b>		Fat	Cierra exclamación (!)
DD	Determinante Demostrativo	Fc	Coma (,)
DP	Determinante Posesivo	Fca	Inicia corchetes ([)
DT	Determinante Interrogativo	Fct	Cierra corchetes (])
DE	Determinante Exclamativo	Fd	Dos puntos (:)
DI	Determinante Indefinido	Fe	Comillas ("")
DA	Determinante Artículo	Fg	Guión Alto (-)
<b>Nombres</b>		Fh	Slash (/)
NC	Nombre Común	Fia	Inicia interrogación (¿)
NP	Nombre Propio	Fit	Cierra Interrogación (?)
<b>Verbos</b>		Fla	Inicia Llaves ({)
VMI	Verbo Principal, Indicativo	Flt	Cierra Llaves (})
VMS	Verbo Principal, Subjuntivo	Fp	Punto (.)
VMM	Verbo Principal, Imperativo	Fpa	Inicia paréntesis ("(")
VMN	Verbo Principal, Infinitivo	Fpt	Cierra paréntesis (")")
VMG	Verbo Principal, Gerundio	Fra	Inicia comillas angulares («)
VMP	Verbo Principal, Participativo	Frc	Cierra comillas angulares (»)
VAI	Verbo Auxiliar, Indicativo	Fs	Tres puntos (...)
VAS	Verbo Auxiliar, Subjuntivo	Ft	Porcentaje (%)
VAM	Verbo Auxiliar, Imperativo	Fx	Punto y coma (;)
VAN	Verbo Auxiliar, Infinitivo	Fz	Otros signos (+, -, *, =)
VAG	Verbo Auxiliar, Gerundio	<b>Números</b>	
VAP	Verbo Auxiliar, Participativo	Z	Cifra
VSI	Verbo Semiauxiliar, Indicativo	<b>Fecha/Hora</b>	
VSM	Verbo Semiauxiliar, Subjuntivo	W	Fecha y hora
VSN	Verbo Semiauxiliar, Infinitivo	<b>Desconocido</b>	
VSG	Verbo Semiauxiliar, Gerundio	X	Desconocido
VSP	Verbo Semiauxiliar, Participativo		
<b>Pronombres</b>			
PP	Pronombre Personal		
PD	Pronombre Demostrativo		
PX	Pronombre Posesivo		
PI	Pronombre Indefinido		
PT	Pronombre Interrogativo		
PR	Pronombre Relativo		
PE	Pronombre Exclamativo		
<b>Conjunciones</b>			
CC	Conjunción Coordinada		
CS	Conjunción Subordinada		

Tabla 21: Anexo Resumen de etiquetas EAGLES

A continuación, se presenta el detalle de cada una de las formas lingüísticas presentes en la nomenclatura EAGLES. Estas están representadas con la siguiente estructura:

- Adjetivos

Adjetivos			
Posición	Atributo	Valor	Código
1	Categoría	Adjetivo	A
2	Tipo	Calificativo	Q
		Ordinal	O
3	Grado	Aumentativo	A
		Diminutivo	D
		Comparativo	C
		Superlativo	S
4	Género	Femenino	M
		Masculino	F
		Común	C
5	Número	Singular	S
		Plural	P
		Invariable	N
6	Función	-	0
		Participio	P

Tabla 22: Anexo Detalle de adjetivos etiquetas EAGLES

- Adverbios

Adverbio			
Posición	Atributo	Valor	Código
1	Categoría	Adverbio	R
2	Tipo	General	G
		Negativo	N

Tabla 23: Anexo Detalle de adverbios etiquetas EAGLES

- Determinante

Determinantes			
Posición	Atributo	Valor	Código
1	Categoría	Determinante	D
2	Tipo	Demostrativo	D
		Posesivo	P
		Interrogativo	T
		Exclamativo	E
		Indefinido	I
		Artículo	A
3	Persona	Primera	1
		Segunda	2
		Tercera	3
4	Género	Masculino	M
		Femenino	F
		Común	C
		Neutro	N
5	Número	Singular	S
		Plural	P
		Invariable	N
6	Poseedor	Singular	S
		Plural	P

Tabla 24: Anexo Detalle de determinantes etiquetas EAGLES

- Nombre

Nombres			
Posición	Atributo	Valor	Código
1	Categoría	Nombre	N
2	Tipo	Común	C
		Propio	P
3	Género	Masculino	M
		Femenino	F
		Común	C
4	Número	Singular	S
		Plural	P
		Invariable	N
5	Clasificación Semántica	Persona	SP
		Lugar	G0
		Organización	O0
		Otros	V0
6	Grado	Aumentativo	A
		Diminutivo	D

Tabla 25: Anexo Detalle de nombre etiquetas EAGLES

- Verbo:

Verbos			
Posición	Atributo	Valor	Código
1	Categoría	Verbo	V
2	Tipo	Principal	M
		Auxiliar	A
		Semiauxiliar	S
3	Modo	Indicativo	I
		Subjuntivo	S
		Imperativo	M
		Infinitivo	N
		Gerundio	G
		Participio	P
4	Tiempo	Presente	P
		Imperfecto	I
		Futuro	F
		Pasado	S
		Condicional	C
		-	0
5	Persona	Primera	1
		Segunda	2
		Tercera	3
6	Número	Singular	S
		Plural	P
7	Género	Masculino	M
		Femenino	F

Tabla 26: Anexo Detalle de verbos etiquetas EAGLES



- Pronombres:

Pronombres			
Posición	Atributo	Valor	Código
1	Categoría	Pronombre	P
2	Tipo	Personal	P
		Demostrativo	D
		Posesivo	X
		Indefinido	I
		Interrogativo	T
		Relativo	R
		Exclamativo	E
3	Persona	Primera	1
		Segunda	2
		Tercera	3
4	Género	Masculino	M
		Femenino	F
		Común	C
		Neutro	N
5	Número	Singular	S
		Plural	P
		Impersonal Invariable	N
6	Caso	Nominativo	N
		Acusativo	A
		Dativo	D
		Oblicuo	O
7	Poseedor	Singular	S
		Plural	P
8	Politeness	Polite	P

Tabla 27: Anexo Detalle de pronombres etiquetas EAGLES

- Conjunción:

Conjunciones			
Posición	Atributo	Valor	Código
1	Categoría	Conjunción	C
2	Tipo	Coordinada	C
		Subordinada	S

Tabla 28: Anexo Detalle de conjunciones etiquetas EAGLES

- Interjección:

Interjecciones			
Posición	Atributo	Valor	Código
1	Categoría	Interjección	I

Tabla 29: Anexo Detalle de interjecciones etiquetas EAGLES

- Preposición:

Proposiciones			
Posición	Atributo	Valor	Código
1	Categoría	Adposición	S
2	Tipo	Preposición	P
3	Forma	Simple	S
		Contraída	C
4	Género	Masculino	M
5	Número	Singular	S

Tabla 30: Anexo Detalle de preposiciones etiquetas EAGLES

- Puntuación:

Puntuación			
Posición	Atributo	Valor	Código
1	Categoría	Puntuación	F

Tabla 31: Anexo Detalle de puntuación etiquetas EAGLES

- Numeración:

Numerales			
Posición	Atributo	Valor	Código
1	Categoría	Cifra	Z
2	Tipo	Partitivo	d
		Moneda	m
		Porcentaje	p
		Unidad	u

Tabla 32: Anexo Detalle de numeración etiquetas EAGLES

- Fecha y Hora:

Fecha y Hora			
Posición	Atributo	Valor	Código
1	Categoría	Fecha/Hora	W

Tabla 33: Anexo Detalle de fecha y hora etiquetas EAGLES

**Anexo 3:** Listado de StopWords utilizados en la limpieza:

de	pero	nos	otra	tu
la	sus	durante	él	tus
que	le	todos	tanto	ellas
el	ya	uno	esa	nosotras
en	o	les	estos	vosotros
y	este	ni	mucho	vosotras
a	sí	contra	quienes	os
los	porque	otros	nada	mío
del	esta	ese	muchos	mía
se	entre	eso	cual	míos
las	cuando	ante	poco	mías
por	muy	ellos	ella	tuyo
un	sin	e	estar	tuya
para	sobre	esto	estas	tuyos
con	también	mí	algunas	tuyas
no	me	antes	algo	suyo
una	hasta	algunos	nosotros	suya
su	hay	qué	mi	suyos
al	donde	unos	mis	suyas
lo	quien	yo	tú	nuestro
como	desde	otro	te	nuestra
más	todo	otras	ti	nuestros

nuestras	estabas	habéis	hubieras	serán
vuestro	estábamos	han	hubiéramos	sería
vuestra	estabais	haya	hubierais	serías
vuestros	estaban	hayas	hubieran	seríamos
vuestras	estuve	hayamos	hubiese	seríais
esos	estuviste	hayáis	hubieses	serían
esas	estuvo	hayan	hubiésemos	era
estoy	estuvimos	habré	hubieseis	eras
estás	estuvisteis	habrás	hubiesen	éramos
está	estuvieron	habrá	habiendo	erais
estamos	estuviera	habremos	habido	eran
estáis	estuvieras	habréis	habida	fui
están	estuviéramos	habrán	habidos	fuiste
esté	estuvierais	habría	habidas	fue
estés	estuvieran	habrías	soy	fuimos
estemos	estuviese	habríamos	eres	fuisteis
estéis	estuvieses	habríais	es	fueron
estén	estuviésemos	habrían	somos	fuera
estaré	estuvieseis	había	sois	fueras
estarás	estuviesen	habías	son	fuéramos
estará	estando	habíamos	sea	fuerais
estaremos	estado	habíais	seas	fueran
estaréis	estada	habían	seamos	fuese
estarán	estados	hube	seáis	fueses
estaría	estadas	hubiste	sean	fuésemos
estarías	estad	hubo	seré	fueseis
estaríamos	he	hubimos	serás	fuesen
estaríais	has	hubisteis	será	siendo
estarían	ha	hubieron	seremos	sido
estaba	hemos	hubiera	seréis	tengo

tienes	tendría	tuviera
tiene	tendrías	tuvieras
tenemos	tendríamos	tuviéramos
tenéis	tendríais	tuvierais
tienen	tendrían	tuvieran
tenga	tenía	tuviese
tengas	tenías	tuvieses
tengamos	teníamos	tuviésemos
tengáis	teníais	tuvieseis
tengan	tenían	tuviesen
tendré	tuve	teniendo
tendrás	tuviste	tenido
tendrá	tuvo	tenida
tendremos	tuvimos	tenidos
tendréis	tuvisteis	tenidas
tendrán	tuvieron	tened

#### Anexo 4: Resultados de clasificación entregados por WEKA:

- Corpus 1 – Clasificador DT

Correctly Classified Instances	583	28.7051 %							
Incorrectly Classified Instances	1448	71.2949 %							
Kappa statistic	0.1018								
Mean absolute error	0.1651								
Root mean squared error	0.3368								
Relative absolute error	91.0531 %								
Root relative squared error	111.8897 %								
Total Number of Instances	2031								
=== Detailed Accuracy By Class ===									
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,041	0,019	0,051	0,041	0,045	0,025	0,568	0,035	alegria
	0,406	0,342	0,303	0,406	0,347	0,059	0,542	0,295	confianza
	0,000	0,013	0,000	0,000	0,000	-0,016	0,563	0,024	miedo
	0,077	0,047	0,120	0,077	0,094	0,037	0,558	0,092	sorpresa
	0,366	0,033	0,512	0,366	0,427	0,389	0,697	0,378	tristeza
	0,104	0,035	0,185	0,104	0,133	0,091	0,530	0,109	aversion
	0,021	0,006	0,077	0,021	0,033	0,028	0,535	0,025	ira
	0,283	0,198	0,288	0,283	0,286	0,086	0,559	0,261	anticipacion
	0,331	0,209	0,296	0,331	0,312	0,117	0,589	0,273	ninguna
Weighted Avg.	0,287	0,189	0,276	0,287	0,277	0,103	0,570	0,243	

Figura 15: Anexo Resultados Clasificación WEKA - Corpus 1 Clasificador DT

- Corpus 1 – Clasificador NB

=== Summary ===									
Correctly Classified Instances	660	32.4963 %							
Incorrectly Classified Instances	1371	67.5037 %							
Kappa statistic	0.1627								
Mean absolute error	0.1616								
Root mean squared error	0.3056								
Relative absolute error	89.1443 %								
Root relative squared error	101.5304 %								
Total Number of Instances	2031								
=== Detailed Accuracy By Class ===									
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,143	0,023	0,132	0,143	0,137	0,115	0,597	0,048	alegria
	0,398	0,289	0,335	0,398	0,364	0,103	0,607	0,351	confianza
	0,175	0,017	0,175	0,175	0,175	0,158	0,723	0,120	miedo
	0,122	0,053	0,161	0,122	0,139	0,079	0,599	0,110	sorpresa
	0,451	0,029	0,594	0,451	0,513	0,479	0,808	0,543	tristeza
	0,264	0,047	0,299	0,264	0,280	0,230	0,706	0,254	aversion
	0,021	0,022	0,023	0,021	0,022	-0,001	0,558	0,032	ira
	0,313	0,207	0,300	0,313	0,306	0,104	0,600	0,324	anticipacion
	0,357	0,156	0,378	0,357	0,367	0,205	0,662	0,350	ninguna
Weighted Avg.	0,325	0,167	0,327	0,325	0,324	0,163	0,641	0,317	

Figura 16: Anexo Resultados Clasificación WEKA - Corpus 1 Clasificador NB

- Corpus 1 – Clasificador SVM

Correctly Classified Instances	803	39.5372 %							
Incorrectly Classified Instances	1228	60.4628 %							
Kappa statistic	0.2271								
Mean absolute error	0.1827								
Root mean squared error	0.2984								
Relative absolute error	100.7544 %								
Root relative squared error	99.1263 %								
Total Number of Instances	2031								
=== Detailed Accuracy By Class ===									
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,061	0,003	0,375	0,061	0,105	0,144	0,598	0,055	alegria
	0,560	0,324	0,388	0,560	0,458	0,215	0,647	0,363	confianza
	0,100	0,002	0,571	0,100	0,170	0,234	0,661	0,090	miedo
	0,090	0,023	0,246	0,090	0,131	0,108	0,571	0,114	sorpresa
	0,440	0,017	0,706	0,440	0,542	0,526	0,809	0,450	tristeza
	0,236	0,021	0,459	0,236	0,312	0,294	0,713	0,231	aversion
	0,021	0,002	0,200	0,021	0,038	0,058	0,565	0,034	ira
	0,373	0,192	0,355	0,373	0,363	0,178	0,638	0,298	anticipacion
	0,465	0,197	0,385	0,465	0,421	0,251	0,684	0,321	ninguna
Weighted Avg.	0,395	0,175	0,400	0,395	0,377	0,233	0,663	0,298	

Figura 17: Anexo Resultados Clasificación WEKA - Corpus 1 Clasificador SVM

- Corpus 2 – Clasificador DT

Correctly Classified Instances	637	31.3639 %							
Incorrectly Classified Instances	1394	68.6361 %							
Kappa statistic	0.0878								
Mean absolute error	0.1705								
Root mean squared error	0.3002								
Relative absolute error	94.0269 %								
Root relative squared error	99.7353 %								
Total Number of Instances	2031								
=== Detailed Accuracy By Class ===									
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,020	0,001	0,333	0,020	0,038	0,078	0,568	0,039	alegria
	0,769	0,717	0,282	0,769	0,413	0,051	0,537	0,289	confianza
	0,000	0,001	0,000	0,000	0,000	-0,003	0,586	0,026	miedo
	0,013	0,005	0,182	0,013	0,024	0,029	0,584	0,096	sorpresa
	0,286	0,006	0,820	0,286	0,424	0,460	0,686	0,372	tristeza
	0,097	0,005	0,583	0,097	0,167	0,218	0,555	0,152	aversion
	0,000	0,000	0,000	0,000	0,000	0,000	0,499	0,023	ira
	0,199	0,115	0,328	0,199	0,248	0,102	0,560	0,277	anticipacion
	0,146	0,070	0,354	0,146	0,206	0,109	0,567	0,284	ninguna
Weighted Avg.	0,314	0,234	0,356	0,314	0,260	0,118	0,567	0,250	

Figura 18: Anexo Resultados Clasificación WEKA - Corpus 2 Clasificador DT



- Corpus 2 – Clasificador NB

Correctly Classified Instances	687	33.8257 %							
Incorrectly Classified Instances	1344	66.1743 %							
Kappa statistic	0.1617								
Mean absolute error	0.1643								
Root mean squared error	0.2985								
Relative absolute error	90.6031 %								
Root relative squared error	99.1515 %								
Total Number of Instances	2031								
=== Detailed Accuracy By Class ===									
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,143	0,031	0,101	0,143	0,119	0,095	0,612	0,053	alegria
	0,617	0,492	0,315	0,617	0,417	0,111	0,631	0,382	confianza
	0,200	0,015	0,211	0,200	0,205	0,190	0,719	0,142	miedo
	0,058	0,031	0,134	0,058	0,081	0,040	0,630	0,112	sorpresa
	0,429	0,025	0,615	0,429	0,505	0,476	0,823	0,547	tristeza
	0,222	0,028	0,381	0,222	0,281	0,251	0,724	0,280	aversion
	0,042	0,022	0,044	0,042	0,043	0,021	0,558	0,032	ira
	0,254	0,122	0,371	0,254	0,302	0,153	0,604	0,322	anticipacion
	0,244	0,080	0,448	0,244	0,316	0,210	0,678	0,386	ninguna
Weighted Avg.	0,338	0,184	0,358	0,338	0,322	0,176	0,657	0,335	

Figura 19: Anexo Resultados Clasificación WEKA - Corpus 2 Clasificador NB

- Corpus 2 – Clasificador SVM

Correctly Classified Instances	785	38.6509 %							
Incorrectly Classified Instances	1246	61.3491 %							
Kappa statistic	0.2116								
Mean absolute error	0.1828								
Root mean squared error	0.2986								
Relative absolute error	100.8114 %								
Root relative squared error	99.1973 %								
Total Number of Instances	2031								
=== Detailed Accuracy By Class ===									
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,041	0,001	0,667	0,041	0,077	0,161	0,563	0,072	alegria
	0,547	0,312	0,391	0,547	0,456	0,215	0,627	0,348	confianza
	0,075	0,002	0,500	0,075	0,130	0,188	0,651	0,102	miedo
	0,058	0,014	0,250	0,058	0,094	0,087	0,587	0,107	sorpesa
	0,434	0,011	0,792	0,434	0,561	0,560	0,800	0,473	tristeza
	0,194	0,011	0,571	0,194	0,290	0,307	0,735	0,270	aversion
	0,000	0,002	0,000	0,000	0,000	-0,006	0,578	0,032	ira
	0,404	0,218	0,344	0,404	0,372	0,176	0,624	0,290	anticipacion
	0,441	0,226	0,342	0,441	0,385	0,198	0,657	0,294	ninguna
Weighted Avg.	0,387	0,182	0,406	0,387	0,366	0,222	0,650	0,292	

Figura 20: Anexo Resultados Clasificación WEKA - Corpus 2 Clasificador SVM



- Corpus 3 – Clasificador DT

Correctly Classified Instances	598	29.4436 %							
Incorrectly Classified Instances	1433	70.5564 %							
Kappa statistic	0.1089								
Mean absolute error	0.1637								
Root mean squared error	0.3322								
Relative absolute error	90.2624 %								
Root relative squared error	110.3485 %								
Total Number of Instances	2031								
=== Detailed Accuracy By Class ===									
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,041	0,016	0,061	0,041	0,049	0,031	0,576	0,038	alegria
	0,459	0,374	0,310	0,459	0,370	0,077	0,557	0,320	confianza
	0,000	0,008	0,000	0,000	0,000	-0,013	0,604	0,030	miedo
	0,115	0,054	0,151	0,115	0,131	0,070	0,565	0,099	sorpresa
	0,349	0,038	0,466	0,349	0,399	0,355	0,714	0,374	tristeza
	0,132	0,033	0,232	0,132	0,168	0,129	0,558	0,115	aversion
	0,042	0,004	0,200	0,042	0,069	0,082	0,512	0,029	ira
	0,248	0,171	0,291	0,248	0,268	0,082	0,558	0,258	anticipacion
	0,317	0,198	0,298	0,317	0,307	0,116	0,604	0,284	ninguna
Weighted Avg.	0,294	0,190	0,284	0,294	0,282	0,110	0,582	0,253	

Figura 21: Anexo Resultados Clasificación WEKA - Corpus 3 Clasificador DT

- Corpus 3 – Clasificador NB

Correctly Classified Instances	646	31.807 %							
Incorrectly Classified Instances	1385	68.193 %							
Kappa statistic	0.1555								
Mean absolute error	0.1624								
Root mean squared error	0.3066								
Relative absolute error	89.5916 %								
Root relative squared error	101.842 %								
Total Number of Instances	2031								
=== Detailed Accuracy By Class ===									
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,143	0,025	0,123	0,143	0,132	0,109	0,574	0,044	alegria
	0,400	0,289	0,337	0,400	0,366	0,106	0,605	0,352	confianza
	0,175	0,020	0,149	0,175	0,161	0,143	0,740	0,091	miedo
	0,135	0,057	0,164	0,135	0,148	0,085	0,599	0,111	sorpresa
	0,434	0,028	0,594	0,434	0,502	0,469	0,798	0,531	tristeza
	0,243	0,050	0,271	0,243	0,256	0,203	0,687	0,247	aversion
	0,042	0,020	0,049	0,042	0,045	0,024	0,566	0,035	ira
	0,315	0,204	0,304	0,315	0,309	0,109	0,596	0,315	anticipacion
	0,326	0,156	0,356	0,326	0,341	0,176	0,662	0,347	ninguna
Weighted Avg.	0,318	0,167	0,322	0,318	0,318	0,157	0,638	0,312	

Figura 22: Anexo Resultados Clasificación WEKA - Corpus 3 Clasificador NB

- Corpus 3 – Clasificador SVM

Correctly Classified Instances	810	39.8818 %							
Incorrectly Classified Instances	1221	60.1182 %							
Kappa statistic	0.2313								
Mean absolute error	0.1827								
Root mean squared error	0.2984								
Relative absolute error	100.7745 %								
Root relative squared error	99.1485 %								
Total Number of Instances	2031								
=== Detailed Accuracy By Class ===									
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,061	0,002	0,500	0,061	0,109	0,169	0,597	0,065	alegria
	0,569	0,300	0,410	0,569	0,477	0,246	0,656	0,369	confianza
	0,075	0,001	0,600	0,075	0,133	0,207	0,692	0,089	miedo
	0,096	0,023	0,254	0,096	0,140	0,115	0,577	0,112	sorpresa
	0,457	0,015	0,748	0,457	0,567	0,556	0,809	0,466	tristeza
	0,222	0,018	0,485	0,222	0,305	0,296	0,701	0,235	aversion
	0,021	0,002	0,200	0,021	0,038	0,058	0,577	0,036	ira
	0,391	0,209	0,346	0,391	0,367	0,174	0,625	0,291	anticipacion
	0,448	0,206	0,367	0,448	0,403	0,226	0,673	0,310	ninguna
Weighted Avg.	0,399	0,174	0,410	0,399	0,381	0,239	0,661	0,298	

Figura 23: Anexo Resultados Clasificación WEKA - Corpus 3 Clasificador SVM

- Corpus 4 – Clasificador DT

Correctly Classified Instances	603	29.6898 %							
Incorrectly Classified Instances	1428	70.3102 %							
Kappa statistic	0.066								
Mean absolute error	0.1723								
Root mean squared error	0.3021								
Relative absolute error	94.9993 %								
Root relative squared error	100.3615 %								
Total Number of Instances	2031								
=== Detailed Accuracy By Class ===									
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,000	0,001	0,000	0,000	0,000	-0,005	0,562	0,032	alegria
	0,754	0,707	0,281	0,754	0,410	0,047	0,549	0,308	confianza
	0,000	0,002	0,000	0,000	0,000	-0,005	0,562	0,033	miedo
	0,026	0,010	0,182	0,026	0,045	0,041	0,579	0,096	sorpesa
	0,211	0,012	0,627	0,211	0,316	0,333	0,667	0,319	tristeza
	0,049	0,003	0,538	0,049	0,089	0,146	0,512	0,096	aversion
	0,000	0,000	0,000	0,000	0,000	0,000	0,518	0,026	ira
	0,203	0,146	0,283	0,203	0,236	0,065	0,531	0,261	anticipacion
	0,124	0,060	0,356	0,124	0,184	0,101	0,582	0,275	ninguna
Weighted Avg.	0,297	0,236	0,319	0,297	0,238	0,090	0,562	0,242	

Figura 24: Anexo Resultados Clasificación WEKA - Corpus 4 Clasificador DT

- Corpus 4 – Clasificador NB

Correctly Classified Instances	702	34.5643 %							
Incorrectly Classified Instances	1329	65.4357 %							
Kappa statistic	0.1692								
Mean absolute error	0.1659								
Root mean squared error	0.2992								
Relative absolute error	91.4783 %								
Root relative squared error	99.412 %								
Total Number of Instances	2031								
=== Detailed Accuracy By Class ===									
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,163	0,033	0,110	0,163	0,131	0,108	0,620	0,053	alegria
	0,637	0,491	0,322	0,637	0,428	0,129	0,629	0,381	confianza
	0,200	0,017	0,195	0,200	0,198	0,181	0,745	0,135	miedo
	0,051	0,033	0,116	0,051	0,071	0,028	0,623	0,111	sorpresa
	0,423	0,018	0,685	0,423	0,523	0,506	0,819	0,539	tristeza
	0,194	0,023	0,389	0,194	0,259	0,238	0,715	0,258	aversion
	0,042	0,019	0,051	0,042	0,046	0,025	0,556	0,033	ira
	0,254	0,124	0,368	0,254	0,301	0,151	0,597	0,309	anticipacion
	0,265	0,080	0,467	0,265	0,338	0,232	0,672	0,379	ninguna
Weighted Avg.	0,346	0,183	0,369	0,346	0,329	0,186	0,653	0,328	

Figura 25: Anexo Resultados Clasificación WEKA - Corpus 4 Clasificador NB

- Corpus 4 – Clasificador SVM

Correctly Classified Instances	790	38.8971 %							
Incorrectly Classified Instances	1241	61.1029 %							
Kappa statistic	0.2137								
Mean absolute error	0.1828								
Root mean squared error	0.2985								
Relative absolute error	100.7913 %								
Root relative squared error	99.1771 %								
Total Number of Instances	2031								
=== Detailed Accuracy By Class ===									
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,041	0,001	0,667	0,041	0,077	0,161	0,572	0,073	alegria
	0,560	0,306	0,402	0,560	0,468	0,233	0,643	0,361	confianza
	0,075	0,002	0,500	0,075	0,130	0,188	0,670	0,105	miedo
	0,045	0,014	0,212	0,045	0,074	0,065	0,586	0,103	sorpresa
	0,417	0,008	0,839	0,417	0,557	0,568	0,805	0,486	tristeza
	0,174	0,010	0,568	0,174	0,266	0,288	0,753	0,257	aversion
	0,000	0,002	0,000	0,000	0,000	-0,006	0,590	0,031	ira
	0,415	0,222	0,346	0,415	0,378	0,182	0,619	0,290	anticipacion
	0,444	0,231	0,338	0,444	0,384	0,194	0,653	0,291	ninguna
Weighted Avg.	0,389	0,182	0,410	0,389	0,366	0,225	0,655	0,294	

Figura 26: Anexo Resultados Clasificación WEKA - Corpus 4 Clasificador SVM