



UJAT

UNIVERSIDAD JUÁREZ
AUTÓNOMA DE TABASCO

"ESTUDIO EN LA DUDA. ACCIÓN EN LA FE"



DIVISIÓN
ACADÉMICA DE
INFORMÁTICA Y
SISTEMAS

Universidad Juárez Autónoma de Tabasco

Tesis Doctoral

Modelo de Análisis de Emociones para evaluar el desempeño docente aplicando algoritmos de Aprendizaje Automático

Que presenta

Guadalupe Obdulia Gutiérrez Esparza

Para obtener el grado de

Doctor en ciencias de la computación

Directores

Dr. Alejandro Padilla Díaz

Dra. Juana Canul Reich

Aguascalientes, Ags., México

Enero 2017



UNIVERSIDAD JUÁREZ
AUTÓNOMA DE TABASCO

"ESTUDIO EN LA DUDA. ACCIÓN EN LA FE"



Universidad Veracruzana



UJAT

UNIVERSIDAD JUÁREZ
AUTÓNOMA DE TABASCO

"ESTUDIO EN LA DUDA. ACCIÓN EN LA FE"



DIVISIÓN
ACADÉMICA DE
INFORMÁTICA Y
SISTEMAS

Universidad Juárez Autónoma de Tabasco

Tesis Doctoral

Modelo de Análisis de Emociones para evaluar el desempeño docente aplicando algoritmos de Aprendizaje Automático

Que presenta

Guadalupe Obdulia Gutiérrez Esparza

Para obtener el grado de

Doctor en ciencias de la computación

Comité Tutorial: **Dr. Alejandro Padilla Díaz**
Dra. Juana Canul Reich
Dr. Julio Cesar Ponce Gallegos

Jurado: Dr. Alejandro Padilla Díaz	Presidente
Dra. Juana Canul Reich	Secretario
Dra. María de Lourdes Margain Fuentes	Vocal
Dr. Julio Cesar Ponce Gallegos	Vocal
Dr. Francisco Javier Álvarez Rodríguez	Vocal

Aguascalientes, Ags., México

Enero 2017



UNIVERSIDAD JUÁREZ
AUTÓNOMA DE TABASCO

"ESTUDIO EN LA DUDA. ACCIÓN EN LA FE"



Universidad Veracruzana

Esta Tesis Doctoral está dedicada...

*A mi madre por todo gran esfuerzo para sacarnos adelante a mi hermano y a mí, a pesar de todas las dificultades por las que paso, por inculcarnos el valor por el estudio y por cuidarnos siempre.
Gracias mamá, eres una mujer admirable.*

Agradecimientos:

A mi familia (mamá, Ricardo, Magaly y mis sobrinos) por su invaluable amor y apoyo brindado todos los días, gracias por estar a mi lado en todos los momentos buenos y malos.

Gracias también a mis amigas: Citla, Diana y Aiko, quienes confiaron en mí y me apoyaron para llevar a cabo este proyecto, a pesar del estrés y noches trabajando. Gracias por su apoyo, esfuerzo, dedicación y sobre todo por su amistad.

A cada uno de mis asesores:

- Dra. Lourdes Margain, no tengo palabras para agradecer todo lo que me has ayudado desde años atrás, gracias a ti he aprendido a ser fuerte y a confiar en mí a pesar de los obstáculos que se han presentado en mi vida personal y laboral. Gracias por tus consejos y el cariño que me has brindado siempre.*
- Dra. Juanita, gracias por su paciencia, por compartir conmigo su conocimiento y por hacerme sentir como en casa en las estancias que hice en Tabasco.*
- Dr. Alejandro Padilla, Dr. Julio Ponce y Dr. Francisco Álvarez, gracias por sus palabras de ánimo, por compartir sus conocimientos, por confiar en desde un principio y ayudarme a solucionar toda situación problemática que se presentaba.*
- Dr. Alejandro de Luna, Alex gracias por tu apoyo, por facilitarme permisos que requería para trabajar en esto, y por la confianza depositada en mí a lo largo de nuestra estancia en ISEI.*

Asimismo agradezco a Marco Álvarez por su apoyo y permitirme el tiempo para terminar sin interrupciones esta tesis doctoral, de igual manera gracias por cada una de tus anécdotas para hacerme sentir motivada y brindarme el aprendizaje de que los nervios no siempre son malos. Más que ser jefe, eres un líder que siempre procura escuchar, comprender y motivar, gracias.

Por último pero no menos importante quiero agradecer a mi amiga Tania Aglaé por todo tu apoyo, por escucharme en mis momentos más estresantes de la Tesis y estar a mi lado en una de mis más fuertes caídas emocionales, gracias por tu cariño, tus palabras de aliento y sobre todo gracias por tu amistad.



UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO

"ESTUDIO EN LA DUDA. ACCIÓN EN LA FE"

DIVISIÓN ACADÉMICA DE INFORMÁTICA Y SISTEMAS

Cunduacán, Tab. Octubre 3 de 2016.

En la Universidad Juárez Autónoma de Tabasco, de acuerdo al Reglamento de Estudios de Posgrado vigente, se revisó el trabajo de investigación titulado "Modelo de Análisis de Emociones para evaluar el desempeño docente aplicando Aprendizaje Automático", realizado por la C. Guadalupe Obdulia Gutiérrez Esparza, estudiante del Doctorado Interinstitucional en Ciencias de la Computación para obtener el Grado de Doctor en Ciencias de la Computación bajo la modalidad de Tesis.

Los integrantes del jurado, después de revisar el trabajo, y en virtud de que se han atendido satisfactoriamente las observaciones y recomendaciones, otorgamos nuestra aprobación para que continúe los trámites correspondientes a la obtención del grado.

Dr. Alejandro Padilla Díaz
Profesor-Investigador

Dra. Juana Canul Reich
Profesora-Investigadora

Dr. Julio Cesar Ponce Gallegos
Profesor-Investigador

Dra. María de Lourdes Margain Fuentes
Profesora-Investigadora

Dr. Francisco Javier Álvarez Rodríguez
Profesor-Investigador



**UNIVERSIDAD JUÁREZ
AUTÓNOMA DE TABASCO**

"ESTUDIO EN LA DUDA. ACCIÓN EN LA FE"



11111000011

DIVISIÓN ACADÉMICA DE INFORMÁTICA Y SISTEMAS

Oficio No. 2058/2016/DAIS-D
24 de noviembre de 2016

C. Guadalupe Obdulia Gutiérrez Esparza

Estudiante del DICC
Cunduacán, Tabasco.

Por este medio me permito Validar las Figuras Académicas participantes en el desarrollo de sus estudios doctorales, en sus distintas responsabilidades de Director, Co-director y Jurado de examen de grado, conforme a las actividades que señala el Programa de estudio del Doctorado Interinstitucional en Ciencias de la Computación (DICC).

Directores:

Dr. Alejandro Padilla Díaz
Dra. Juana Canul Reich

Comité Tutorial:

Dr. Julio César Ponce Gallegos
Dra. María de Lourdes Yolanda Margain Fuentes
Dr. Francisco Javier Álvarez Rodríguez

Sin otro particular reciba Usted un cordial saludo.

Atentamente

MATI, Eduardo Cruces Gutiérrez
Director

C.c.p. Dr. Jesús Hernández del Real.- Encargado del Despacho de la Coordinación de Posgrado
Archivo.
Consecutivo

Miembro CUMEX desde 2008
**Consortio de
Universidades
Mexicanas**
UNA ALIANZA DE CALIDAD POR LA EDUCACIÓN SUPERIOR

Carretera Cunduacán-Jalpa Km. 1, Colonia Esmeralda, C.P. 86690, Cunduacán, Tabasco, México.
E-mail: direccion.dais@ujat.mx
Teléfonos: (993) 358 1500 ext. 6727; (914) 336 0616; Fax: (914) 336 0870



UNIVERSIDAD JUÁREZ
AUTÓNOMA DE TABASCO

"ESTUDIO EN LA DUDA. ACCIÓN EN LA FE"



DIVISIÓN ACADÉMICA DE INFORMÁTICA Y SISTEMAS

Oficio No. 2059/2016/DAIS-D
24 de noviembre de 2016

C. Guadalupe Obdulia Gutiérrez Esparza
Estudiante del DICC
Cunduacán, Tabasco

Conforme a la sección 10, punto 3 del Programa de estudios del Doctorado Interinstitucional en Ciencias de la Computación (DICC), por este medio me permito hacer de su conocimiento, que las publicaciones derivadas de su trabajo de tesis doctoral, que se muestran en anexo 1, cumplen con los requisitos de obtención de grado.

Sin otro particular reciba Usted un saludo cordial.

Atentamente

MAT. Eduardo Cruces Gutiérrez
Director

C.c.p. Dr. Jesús Hernández del Real.- Encargado del Despacho de la Coordinación de Posgrado
Archivo
Consecutivo



UNIVERSIDAD JUÁREZ
AUTÓNOMA DE TABASCO

"ESTUDIO EN LA DUDA. ACCIÓN EN LA FE"



DIVISIÓN ACADÉMICA DE INFORMÁTICA Y SISTEMAS

ANEXO 1

Producto 1

Design of a SemanticLexiconAffectiveApplied to an Analysis Model Emotionsfor TeachingEvaluation.

Evidencias del arbitraje http://link.springer.com/chapter/10.1007/978-3-319-19638-1_25

Evidencias del índice

Estado publicado. Distributed Computing and Artificial Intelligence, 12th International Conference Volume 373 of the series Advances in IntelligentSystems and Computing pp 219- 226. La conferencia se realizó del 3 al 5 de junio en Salamanca, España.

Relación con el trabajo doctoral que describe los principios de diseño de la arquitectura para el modelo propuesto en la tesis doctoral, usando un léxico afectivo semántico.

Producto 2

A Sentiment AnalysisModel: To ProcessSubjective Social Corpus throughthe Adaptation of anAffective SemanticLexicon.

Evidencias del arbitraje http://link.springer.com/chapter/10.1007%2F978-3-319-13647-9_22

Evidencias del índice

Estado publicado. Human-Inspired Computing and ItsApplications. Volume 8856 of the series Lecture Notes in ComputerSciencepp 233-244. La conferencia se realizó del 16 al 22 de noviembre de 2014, en Tuxtla Gutiérrez, Chiapas, México.

Relación con el trabajo doctoral que describe los avances en el desarrollo del léxico afectivo semántico implantado en el modelo propuesto de la tesis doctoral, así como los primeros resultados.



UNIVERSIDAD JUÁREZ
AUTÓNOMA DE TABASCO

"ESTUDIO EN LA DUDA. ACCIÓN EN LA FE"



DIVISIÓN ACADÉMICA DE INFORMÁTICA Y SISTEMAS

Producto 3

Proposal of a SentimentAnalysisModel in Tweets forImprovement of theTeaching - LearningProcess in theClassroomUsing a Corpus of Subjectivity.

Evidencias del arbitraje

<http://ijcopi.org/ojs/index.php?journal=ijcopi&page=about&op=editorialTeam>

Evidencias del índice

<http://ijcopi.org/ojs/index.php?journal=ijcopi&page=about&op=editorialPolicies#custom0>

Estado publicado en el número correspondiente al periodo mayo – agosto de 2016.
Vol. 7, num. 2.

Relación con el trabajo doctoral que describe los resultados del experimento realizado con el algoritmo de NaïveBayes en combinación el léxico afectivo desarrollado para el modelo propuesto en la tesis doctoral con el fin de clasificar comentarios en polaridades negativa y positiva.

Miembro CUMEA desde 2008

Consortio de
Universidades
Mexicanas
UNA ALIANZA DE CALIDAD POR LA EDUCACIÓN SUPERIOR

Carretera Cunduacán-Jalpa Km. 1, Colonia Esmeralda, C.P. 86690. Cunduacán, Tabasco, México.

E-mail: direccion.dais@ujat.mx

Teléfonos: (993) 358 1500 ext. 6727; (914) 336 0616; Fax: (914) 336 0870



UNIVERSIDAD JUÁREZ
AUTÓNOMA DE TABASCO

"ESTUDIO EN LA DUDA. ACCIÓN EN LA FE"



11111000011

DIVISIÓN ACADÉMICA DE INFORMÁTICA Y SISTEMAS

Oficio No. 2878/16/DAIS/D
09 de diciembre 2016

Mtra. Guadalupe Obdulia Gutiérrez Esparza

En virtud de que cumple satisfactoriamente los requisitos establecidos en el Reglamento General de Estudio de Posgrado vigente en la Universidad, informo a Usted que se autoriza la impresión del trabajo recepcional **"Modelo de Análisis de Emociones para evaluar el desempeño docente aplicando algoritmos de Aprendizaje Automático"** para presentar examen y obtener el Grado de Doctor en Ciencias de la Computación bajo la modalidad de Tesis.

Sin otro particular, aprovecho la oportunidad para saludarle.

Atentamente

MAT. Eduardo Cruces Gutiérrez
Director

C.c.p. Director de Tesis 1: Dr. Alejandro Padilla Díaz
Archivo
Consecutivo

M.A.T.I.*ECG/Dr*JHR

Miembro CUMEX desde 2008

Consortio de
Universidades
Mexicanas

UNA ALIANZA DE CALIDAD POR LA EDUCACIÓN SUPERIOR

Carretera Cunduacán-Jalpa Km. 1, Colonia Esmeralda, C.P. 86690. Cunduacán, Tabasco, México.

E-mail: direccion.dais@ujat.mx

Teléfonos: (993) 358 1500 ext. 6727; (914) 336 0616; Fax: (914) 336 0870

Cesión de Derechos

Cunduacán, Tabasco, a 28 de noviembre de 2016

A quien corresponda:

Los abajo firmantes, declaramos que el trabajo de tesis doctoral titulado, "**Modelo de Análisis de Emociones para evaluar el desempeño docente aplicando algoritmos de Aprendizaje Automático**" es de nuestra autoría intelectual y por lo tanto cedemos los derechos de comunicación pública, reproducción, distribución, difusión en general y puesta a disposición electrónica de la citada tesis doctoral, de forma gratuita y no exclusiva, a la Universidad Juárez Autónoma de Tabasco, a la cual relevamos de cualquier sanción y asumimos responder a cualquier reclamo de derechos de autor ante las autoridades competentes.

Atentamente

Autores

Nombre	Dirección	Firma
Mtra. Guadalupe Obdulia Gutiérrez Esparza	Priv. Benjamin Méndez 102 int. 17, Col. Unidad Ganadera, C.P. 20130. Aguascalientes, Ags.	
Dr. Alejandro Padilla Díaz	Flor de noche buena 103 int. 8, Fracc. Villasur, C.P. 20296, Aguascalientes, Ags.	

A quien corresponda:

La que suscribe, autoriza por medio del presente escrito a la Universidad Juárez Autónoma de Tabasco para que utilice tanto física como digitalmente la tesis de grado denominada **“Modelo de Análisis de Emociones para evaluar el desempeño docente aplicando algoritmos de Aprendizaje Automático”**, de la cual soy autora y titular de los Derechos de Autor.

La finalidad del uso por parte de la Universidad Juárez Autónoma de Tabasco de la tesis antes mencionada, será única y exclusivamente para difusión, educación y sin fines de lucro; autorización que se hace de manera enunciativa más no limitativa para subirla a la Red Abierta de Bibliotecas Digitales (RABID) y a cualquier otra red académica con las que la Universidad tenga relación institucional.

Por lo antes manifestado, libero a la Universidad Juárez Autónoma de Tabasco de cualquier reclamación legal que pudiera ejercer respecto al uso y manipulación de la tesis mencionada y para los fines estipulados en éste documento.

Se firma la presente autorización en la ciudad de Villahermosa, Tabasco a los 28 días del mes de noviembre de 2016.

Autorizo


C. Guadalupe Obdulia Gutiérrez-Esparza

Tabla de Contenido

Índice de figuras	III
Índice de tablas	IV
Publicaciones	V
Resumen	1
Capítulo I	3
Introducción	3
1.1 Introducción.....	3
1.2 Planteamiento del problema.....	6
1.2.1 Problemática específica de la investigación.....	7
1.3 Objetivos.....	10
1.3.1 Objetivo general.....	10
1.3.2 Objetivos específicos.....	10
1.4 Hipótesis y preguntas de investigación.....	11
1.4.1 Hipótesis.....	11
1.4.2 Preguntas de investigación.....	11
1.5 Contribución.....	12
1.6 Organización del documento.....	13
Capítulo II	14
Marco Teórico y Estado del Arte	14
2.1 Marco Teórico.....	14
2.1.1 Análisis de Emociones.....	14
2.1.2 Análisis de Subjetividad.....	15
2.1.3 Léxicos Afectivos.....	17
2.1.4 Niveles del Análisis de Emociones.....	19
2.1.5 Minería de Textos.....	20
2.1.6 Aprendizaje Automático.....	21
2.1.6.1 Máquina de Soporte Vectorial (SVM).....	22
2.1.6.2 Naïve Bayes.....	25
2.1.6.3 Bosques aleatorios (Random Forest).....	26
2.1.7 Métodos para Selección de términos distintivos (<i>features</i>).....	27
2.1.8 Métricas para evaluar el desempeño de clasificación.....	30
2.2 Estado del Arte.....	35

2.2.1 Trabajos relacionados enfocados a la mejora de la educación	35
2.2.2 Trabajos relacionados que aplican aprendizaje automático	38
Capítulo III	43
Modelo SocialMining propuesto	43
3.1 Modelo SocialMining aplicando <i>Naïve Bayes</i> y Subjetividad.....	43
3.1.1 Descripción del Modelo	43
3.1.2 Datos y diseño experimental	47
3.1.2.1 Datos	47
3.1.2.2 Diseño experimental.....	49
3.1.3 Resultados.....	51
3.1.4 Caso de estudio	56
3.1.5 Discusión de resultados	58
3.2 Modelo SocialMining aplicando SVM y <i>Random Forest</i>	60
3.2.1 Descripción del Modelo	60
3.2.2 Datos y diseño experimental	62
3.2.2.1 Datos	62
3.2.2.2 Diseño experimental.....	63
3.2.3 Configuración de parámetros	63
3.2.4 Resultados de optimización de parámetros	64
3.2.5 Optimización de <i>features</i>	67
3.2.6 Resultados de clasificación con base a optimización de <i>features</i>	74
3.2.7 Discusión de Resultados	80
3.2.8 Algoritmo general del Modelo SocialMining	82
Capítulo IV	83
Conclusiones y trabajo futuro	83
4.1 Conclusiones	83
4.2 Trabajo futuro	87
Anexo A	88
Glosario de términos	88
Anexo B	91
Preguntas de la Evaluación Docente de la UPA	91
Anexo C	93
Evidencias de Evaluación docente en Twitter.....	93
Bibliografía	95

Índice de figuras

Figura 1. Panorama general de la computación del afecto	5
Figura 2. Interfaz de la interfaz de EED	9
Figura 3. Representación gráfica de la prueba de Turing.....	14
Figura 4. Representación de un modelo de aprendizaje automático	22
Figura 5. Representación de hiperplano óptimo en SVM	23
Figura 6. Representación del algoritmo Random Forest	27
Figura 7. El método contenedor en la selección de features	29
Figura 8. Métodos embebidos en la selección de features	29
Figura 9. Tipos de curva ROC.....	33
Figura 10. Arquitectura general del Modelo de SocialMining	43
Figura 11. Actividades del proceso de limpieza	44
Figura 12. Algoritmo del proceso de clasificación.....	47
Figura 13. Nube de palabras general (a), nube de palabras limpia (b).	49
Figura 14. Gráfico de resultados de clasificación de la versión 1	51
Figura 15. Gráfico de resultados de clasificación de la versión 2	52
Figura 16. Gráfico de resultados de clasificación de la versión 3.....	53
Figura 17. Gráfico de valores predictivos.....	55
Figura 18. Curva ROC en SocialMining con Naïve Bayes y el corpus de subjetividad	56
Figura 19. Gráfico de valores predictivos.....	57
Figura 20. Arquitectura de SocialMining aplicando aprendizaje automático	60
Figura 21. Resultados de remuestreo utilizando ROC.....	65
Figura 22. Metodología propuesta para la optimización de features	67
Figura 23. Nube de palabras de corpus de comentarios	68
Figura 24. Resultados de la ejecución en la optimización de parámetros en SVM - radial.....	70
Figura 25. Promedio de los resultados de ajuste de las SVM y Random Forest	72
Figura 26. Resultados de diferencia en SVMs y Random Forest	74
Figura 27. Comparación de desempeño del núcleo lineal.....	75
Figura 28. Comparación de desempeño del núcleo radial	76
Figura 29. Comparación de desempeño del núcleo polinomial	76
Figura 30. Comparación de desempeño del núcleo radial	77
Figura 31. Curva de ROC en SocialMining aplicando SVM y Random Forest	78
Figura 32. Enfoques diversos de aprendizaje automático según los materiales a utilizar	80
Figura 33. Diagrama general del Modelo SocialMining aplicando aprendizaje automático	82
Figura 34. Algoritmo para optimización de features.....	82
Figura 35. Algoritmo para optimización de parámetros	83
Figura 36. Emociones en Facebook	84

Índice de tablas

Tabla 1. Recursos disponibles para el análisis de subjetividad	15
Tabla 2. Recursos disponibles para el análisis de emociones	17
Tabla 3. Parámetros de núcleos	23
Tabla 4. Matriz de confusión para clasificación binaria	31
Tabla 5. Resumen de trabajos relacionados	41
Tabla 6. Relación de hashtag para evaluación docente en Twitter	48
Tabla 7. Extracto de corpus de subjetividad	50
Tabla 8. Matices de confusión de las versiones de SocialMining	53
Tabla 9. Resultados de VRP, ACC, ACCW y SPC	54
Tabla 10. Extracto de la relación de valores predictivos	54
Tabla 11. Matriz de confusión	57
Tabla 12. Resultados de la precisión y la sensibilidad	57
Tabla 13. Extracto de features utilizadas en SocialMining	63
Tabla 14. Resumen de resultados de optimización de parámetros	65
Tabla 15. Resultados promedio de clasificación	66
Tabla 16. Extracto de features con random forest	68
Tabla 17. Resultados de ajuste de SVM con núcleo Lineal	69
Tabla 18. Resultados de ajuste de SVM con núcleo Radial	69
Tabla 19. Resultados de ajuste de SVM con núcleo Polinomial	71
Tabla 20. Resultados de ajuste de Random Forest	71
Tabla 21. Promedio de los resultados de ajuste de las SVM y Random Forest	72
Tabla 22. Resultados de diferencia en SVMs y Random Forest	73
Tabla 23. Resultados promedios de métricas	74
Tabla 24. Extracto de valores predictivos en núcleo Radial de SVM	78

Publicaciones

Gutiérrez Guadalupe¹⁷, Margain Lourdes, Padilla Alejandro, Canul-Reich Juana and Ponce Julio. *Design of a Semantic Lexicon Affective Applied to an Analysis Model Emotions for Teaching Evaluation*. Distributed Computing and Artificial Intelligence, 12th International Conference Volume 373 of the series Advances in Intelligent Systems and Computing, pp. 219-226.

Guadalupe Gutiérrez, Lourdes Margain, Carlos de Luna, Alejandro Padilla, Julio Ponce, Juana Canul and Alberto Ochoa. *A Sentiment Analysis Model: To Process Subjective Social Corpus through the Adaptation of an Affective Semantic Lexicon*. Human-Inspired Computing and Its Applications Volume 8856 of the series Lecture Notes in Computer Science, pp. 233-244.

Guadalupe Gutiérrez, Alejandro Padilla, Juana Canul-Reich⁷, Alejandro De Luna and Julio Ponce. *Proposal of a Sentiment Analysis Model in Tweets for Improvement of the Teaching - Learning Process in the Classroom Using a Corpus of Subjectivity*. International Journal of Combinatorial Optimization Problems and Informatics, Vol. 7, No. 2, May-Aug 2016, pp. 22-34. ISSN: 2007-1558.

Resumen

El análisis de emociones conocido en inglés como *Sentiment Analysis*, el cual surge de áreas como el procesamiento de lenguaje natural y minería de datos, se ha convertido en un área clave para la sociedad debido a que gracias a ésta es posible identificar emociones expresadas en textos. El análisis de emociones involucra diversas tareas de investigación, entre las cuales destaca la detección de subjetividad [1, 2], clasificación por polaridad [3, 4], intensidad en la clasificación [5, 6] y la identificación de emociones [7-9].

El creciente auge del análisis de emociones se debe principalmente al interés de las empresas para conocer rápidamente opinión de los consumidores sobre sus productos y servicios como un medio para mejorar sus estrategias de publicidad y venta [10]. El área de la educación no se ha quedado atrás y ha planteado soluciones para mejorar la educación a distancia, cursos, talleres y clases presenciales [11, 12].

En el presente trabajo se describe el proceso de diseño, desarrollo y pruebas de un Modelo denominado SocialMining, con el propósito de apoyar al proceso de la evaluación docente. La recolección de comentarios fue realizada de dos maneras, la primera es que se recolectaron comentarios de una evaluación docente realizada en el 2015 y la segunda se realizó en Twitter mediante la participación de los alumnos de la carrera de Ingeniería en Sistemas Estratégicos de Información de la Universidad Politécnica de Aguascalientes.

Una vez recolectados los comentarios se efectuó una limpieza y revisión de los mismos para conformar un corpus de comentarios que apoyó a la generación de un corpus de subjetividad, así como a la selección y optimización de términos distintivos.

Para el proceso de clasificación de comentarios en el Modelo SocialMining se aplicaron algoritmos de Aprendizaje Automático (conocido en inglés como *Machine Learning*), como: *Naïve Bayes*, Máquinas de Soporte Vectorial (*Support Vector Machine*, *SVM por sus siglas en inglés*) y *Random Forest*, además métricas como precisión total, precisión equilibrada, sensibilidad, especificidad y el análisis ROC fueron utilizadas para medir el desempeño de los algoritmos clasificadores y en su caso mejorar. Los resultados obtenidos en clasificación en el Modelo SocialMining se consideran factibles, puesto que sobrepasan un 80%, valor que en la literatura [13] reportan aceptable.

Cabe resaltar que a pesar de que en la actualidad numerosas empresas y organizaciones aplican análisis de emociones para diferentes propósitos, la mayoría de las aplicaciones se concentran para el idioma inglés, así como para la inteligencia de negocios y publicidad. Hasta el momento no se reporta una herramienta o modelo dedicada al apoyo de la evaluación docente, asimismo tampoco se reporta un corpus afectivo o de subjetividad en español enfocado a procesos del área de la educación.

Capítulo I

Introducción

1.1 Introducción

Desde tiempos remotos el ser humano ha hecho uso de medios de comunicación para expresar sus necesidades, preferencias y emociones. La red de Internet ha sido uno de los medios más utilizados para posibilitar la comunicación y encontrar información de interés. Baldoni [14] resalta que en los últimos años la Internet ha dejado de ser solo una web de documentos y se ha convertido en una red de usuarios, asimismo el surgimiento de redes sociales como: Facebook, LinkedIn, Google+, Pinterest, Twitter y otras, ha generado una vasta cantidad de información acerca de las preferencias y comportamientos de los usuarios.

Un informe publicado por la consultora Nielsen [15], indica que un 70% de los usuarios de redes sociales presta atención a los comentarios de otros, un 65% busca información antes de hacer una elección, un 50% expresa opiniones positivas de marcas de consumo y otro 50% hace uso de las redes sociales para quejarse o hacer alguna reclamación. Twitter es una de las redes sociales usadas por los mexicanos para compartir información, realizar comentarios o mantener comunicación, de 65 millones de usuarios que hay en México en 2016, el 55% son usuarios activos de Twitter, sin embargo la Asociación Mexicana de Internet (AMIPICI¹) señala que en promedio, cada usuario está inscrito mínimo a cinco redes sociales. Un de las ventajas de Twitter para es que incorpora una plataforma de análisis gratis para los usuarios, razón por la cual, en la actualidad se ha destacado por ser de utilidad para obtener datos de los usuarios a nivel mundial. Debido a que en Twitter el tamaño de las publicaciones se limita a 140 caracteres, se cree que los comentarios denominados tweets pueden estar en su mayoría

¹ <https://www.amipci.org.mx/es/>

libres de spam y expresar una opinión más significativa a diferencia de otras redes sociales [16-19]. Los datos que continuamente se generan en las redes sociales en su mayoría, puede ser información valiosa ya que contienen las percepciones, opiniones y tendencias de los usuarios hacia objetos, personas o servicios. Por lo cual se considera una oportunidad para monitorear la opinión pública en una gran variedad de temas.

En un principio el tratamiento de textos era llevado a cabo mediante la aplicación del procesamiento de lenguaje natural, sin embargo para la explotación de información textual fue necesario la inclusión de otra área denominada minería de textos, la cual surge también de la minería de datos.

La combinación de minería de textos y procesamiento de lenguaje natural se aplicó para el procesamiento de información y clasificación de textos. No obstante la necesidad de explotar la gran cantidad de información en redes sociales y otros medios de Internet, provoca la aparición de una nueva área llamada Análisis de Emociones (conocida en inglés como Sentiment Analysis), término que comenzó a usarse alrededor del año 2001[20, 21].

El análisis de emociones también denominado minería de opiniones (en inglés *Opinion Mining*) ha sido definido como un área de estudio computacional de opiniones, sentimientos y emociones expresadas en textos [22]. Esta área en un principio estuvo dirigida principalmente a los negocios inteligentes (*Business Intelligence*) pero en la actualidad se ha ido extendido a otras áreas como la política, la medicina y educación, entre otras. Bing [23] sostiene que la mayoría del tiempo los individuos buscan opiniones antes de tomar una decisión, por lo cual la información resultante después de aplicar un análisis de emociones puede permitir identificar clientes potenciales a una empresa, o bien el éxito de un producto o servicio. Por su parte Balahur [24] dice *“Esta nueva realidad ha dado lugar a importantes transformaciones en la forma, extensión y rapidez de circulación de las noticias y sus opiniones asociadas, dando lugar a fenómenos sociales, económicos y psicológicos nuevos y desafiantes”*.

A pesar de que análisis de opiniones según Pang [2] comienza a tomar impulso alrededor del año 2001, existen estudios que datan de tiempo atrás como el de Ortoni [25] quien en 1988 trató de encontrar las bases para un modelo computacional para analizar textos con el fin de identificar emociones, no obstante

a pesar de su esfuerzo, nunca pudo llevar a cabo su investigación debido a que no contó con la tecnología requerida. Otro investigador que trato de definir el análisis de emociones fue Picard [26] quien en el año de 1995 definió la computación del afecto, como un área en donde influyen tareas como la desambiguación del sentido de las palabras y el análisis de la subjetividad mediante técnicas de aprendizaje automático, métodos estadísticos, reglas, léxicos, entre otros.

Debido a la importancia que ha generado la aplicación del análisis de emociones en diferentes tipos de estudios con el propósito de explotar información y encontrar patrones de comportamiento, así como las preferencias y necesidades de los usuarios, se ha combinado con otras áreas como aprendizaje automático, además del procesamiento de lenguaje natural y la minería de textos. El área de aprendizaje automático se ofrece un conjunto de algoritmos como una solución, cuando se desea agilizar el procesamiento de tareas como la clasificación de información, tal es el caso en esta investigación. La figura 1 presenta algunos de los algoritmos de aprendizaje automático empleados para procesos de clasificación, los cuales pueden ser divididos en dos clases, clasificadores simples (k-vecinos más cercanos (k-NN), SVM, redes neuronales (ANN) y análisis lineal discriminante (LDA) y métodos combinados de aprendizaje (boossting, bagging, Random forest (bosques aleatorios en español), Random subspace, C5.0). En este trabajo se hace uso de la SVM, Naïve Bayes y Random Forest.

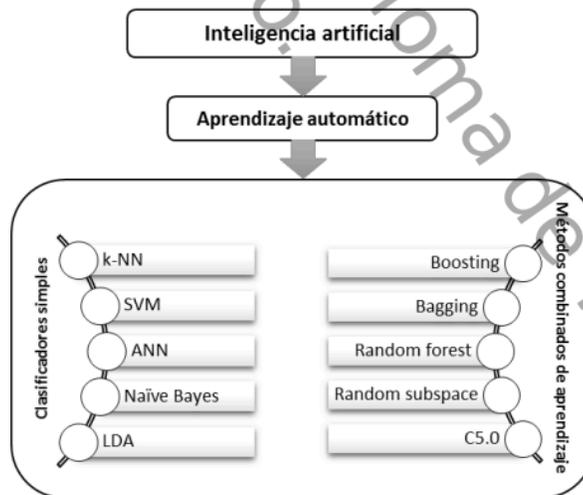


Figura 1. Panorama general de la computación del afecto

La aplicación de aprendizaje automático en análisis de emociones ha obtenido resultados factibles en diferentes contextos como la inteligencia de negocios o empresarial, la política y la educación.

La educación es una de las áreas que en los últimos años ha mostrado interés por analizar los comentarios de los estudiantes con el fin de que los profesores mejoren sus técnicas de enseñanza propiciando un adecuado aprendizaje en los estudiantes.

1.2 Planteamiento del problema

Existe un exponencial interés por parte de empresas de diversos ámbitos para analizar los comentarios u opiniones de los usuarios de redes sociales, con el fin de identificar las emociones que expresan. Por lo cual se han aplicado diferentes tipos de técnicas y herramientas para realizar este tipo de análisis. Sin embargo el problema general que sostienen autores como Wang y Feldman [27, 28] es que *“No existe un modelo o metodología para realizar un análisis de emociones que permita obtener resultados factibles en variados contextos e idiomas”*. En la siguiente sección se presenta una revisión con base en la literatura sobre los problemas que se plantean para el análisis de emociones.

Algunas de las necesidades del área de análisis de emociones según investigaciones de Feldman [28] son las siguientes:

- Es necesario la creación de un modelo para el análisis de emociones que permita identificar si una expresión es objetiva o subjetiva.
- Se considera necesario mejorar el proceso de identificación de texto relevante para cada objeto cuando se habla de varios al mismo tiempo, según Feldman [28] este proceso carece aún de ser factible.
- La identificación del sarcasmo es otra tarea que requiere de atención ya que a pesar de que existen investigaciones como la de Tsur [29] (quien identifica el sarcasmo con base en métodos de clasificación), aún no se ha integrado exitosamente con sistemas autónomos de análisis de opiniones.
- Otra tarea desafiante es la supresión del ruido en textos, es decir la mala ortografía, errores gramaticales, problemas de puntuación y spam.
- La concatenación de sentencias subjetivas previamente procesadas y analizadas para la identificación de emociones con aquellas que son

objetivas, se considera aun una necesidad sobre todo para los artículos de hechos o noticias.

Por su parte Wang [27] menciona que a pesar de que en la actualidad ha habido un creciente interés en combinar técnicas de aprendizaje para mejorar el proceso de clasificación, los resultados obtenidos aún no son confiables. De igual forma indica que no hay una metodología a adoptar dado un problema en un determinado dominio, por lo cual es común que se desarrollen múltiples clasificadores o sistemas híbridos.

Otra de las investigadoras fuertemente ligada al área de análisis de emociones es Balahur [24] quien detecta los siguientes problemas:

- Debido a que no hay un proceso definido en el análisis de emociones no está claro la manera en que los distintos investigadores puedan comparar el rendimiento de sus sistemas, puesto que puede haber variaciones en los textos que evalúan.
- No existen recursos para todos los contextos o idiomas en el análisis de emociones.
- La mayoría de sistemas trabajan a nivel léxico, utilizando reglas, léxicos, métodos estadísticos o aprendizaje automático. La investigación que se ha hecho hasta ahora no toma en cuenta otros niveles de análisis, como el análisis sintáctico o semántico.

De acuerdo a lo explicado en esta sección se puede comprender que existen varios problemas a solucionar en el área del análisis de emociones, abarcarlos todos para ofrecer una solución involucra décadas de investigación, inversión en equipo de alta tecnología y recursos humanos. Debido a esto el análisis de emociones es aplicado para solucionar problemas en áreas específicas, como por ejemplo analizar las preferencias de la población para la elección de un político [30], o bien, analizar las necesidades de usuarios para el lanzamiento de un nuevo producto o servicio que pueda generar altas ventas [31].

1.2.1 Problemática específica de la investigación

Propiciar el aprendizaje en los estudiantes es una tarea que requiere la atención del profesor en los alumnos debido a diferentes problemas que rodean al estudiante, como la capacidad cognitiva de cada uno de ellos, los

comportamientos distractores e incluso la metodología que el profesor emplea para enseñar puede provocar que no exista un aprendizaje. Evaluar el desempeño de un profesional dedicado a la educación es un tema considerado importante para las instituciones de educación superior. Este tipo de evaluación es uno de los procesos con mayor complejidad en cualquier casa de estudios, dado a diversos factores y criterios que deben reunirse para ser concentrados y emitir una evaluación final al profesional.

Algunos de los instrumentos utilizados para evaluar al docente pueden ser la guía de observación, una rúbrica con diferentes criterios de evaluación o pares académicos especializados en el área de pedagogía. Sin embargo, cuando el desempeño docente es evaluado considerando las opiniones de los alumnos, surge la necesidad de usar métodos avanzados para el análisis de los comentarios de los alumnos para ofrecer en un menor tiempo una solución a posibles problemas o situaciones que requieran un apoyo inmediato.

En la Universidad Politécnica de Aguascalientes (UPA) es aplicada una encuesta de evaluación docente mediante un sistema informático, la cual es aplicada dos veces en un ciclo cuatrimestral. La encuesta de evaluación docente aplicada (denominada de aquí en adelante como EED) contiene un total de 21 preguntas, donde las primeras 20 son opción múltiple y en la pregunta 21 se permite al alumno ingresar un comentario u opinión del docente como sugerencia constructiva (ver figura 2).

Una vez finalizado el tiempo de apertura de la EED, los resultados son exportados a un archivo en Excel que puede llegar a contener alrededor de 12,000 comentarios, dependiendo de la matrícula de alumnos y profesores que imparten clase. El Departamento de Superación Docente considera los resultados de las primeras veinte preguntas y realiza un informe, que distribuye a cada Director de Programa Académico (DPA), en el caso de los comentarios que publican los estudiantes, son leídos por el DPA y el profesor correspondiente. Sin embargo no se realiza un análisis detallado de los mismos.

Tradicionalmente la evaluación docente ha sido empleada para calificar al profesor en forma global, indicando si es un profesor es “bueno” o “malo”, sin embargo se busca analizar la opinión de los alumnos con la finalidad de provocar una mejora continua en las clases detectando situaciones donde sea posible

recomendar al docente cursos o talleres para mejorar su proceso de enseñanza. O incluso detectar situaciones donde se requiera el apoyo de formación integral o atención especial al grupo o estudiante.

Figura 2. Interfaz de la interfaz de EED

Debido a la gran cantidad de comentarios que pueden ser generados de una evaluación docente pudiera ser complicado detectar si existe en algún grupo un problema que requiera atención inmediata, o bien situaciones donde es necesario tomar una decisión respecto al grupo o el profesor. Otro problema que se desprende de la evaluación docente es que ésta es aplicada dos veces en un ciclo escolar, por lo cual en ocasiones no es posible detectar posibles problemas con los estudiantes o el profesor, por esta razón se propone el uso de redes sociales. Novak [27] comenta que hay un **crecimiento exponencial en el uso de redes sociales como Twitter**, para que las **universidades a nivel mundial mantengan comunicación con** sus alumnos, lo cual ha permitido identificarlas como un recurso educativo potencial. Sin embargo de igual manera se requiere usar métodos avanzados para el análisis de los comentarios de los alumnos.

Por lo tanto con base al problema descrito es necesario el desarrollo de un modelo para el análisis de emociones en textos, con el fin de semi-automatizar el proceso de clasificación de comentarios, para apoyar a la evaluación docente en las instituciones de educación.

1.3 Objetivos

1.3.1 Objetivo general

Diseñar un Modelo para Analizar Emociones en textos en español aplicando algoritmos de aprendizaje automático para apoyar en el proceso de la evaluación del desempeño docente.

1.3.2 Objetivos específicos

Objetivo específico 1

Identificar los términos distintivos de la evaluación docente realizada por alumnos, mediante un análisis de los comentarios para conformar un corpus de subjetividad.

Objetivo específico 2

Aplicar el algoritmo de Naïve Bayes en combinación con los términos distintivos obtenidos a partir del análisis de comentarios para clasificarlos dentro de tres clases (positiva, negativa y neutral).

Objetivo específico 3

Aplicar algoritmos de aprendizaje automático en el proceso de clasificación de comentarios de la evaluación docente.

Objetivo específico 4

Evaluar el desempeño de los algoritmos implementados en el Modelo SocialMining.

1.4 Hipótesis y preguntas de investigación

1.4.1 Hipótesis

Este proyecto está sujeto a las siguientes hipótesis de investigación:

Hipótesis 1:

La combinación del corpus de subjetividad con el algoritmo de Naïve Bayes contribuye al Modelo SocialMining para alcanzar un 90% en precisión total y equilibrada, en la clasificación de comentarios de la evaluación docente.

Hipótesis 2:

El análisis de términos distintivos (*features*) permite a los algoritmos de aprendizaje automático disminuir el margen de error en la clasificación de comentarios.

1.4.2 Preguntas de investigación

La primera pregunta que motiva el desarrollo de esta investigación es:

- ¿Cuáles son los términos distintivos que deben incluirse en el corpus de subjetividad y que fuerza subjetiva debe tener cada uno de éstos, para alcanzar un 90% de precisión en clasificación de comentarios de la evaluación docente?
- ¿Qué técnica de aprendizaje automático apoya al Modelo SocialMining para clasificar cometarios con el menor margen de error?

1.5 Contribución

A pesar de que el análisis de emociones no es relativamente nuevo ya que comienza a tomar impulso alrededor del año 2001 según indica Bing [23], no existe un modelo o metodología a seguir para implementar en diferentes áreas o contextos. La educación es una de las áreas que en los últimos años se ha mostrado interesada por aplicar el análisis de emociones para mejorar el proceso de enseñanza – aprendizaje, sin embargo esta área aún se considera poco abordada debido a que son pocos los investigadores que han logrado desarrollar alguna herramienta, aplicación o modelo para fines de análisis de emociones.

Mediante el desarrollo del Modelo SocialMining se brinda un aporte al área de la educación para apoyar al proceso de la evaluación docente, analizando los comentarios de los alumnos con el fin de realizar una clasificación bipolar (comentarios positivos y negativos) considerando las emociones en los comentarios.

Una de las contribuciones del Modelo SocialMining es la creación de un corpus de subjetividad utilizado como apoyo para la identificación de comentarios positivos y negativos. Dicho corpus de subjetividad contiene términos seleccionados mediante un análisis de comentarios de la evaluación docente, los términos fueron etiquetados con base a su polaridad (positivo, negativo), además de una fuerza subjetiva para indicar la influencia del término en la clasificación. En este trabajo el Modelo SocialMining fue probado e implementado en la Universidad Politécnica de Aguascalientes, sin embargo es posible aplicarlo en otras instituciones de educación superior, no obstante es importante considerar los coloquialismos de la región, por lo cual se recomienda seguir actualizando el corpus de subjetividad para futuras aplicaciones.

Cabe resaltar que no existen corpus en español dedicados a la educación. La mayoría de los corpus disponibles en internet están en inglés, además de orientados a otros contextos y aunque es posible encontrar algunos en español, muchos de éstos han sido traducidos por lo que es necesario adecuarlos al contexto requerido y revisar la estructura y gramática según el idioma con el cual se va trabajar.

Otra de las contribuciones del Modelo SocialMining es la aplicación de algoritmos de aprendizaje automático, ya que en la actualidad existe un creciente interés para analizar emociones de comentarios, sin embargo la mayoría de los usuarios utiliza herramientas disponibles en internet. La aplicación de algoritmos de aprendizaje automático en este trabajo se considera importante debido a que complementan el Modelo SocialMining propuesto para la mejora del desempeño en clasificación de comentarios, el cual es validado mediante el uso de métricas.

1.6 Organización del documento

El documento está estructurado como sigue:

- El **Capítulo I** explica el problema a resolver, plantea los objetivos de la investigación e incluye las preguntas de investigación derivadas de las hipótesis, así como la presentación de la contribución doctoral.
- En el **Capítulo II** se expone el estado del arte que permite ubicar esta propuesta en el contexto de las Ciencias de la Computación, específicamente en el área de la Inteligencia Artificial, además se introducen los conceptos relacionados con el análisis de emociones, algoritmos de aprendizaje automático y las métricas utilizadas para medir el desempeño en la clasificación de los algoritmos utilizados.
- En el **Capítulo III** se presenta la descripción del diagrama de la arquitectura general del Modelo, la naturaleza de los datos, así como los algoritmos o técnicas de aprendizaje automático utilizados.
- En el **Capítulo IV** se describen las pruebas realizadas y los resultados obtenidos.
- Finalmente en el **Capítulo V** se presentan las Conclusiones y trabajos futuros plasman las contribuciones y resultados obtenidos en esta investigación.

Capítulo II

Marco Teórico y Estado del Arte

2 Marco Teórico

2.1.1 Análisis de Emociones

El análisis de emociones es el área de estudio que analiza las opiniones, sentimientos, valoraciones, actitudes y emociones [2] de los usuarios hacia productos, servicios, organizaciones, personalidades, eventos, entre otros. Esta área se enfoca en comentarios que expresan o implican sentimientos positivos o negativos dando retroalimentación a algún tema [23]. Picard [26] menciona que la prueba de Turing (ver figura 3) podría considerarse uno de los primeros intentos en analizar las emociones del usuario mediante textos, debido a que fue diseñada para establecer una comunicación a través de chat (mensajería instantánea) dejando de lado la expresión facial y la entonación de voz.



Figura 3. Representación gráfica de la prueba de Turing

La prueba de Turing consiste en un desafío, donde un juez, la máquina y el ser humano se encuentran cada uno en diferentes habitaciones, la finalidad del juez es identificar al ser humano y a la máquina mediante una serie de preguntas y respuestas por escrito, donde al ser humano y a la máquina se les permite mentir o equivocarse al contestar las preguntas. Por esta razón se podría decir que la subjetividad en los textos está presente.

2.1.2 Análisis de Subjetividad

La subjetividad se refiere a la perspectiva, sentimientos, creencias o deseos y no ha hechos. En el área de procesamiento de lenguaje natural, Wiebe [32] define la subjetividad como: “*La expresión lingüística de alguien en cuanto a opiniones, sentimientos, emociones, evaluaciones, creencias y especulaciones*”, cabe resaltar que para esta definición el autor se inspiró en el trabajo de Banfield [33] quien definió la subjetividad como: “*Oraciones que toman el punto de vista de un personaje*”. Esta definición también fue la base para que Quirk [34] presentará los estados privados como la actitud de una persona hacia un objeto. Con base a las anteriores definiciones la diferencia entre subjetividad y objetividad es que la primera expresa una opinión personal considerando creencias y emociones acerca de un objeto, en cambio la segunda expresa información factual de un objeto, Wiebe [35] plantea el siguiente ejemplo:

Oración subjetiva: *este libro es asombroso*
Oración objetiva: *este libro costó 10 euros en el sitio de Amazon.com*

Para apoyar en el proceso de la detección de la subjetividad se han creado una variedad de léxicos², la tabla 1 presenta una breve descripción acerca de algunos de éstos.

Tabla 1. Recursos disponibles para el análisis de subjetividad

Nombre	Descripción
General Inquirer [36]	Los textos subjetivos están acorde a la polaridad y emoción. Contiene 11,788 términos desambiguados, fue elaborado de forma manual para el idioma inglés.

¹⁰

² Vocabulario, conjunto de las palabras de un idioma, o de las que pertenecen al uso de una región, a una actividad determinada, a un campo semántico dado

Comlex [37]	Diccionario especialmente elaborado para sistemas de PLN. Asimismo también contiene un gran número de adverbios de actitud. Contiene 38,000 términos desambiguados con sentido, fue elaborado de forma manual para el idioma inglés.
Lexicon por Yu. [38]	Otro esquema de anotación y corpus para la subjetividad frente a la objetividad de clasificación, así como la determinación de la polaridad a nivel de la oración. Parte de un conjunto de 1,336 palabras. Fue elaborado de forma semiautomática para el idioma inglés.
Lexicon of Appraisal terms [9]	Corpus basado en la teoría de Appraisal [9], la cual dice que la emoción se compone de actitud (afecto, aprecio, juicio), de graduación (fuerza y enfoque), de orientación (positiva y negativa) y de la polaridad (que puede estar marcado o no marcado). Este corpus contiene 1,329 términos en inglés y fue elaborado de forma semiautomática.
Corpus con dos diccionarios bilingües [28]	En este corpus se usa el léxico de Wilson [39] denominado "Opinion Finder lexicon" así como dos diccionarios bilingües Inglés-Rumano para traducir las palabras. Contiene alrededor de 8,000 palabras y soporta el idioma inglés y rumano.
MPQA [40]	Se trata de un corpus que ha sido traducido para cinco idiomas: inglés, francés, árabe, alemán, rumano y español, contiene 9,700 frases y fue elaborado de forma semiautomática.

Pang [41] ha demostrado que el análisis de subjetividad es el primer paso para obtener buenos resultados en el análisis de emociones. Balahur [24] indica que el texto (opinión) debe pasar por un primer análisis para determinar si es subjetivo u objetivo y posteriormente debe pasar por otro análisis para identificar el tipo de emoción que pudiera contener, donde dependiendo del tipo de análisis será posible concluir si se trata de una opinión positiva, negativa o neutral, o bien, con base a un análisis más detallado es posible identificar emociones como: tristeza, alegría, decepción, entre otras.

Además de léxicos desarrollados para análisis de subjetividad, también existen léxicos especializados para el análisis de emociones, los cuales han sido nombrados como léxicos afectivos.

2.1.3 Léxicos Afectivos

Un léxico afectivo puede entenderse como un conjunto de palabras asociadas con una etiqueta que posee información de tipo emocional o afectivo. La tabla 2 describe en términos generales algunos de estos léxicos.

Los léxicos afectivos pueden estar basados en palabras o términos, en donde los primeros se basan en palabra etiquetadas con información emocional, y los segundos se basan en conceptos, los cuales con base al significado de las palabras etiquetan el concepto acorde a la emoción que expresa. Ambos tipos de léxicos se basan en dos teorías, la teoría psicológica de las categorías emocionales y la teoría psicológica de las dimensiones emocionales. La teoría psicológica de las categorías emocionales busca etiquetar las entradas con emociones básicas como tristeza o miedo, mientras que los basados en dimensiones emocionales etiquetan las entradas por polaridad (positivo o negativo).

Tabla 2. Recursos disponibles para el análisis de emociones

Recurso	Año	Descripción
WordNet Affect [42]	2004	Corpus de términos clasificados acorde con las seis categorías básicas de emociones (alegría, tristeza, miedo, sorpresa, ira y disgusto).
Emotion triggers [43]	2008	Se trata de una base de datos de conceptos que expresan una idea dependiendo del contexto, social, cultural y educativo. Busca introducir términos relacionados con la vida real basado en dos léxicos, para inglés usa ConceptNey y para español usa Larousse Ideologic Dictionary of the Spanish Language.
MicroWNOp [44]	2007	Base de datos que contiene palabras con su polaridad asociada, se construyó sobre la base de un conjunto de términos extraídos del léxico "General Inquirer" y posteriormente se agregaron synsets ³ .
General Inquirer [36]	1966	Contiene palabras desambiguas con sentido. Los subjetivos están clasificados por polaridad y emoción.

³ Conjunto de sinónimos que comparten significado común.

La mayoría de los léxicos basados en dichas teorías son a nivel palabra, en donde se destaca el *LIWC Dictionary*⁴ (*Linguistic Inquiry and Word Count Dictionary* por sus siglas en inglés) para aquellos sustentados en la teoría psicológica de las categorías emocionales y el *General Inquire*⁵ para los asentados en la teoría de dimensiones emocionales. En los léxicos basados en conceptos resaltan *WordNet Affect*⁶ y *SentiWordNet*⁷.

WorNet Affect es un corpus que contiene los conceptos de *WordNet* que poseen un significado afectivo, los conceptos son etiquetados con base a una jerarquía emocional. Ha sido realizado mediante un proceso semi-automático, donde se etiquetó manualmente con categorías emocionales un conjunto inicial de *synsets*, expandiéndolo automáticamente mediante las relaciones de éstos posteriormente.

Está diseñado para su uso con categorías emocionales, sin embargo las categorías son extensas y no acorde con las teorías psicológicas de las categorías emocionales. Asimismo contiene conceptos etiquetados con más de una categoría emocional, lo que puede ocasionar inconvenientes al determinar la emoción correcta sin un proceso de desambiguación específico. Además No hay una versión actualizada debido a que la última versión de *WordNet* para *Windows* es la 2.1, mientras que *WordNet Affect* está basado en la versión 1.6.

En *SentiWordNet* cada *synset* ha sido etiquetado con tres valores numéricos en las categorías Pos(s), Neg(s) y Obj(s), representando, la polaridad positiva, la polaridad negativa y la neutralidad del concepto. Ha sido desarrollado mediante la clasificación automática de los *synsets* de *WordNet*.

A pesar de que *SentiWordNet* posee la ventaja de cubrir todos los *synsets* de *WordNet*, es también una debilidad debido a que en varios casos, los conceptos que no estén bien relacionados con el conjunto inicial de *synsets* etiquetados manualmente no obtendrán puntuaciones muy acordes.

Tanto *WordNet Affect* como *SentiWordNet* contienen una variedad de problemas referentes a actualizaciones, desambiguación y otros, asimismo

⁴ <http://www.liwc.net/>

⁵ <http://www.wjh.harvard.edu/~inquirer/>

⁶ <http://wdomains.fbk.eu/>

⁷ <http://sentiwordnet.isti.cnr.it/>

ninguno está disponible para el idioma Español. Por lo cual en la literatura es posible encontrar léxicos creados para llevar a cabo investigaciones específicas en diferentes idiomas [44-46].

Una de las consideraciones más importantes para el desarrollo de léxicos afectivos es identificar el nivel de análisis de emociones que se desea abordar, los cuales se explican en la siguiente sección.

2.1.4 Niveles del Análisis de Emociones

Según Feldman [28] el análisis de emociones puede ser clasificado en niveles, los cuales se describen a continuación.

- **Análisis de emociones a nivel documento.** En este nivel se asume que el documento a analizar posee una opinión del autor acerca de un determinado objeto. Algunas de las técnicas de clasificación comúnmente usadas para este tipo de análisis según Feldman [28] son: la Máquina de Soporte Vectorial (*Support Vector Machines*, SVM por sus siglas en inglés), *Naïve Bayes*, Regresión Logística y *k-NN*.
- **Análisis de emociones a nivel sentencia.** El análisis de emociones a nivel sentencia analiza las diferentes opiniones sobre un objeto. Este nivel es usado para analizar las opiniones de los usuarios de redes sociales, blogs, chats y foros, sin embargo uno de los problemas más comunes de este nivel es la aparición del sarcasmo. Donde investigadores como Tsur [29] buscan aportar soluciones a este problema.
- **Análisis de emociones a nivel atributo.** En este nivel al igual que los dos anteriores se realiza un análisis para determinar la emoción dentro de un documento o sentencia sobre un determinado objeto, considerando algunos atributos del objeto [31].
- **Análisis de emociones a nivel comparativo.** Este nivel se aplica cuando los usuarios no provén una opinión directa sobre un objeto, sino más bien hacen una comparación con algún otro. Jindal [47] fue uno de los primeros investigadores en el análisis de emociones a nivel comparativo y propuso una solución para cubrir hasta en un 98% las opciones comparativas en el

idioma inglés usando: adjetivos y adverbios comparativos (*more, less and words ending "ing"*), adjetivos y adverbios superlativos (*most, least and words ending "est"*), haciendo uso además de palabras clave como: preferido, superior, superar, contra, y otras.

Otros investigadores consideran que el análisis de emociones puede tener dos enfoques: recuperación de emociones y la extracción de emociones.

- **Extracción de emociones**, dado un conjunto de documentos textuales, frases o fragmentos de textos escritos que expresan actitudes, determina la polaridad (positivo, negativo) de la actitud que expresa, o bien, la emoción que expresa (tristeza, alegría, ira).
- **Recuperación de emociones**, identifica aquellos los comentarios que expresan una actitud hacia un determinado elemento. Este enfoque es utilizado por empresas para conocer la opinión de los usuarios en cuanto a una característica de un producto, por ejemplo la preferencia por dispositivos móviles de una misma marca [31].

El análisis de emociones en combinación con otras áreas como la minería de textos y el aprendizaje automático ha tenido buenos resultados en procesos de clasificación y predicción. Estas áreas se describen de forma general en las siguientes secciones.

2.1.5 Minería de Textos

¹⁹ La minería de textos es un área de investigación dedicada al procesamiento automático de la información, se refiere al proceso de extracción de patrones o conocimientos clave desde documentos no estructurados [48, 49]. La minería de datos se define como el proceso de analizar textos o documentos no estructurados con la finalidad de extraer información útil para propósitos particulares [50]. Es posible realizar una analogía con el área de la minería de datos, la cual busca a diferencia de la minería de textos, se enfoca en descubrir conocimientos en bases de datos [51].

² Debido a que la forma más natural de almacenar información es mediante texto, la minería de textos tiene un potencial más alto que incluso la minería de

datos. Un estudio reciente indica que el 80% de la información de una empresa está contenida en documentos de texto.

La minería de datos puede ser empleada para inteligencia del gobierno, detección de actividades terroristas, inteligencia de negocios y análisis de emociones [52]. Es posible combinar esta área con algunas otras el aprendizaje automático [53] para mejorar los resultados de desempeño en clasificación o predicción.

2.1.6 Aprendizaje Automático

El aprendizaje automático es un método de análisis de datos que permite automatizar la construcción de un modelo analítico. Usando algoritmos que aprenden de forma iterativa mediante el procesamiento de conjuntos de datos, el aprendizaje automático permite encontrar patrones ocultos sin programar explícitamente la búsqueda [54]. El aspecto iterativo del aprendizaje automático es importante debido a que los modelos o algoritmos están expuestos a nuevos datos, y estos modelos son capaces de adaptarse y procesar la información de forma independiente para aprender de los cálculos anteriores y predecir de forma confiable los resultados más óptimos de acuerdo a las circunstancias [55].

En la actualidad, el aprendizaje automático es utilizado principalmente en: detección de fraudes, resultados de búsquedas web, análisis de emociones p sentimientos en textos, reconocimiento de imágenes y patrones, así como filtrado de spam en correo electrónico.

Dos de los métodos de aprendizaje automático más ampliamente adoptados, son: aprendizaje supervisado y aprendizaje no supervisado. Sin embargo, también existen otras tecnologías como aprendizaje semi-supervisado y aprendizaje por refuerzo.

Los principales algoritmos de aprendizaje automático son: redes neuronales artificiales (*Artificial Neural Network*), árboles de decisión (*Decisión tres*), SVM, mapeo de vecino más cercano (*Nearest Neighbors*), técnicas de optimización de búsqueda local (por ejemplo: algoritmos genéticos), redes bayesianas y *Random Forest*, entre otros.

La figura 4 presenta una simulación del funcionamiento de un algoritmo de aprendizaje supervisado [56], como se puede observar es necesario realizar un

entrenamiento mediante documentos, imágenes o conjuntos de datos previamente limpios, asimismo se requiere cierto proceso de etiquetamiento y una selección de *features*. El resultado es un modelo predictivo que obtiene las etiquetas esperadas en los documentos o imágenes, dependiendo de la función para la que fue diseñado.

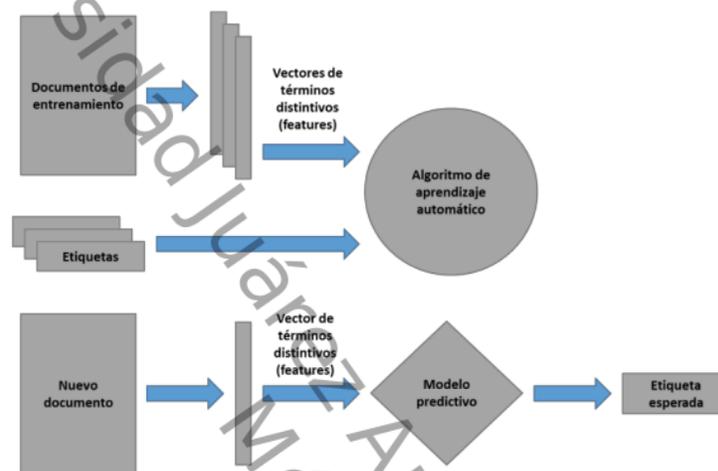


Figura 4. Representación de un modelo de aprendizaje automático [56]

En este trabajo los algoritmos de aprendizaje automático utilizados son: SVM y Random Forest.

2.1.6.1 Máquina de Soporte Vectorial (SVM)

SVM es un algoritmo de aprendizaje supervisado que se ha destacado por su calidad en la clasificación de textos [57] incluso como alternativa a las redes neuronales. La SVM se ha utilizado con éxito en áreas como la recuperación de información, la categorización de textos, el reconocimiento de escritura o la clasificación de imágenes. La SVM surgió como un método de clasificación basado en la teoría de minimización del riesgo estructural de Vapnik [58], es principalmente un método de clasificación que se basa en la construcción de hiperplanos en un espacio multidimensional.

En la figura 5 se observa la representación de un hiperplano óptimo y un margen máximo que separa dos clases de figuras. Vapnick define el margen máximo como la distancia entre los dos hiperplanos, paralelos al hiperplano óptimo. En la figura 5 el margen es la distancia entre las líneas punteadas.

Asimismo Vapnick asegura que mientras más lejos esté el hiperplano de los puntos a los que clasifica, los resultados disminuirán el margen de error en la clasificación.

En dado caso que los atributos no pudieran separarse mediante un hiperplano en SVM, es posible transformar el espacio de los atributos, lo cual se denomina como núcleo (conocido como *kernel* en inglés).

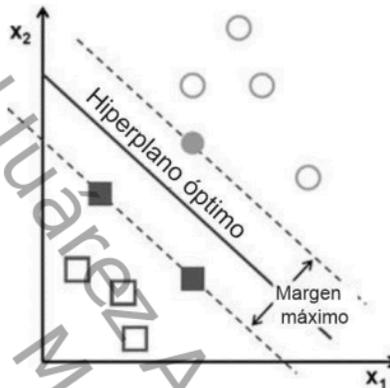


Figura 5. Representación de hiperplano óptimo en SVM

Existen varios núcleos que se pueden utilizar en los modelos SVM, los básicos son: núcleo lineal, núcleo polinomial, núcleo radial y núcleo sigmoideal. Es importante determinar qué tipo de núcleo es el indicado para cada clasificación que se va a realizar, ya que este es el que define el espacio de trabajo transformado donde se realizará el entrenamiento y clasificación del trabajo, asimismo se debe considerar la configuración (proceso de tuneo) de los parámetros dependiendo del tipo de núcleo [59]. La tabla 3 presenta los parámetros configurables de cada uno de los núcleos.

Tabla 3. Parámetros de núcleos

Núcleo	Parámetro
Lineal	C
Radial	C, sigma
Polinomial	C, degree, coef0

A continuación se presenta una descripción general acerca de los tipos de núcleos.

- **Núcleo lineal.** El núcleo lineal es recomendable utilizar cuando hay una gran cantidad de características puesto que el entrenamiento es más rápido, a diferencia de los otros núcleos, ya que solo se busca un parámetro. Los parámetros configurables son: *cost*.

Ecuación 1

$$K(x, y) = x^T y + C$$

Donde: x y y son vectores en el espacio de entrada y C es el parámetro de penalti.

- **Núcleo polinomial.** Ayuda a mejorar la precisión de clasificación mediante funciones polinomiales. Permite un aprendizaje a los modelos no lineales. Los parámetros configurables son: *coef0*, *degree*.

Ecuación 2

$$K(x, y) = (x^T y + \text{coef})^d$$

Donde: x y y son vectores en el espacio de entrada, d es el grado del polinomio, y $\text{coef} > 0$.

- **Núcleo radial.** Se puede utilizar para encontrar un conjunto de pesos para realizar un ajuste en problemas de curvas. El aprendizaje es equivalente a encontrar una superficie de alta dimensión dentro de un espacio que proporciona el mejor ajuste a los datos de entrenamiento. Los parámetros configurables son: *cost*, *gamma* y *sigma*.

Ecuación 3

$$K(x, y) = \exp(-\gamma \|x - y\|^2), \gamma > 0$$

Donde: x e y son vectores en el espacio de entrada, γ controla la desviación estándar de la función de Gauss.

- **Núcleo Sigmoidal.** Es uno de las más precisas para clasificar grandes números de datos, pero a su vez es más compleja. Es comúnmente encontrado en redes neuronales. Los parámetros configurables son: $coef0$.

Ecuación 4

$$K(x, y) = \tanh(\gamma x_i^T y_j + r)$$

Donde: x e y son vectores en el espacio de entrada.

Vapnik [60] afirma que el uso de la SVM en el análisis de emociones puede obtener resultados favorables en la clasificación de textos. Algunos de los principales trabajos en el campo de análisis de emociones utilizando la SVM se llevó a cabo por Pang [20]. De acuerdo a la literatura de SVM se puede observar que esta técnica tiene la capacidad de producir resultados con alta precisión en el proceso de clasificación. Sin embargo la principal desventaja es la complejidad que hace que sea difícil obtener una comprensión sólida de cómo funciona exactamente cuándo se compara con algunos algoritmos más simples como el algoritmo *Naïve Bayes*.

2.1.6.2 Naïve Bayes

El clasificador *Naïve Bayes* está basado en el teorema de Bayes [61] y en la premisa de independencia de los atributos dada una clase. Es uno de los métodos de aprendizaje supervisado más usados para el análisis de emociones, debido a su facilidad de implementación y a los porcentajes altos de exactitud en la clasificación de comentarios. Es un método de clasificación exento de reglas, usa la probabilidad matemática para encontrar el resultado más viable o los posibles clasificadores. La ventaja de este algoritmo es la simplicidad y rapidez para predecir las probabilidades posibles. Se basa en una técnica supervisada ya que se debe entrenar con casos anteriores para que el método sea capaz de predecir un futuro evento.

Ecuación 5

$$P(C_i|D) = \frac{P(C_i)P(D|C_i)}{P(D)}$$

Donde D es un documento, o en este caso comentario del conjunto de datos de entrenamiento, C_i es cada una de las posibles clases y $P(C_i|D)$ es la probabilidad

de que D pertenezca a la clase C_i . La clase seleccionada será la que maximice la probabilidad.

Se puede descomponer $P(C_i|D)$ según la premisa de independencia de los atributos,

Ecuación 6

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

Donde f_k son los términos del comentario y $P(f_k|C_i)$ es la probabilidad de ocurrencia del feature en la clase dada. Por lo tanto, de igual manera la clase seleccionada será la que maximice la probabilidad. Distintas implementaciones del algoritmo de *Naïve Bayes* difieren principalmente en la aproximación de $P(f_k|C_i)$. Una manera de calcular las probabilidades anteriores es a través del logaritmo, sumar en vez de multiplicar las probabilidades condicionales, así la probabilidad $P(f_k|C_i)$ es un peso que indica que tan bueno es el término f_k para predecir una clase C_i .

2.1.6.3 Bosques aleatorios (*Random Forest*)

El algoritmo de *Random Forest* introducida por Breiman y Cutler [62], se trata de un conjunto de árboles de decisión que da salida a un valor de predicción, se le clasifica como un método combinado de aprendizaje (en inglés *Ensemble Learning*). Cada árbol de decisión se construye mediante el uso de un subconjunto al azar de los datos de entrenamiento, asimismo se genera una gran cantidad de árboles que votan a favor de la clase más popular. El algoritmo *Random Forest* surge a partir de *bagging* [63] y los árboles de clasificación y regresión (*CART-split*). El rendimiento del algoritmo *Random Forest* es muy similar a la del *boosting* [64], con la ventaja de que facilita el entrenamiento, mejora la estabilidad, la precisión y además es una técnica robusta frente al sobreajuste (problema conocido en inglés como *overfitting*), el cual es el efecto de sobreentrenar el algoritmo con los datos de entrenamiento aprendiendo características muy específicas de los mismos y por lo tanto obteniendo resultados bajos con nuevos datos de entrenamiento.

La figura 6 muestra una representación del algoritmo *Random Forest* en la cual todos los árboles (x) son entrenados individualmente, posteriormente juntos votan en la optimización de un valor (y).

Random Forest es un clasificador que consiste en una colección de clasificadores con estructura de árbol ($h(x, \theta_k), k = 1, \dots$) donde la θ_k son vectores aleatorios independientes idénticamente distribuidas y cada árbol emite un voto unidad para la clase más popular en la entrada x .

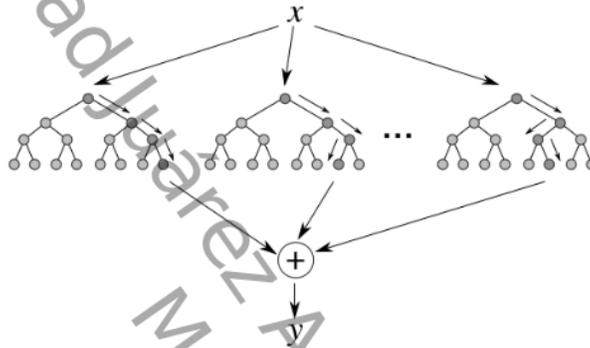


Figura 6. Representación del algoritmo Random Forest

Random Forest posee dos parámetros importantes que se describen a continuación:

- $mtry \in \{1, \dots, p\}$: el número de variables de entrada elegido al azar en cada división.
- $ntree$: el número de árboles.

Para lograr un desempeño alto en procesos de clasificación mediante algoritmos de aprendizaje automático, es importante realizar un proceso para la selección de términos distintivos en el proceso de clasificación.

2.1.7 Métodos para Selección de términos distintivos (*features*)

El proceso de selección de términos distintivos (denominadas *features*) se considera importante en el área de aprendizaje automático para realizar el entrenamiento de los algoritmos, trabajar con *features* irrelevantes pueden provocar resultados bajos en clasificación y predicción [65].

Existen métodos para la selección de *features* que se pueden utilizar para identificar y eliminar los atributos innecesarios, irrelevantes y redundantes de datos que no contribuyen a la mejora de un algoritmo o modelo predictivo. Asimismo la selección de *features* es preferible para reducir la complejidad del modelo. Existen tres clases generales para clasificar los métodos de selección de *features* en los algoritmos de aprendizaje: métodos de filtrado (*Filter Methods*), métodos de contenedor (*Wrapper Methods*) y métodos embebidos (*Embedded Methods*) [66].

- **Métodos de filtrado.** En este método se aplica una medida estadística para asignar una puntuación a cada término [65, 67]. Las *features* se seleccionan por una medida o puntuación obtenida individualmente en cada función. Aquellos términos con baja puntuación se descartan y sólo un subconjunto de los términos seleccionados como *features* se da como entrada para el algoritmo de aprendizaje. Una de las ventajas de filtros es que son computacionalmente eficientes para tratar con altos conjuntos de datos dimensionales, además de que facilita el proceso de calcular. Una desventaja es que los filtros no tienen en cuenta las dependencias de funciones.
- **Métodos de contenedor.** Los métodos de contenedor consideran la selección de un conjunto de *features* como un problema de búsqueda, donde se preparan diferentes combinaciones y posteriormente se evalúan y comparan con otras combinaciones. Se utiliza un modelo predictivo para evaluar las combinaciones de *features* y se asigna una puntuación basada en la precisión del modelo [65, 68]. Se usa un modelo predictivo para evaluar una combinación de *features* y asignar una puntuación basada en la precisión del modelo. Una desventaja es que estos métodos son computacionalmente intensivos y requieren el entrenamiento de un modelo para cada subconjunto de *features* potenciales [69, 70]. En la figura 7 se presenta el modo de trabajo de un método contenedor que usa el algoritmo de aprendizaje como una caja negra al evaluar subconjuntos de *features*.

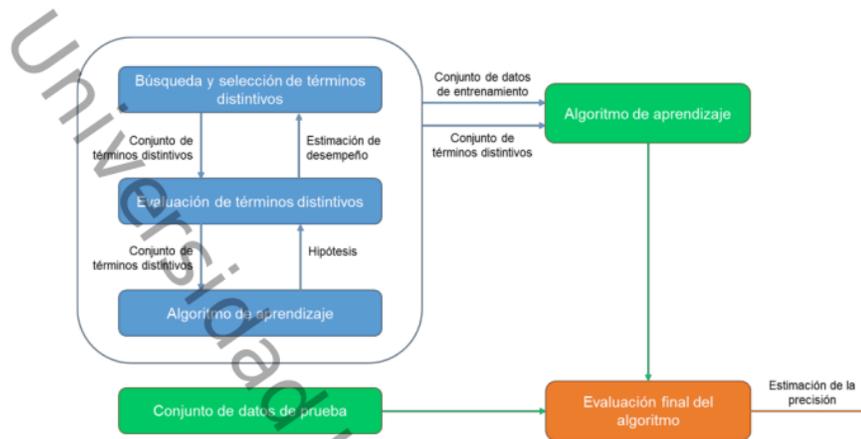


Figura 7. El método contenedor en la selección de features [65, 66].

- Métodos embebidos.** Este tipo de métodos aprenden cuales son las mejores *features* que contribuyen a mejorar la precisión del modelo mientras el modelo está siendo creado [71, 72]. Sin embargo, como el número de *features* aumenta el número de posibles subconjuntos de *features* crece exponencialmente. Este tipo de métodos a diferencia de los métodos contenedor son computacionalmente menos intensos. La figura 8 muestra el modo de trabajo de métodos embebidos que incorporan la búsqueda del subconjunto de *features* y evaluación mientras construyen el clasificador.



Figura 8. Métodos embebidos en la selección de features [66].

Uno de los métodos viables para selección de *features* es utilizando el algoritmo *Random Forest*, debido a que los parámetros a configurar son pocos y puede trabajar sin problema con conjuntos de datos que contengan números y categorías. En este trabajo se utilizan métodos de filtrado para la selección de

features, siendo *Random Forest Variable Importance Measure* (VIM por sus siglas en inglés) el método elegido [73, 74] como guía para la selección de las *features*.

El desempeño de VIM ha sido criticado por algunos autores como Verikas [75], quien manifiesta que el algoritmo k-NN proporciona comparables y, a veces mejores resultados con respecto a la precisión de la predicción de *features*. El algoritmo k-NN puede modelar fácilmente los problemas no lineales y sólo hay un parámetro de ajuste. Sin embargo, el algoritmo k-NN puede tener dificultades con el uso de categorías ya que no hay ningún método general sobre cómo asignar las distancias entre éstas [76]. Asimismo otros autores [77] manifiestan haber obtenido resultados satisfactorios aplicando el método VIM para la selección de *features*, comparando el desempeño obtenido con otros algoritmos como: k-NN, SVM y LDA.

Una vez obtenidas las *features* para entrenar los algoritmos es necesario medir el desempeño de los mismos en el proceso de clasificación, una de las métricas usadas para este propósito, es la curva ROC [78], también denominada análisis ROC, la cual en un principio se aplicó en áreas como la medicina, la radiología y la psicología. No obstante debido a la fiabilidad de este método se comenzó a aplicar en áreas como aprendizaje automático y minería de datos [79].

La curva ROC se compone de otras métricas como: la matriz de confusión, la sensibilidad, la especificidad, valores predictivos positivos y valores predictivos negativos, las cuales se describen en la siguiente sección.

2.1.8 Métricas para evaluar el desempeño de clasificación

- **Matriz de confusión**

La matriz de confusión (presentada en la tabla 4) también conocida como tabla de contingencia o matriz de clasificación, muestra el rendimiento de la predicción en clasificación del algoritmo de aprendizaje automático utilizado. Esta tabla se deriva de la comparación de los resultados de un modelo predictivo contra los valores reales. Las filas de la matriz de confusión representan los resultados de valores de predicción, mientras que las columnas representan los valores reales.

La matriz de confusión se considera el punto de partida para el cálculo de la medición del desempeño de un modelo predictivo.

Tabla 4. Matriz de confusión para clasificación binaria

Valores predictivos	Valor actual	
	VP	FP
	FN	VN

Donde: VP = verdaderos positivos, FP = falso positivo, FN = falso negativo, VN = verdadero negativo.

- **Precisión (ACC)**

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

Ecuación 7

$$ACC = \frac{(VP + VN)}{(VP + FN + VN + FP)}$$

Donde: VP = verdaderos positivos, FP = falso positivo, FN = falso negativo, VN = verdadero negativo.

- **Precisión balanceada (WACC)**

1 Es una mejor estimación del rendimiento del clasificador cuando está presente en un conjunto de datos de una distribución desigual de las dos clases.

Ecuación 8

$$WACC = \frac{\frac{VP}{VP + FN} + \frac{VN}{FP + VN}}{2}$$

Donde: VP = verdaderos positivos, FP = falso positivo, FN = falso negativo, VN = verdadero negativo.

- **Margen de error**

El margen de error mide la proporción de errores del clasificador en una serie de casos. Es el complemento de la precisión, es decir 100% - precisión. La tasa de error mide la proporción de errores de un clasificador sobre un conjunto de instancias. Es el complemento de la precisión [80].

- **Sensibilidad**

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

Ecuación 9

$$VRP = \frac{VP}{(VP + FN)}$$

Donde: VP = verdaderos positivos, FN = falso negativo, VN = verdadero negativo.

- **Especificidad**

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

Ecuación 10

$$SPC = \frac{VN}{(FP + VN)}$$

Donde: FP = falso positivo, VN = verdadero negativo.

- **Curva ROC**

El análisis ROC (Receiver Operating Characteristic) curva mide el rendimiento del clasificador binario. Representa la razón o ratio de verdaderos positivos (VPR) frente a la razón o ratio de falsos positivos (FPR). El rendimiento se mide por la AUC (Área Bajo la Curva). AUC a partir de una curva ROC oscila [0,1], donde 1 es una clasificación perfecta.

Ecuación 11

$$VRP = \frac{VP}{VP + FN}$$

Ecuación 12

$$FPR = \frac{FP}{FP + VN}$$

La curva ROC puede tomar diferentes formas dependiendo de los valores obtenidos para sensibilidad y especificidad, un extracto de las posibilidades se muestra en la figura 9, donde el primer par de imágenes muestran el cruce de las funciones de densidad, lo cual genera una curva ROC habitual, mientras que el segundo par de gráficas corresponden a un experimento catalogado como “no informativo” debido a que el clasificador no puede diferenciar una clase de otra. Por otra parte, el tercer par de imágenes indican un conjunto de funciones de densidad simétricamente distantes, indicando que, de acuerdo a los datos de especificidad y sensibilidad obtenidos, se logra un experimento perfecto.

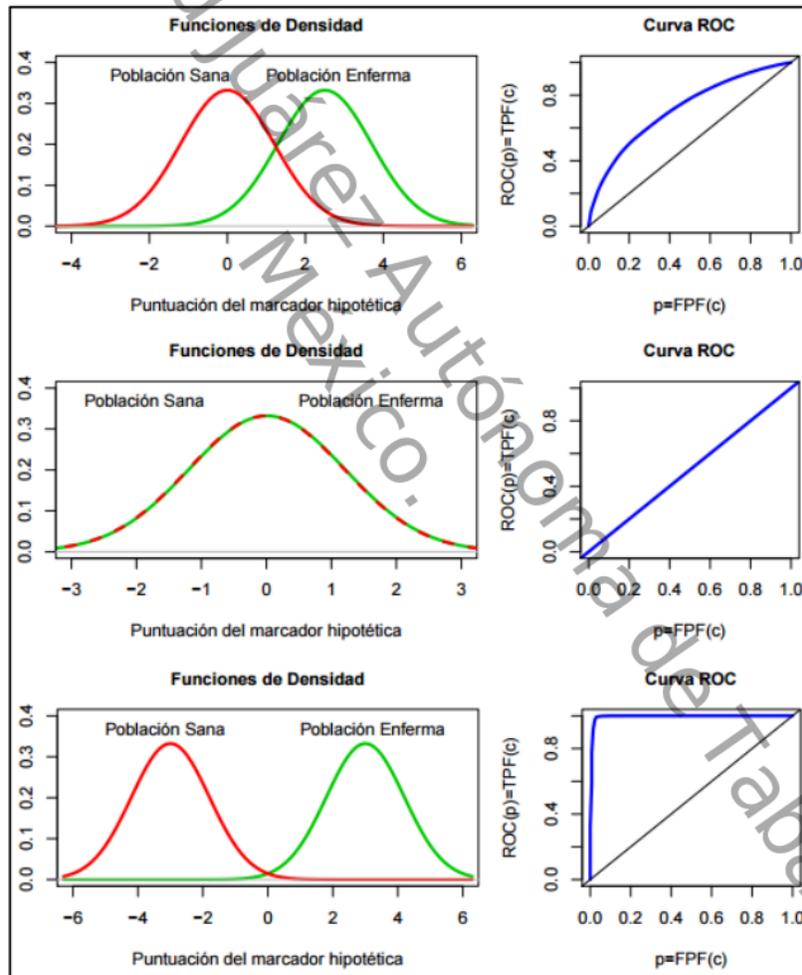


Figura 9. Tipos de curva ROC

- **Probabilidad del suceso contrario en un suceso condicionado**

Permite calcular la probabilidad de un suceso contrario a partir de probabilidades condicionadas.

Ecuación 13

$$P(\bar{A}|B) = 1 - P(A|B)$$

- **Teorema de la probabilidad total**

Permite calcular la probabilidad de un suceso a partir de probabilidades condicionadas.

Ecuación 14

$$P(B) = P(B|E_1)P(E_1) + P(B|E_k)P(E_k)$$

- **Validación cruzada de 10 iteraciones**

En el método de validación cruzada de 10 iteraciones (*10-fold cross-validation* en inglés) los datos se dividen en 10 subconjuntos. Uno de los subconjuntos se utiliza como datos de prueba y el resto como datos de entrenamiento. El proceso de validación cruzada es repetido durante 10 iteraciones, con cada uno de los posibles subconjuntos de datos de prueba [81, 82], finalmente se realiza la media aritmética de los resultados de cada iteración para obtener un único resultado. Este método es muy preciso puesto que se evalúa a partir de 10 combinaciones de datos de entrenamiento y de prueba, pero aun así tiene una desventaja, ya que es lento desde el punto de vista computacional.

En este trabajo se utiliza el método de validación cruzada de 10 iteraciones para realizar el entrenamiento de los algoritmos SVM y *Random Forest*. Asimismo se hace uso de las métricas explicadas en la presente sección con el fin de medir el desempeño de los mismos en el proceso de predicción y clasificación.

En la siguiente sección se presenta una revisión de algunos trabajos relacionados al Modelo SocialMining para ofrecer diferentes perspectivas de investigaciones dedicados a la mejora de la educación aplicando diferentes métodos y herramientas, además de algoritmos de aprendizaje automático.

2.2 Estado del Arte

En esta sección se describen algunos trabajos relacionados que aplican análisis de emociones con fines de mejorar el proceso de enseñanza – aprendizaje de los estudiantes. Asimismo se introducen trabajos que no están enfocados al contexto de la educación, pero utilizan algoritmos de aprendizaje automático. Estos trabajos se incluyen con la finalidad de conocer los resultados en precisión de los algoritmos de aprendizaje automático utilizados.

2.2.1 Trabajos relacionados enfocados a la mejora de la educación

Trabajo 1. Learning sentiment from students' feedback for real-time interventions in classrooms

Altrabsheh [83] sostiene que la aplicación del análisis de emociones en la educación contribuye para conocer la situación emocional del estudiante y hacer frente a problemas académicos que pueden derivarse de la confusión y aburrimiento. Por lo cual propone un sistema para analizar los comentarios de los estudiantes.

Los datos utilizados para el estudio se conforman por 1036 comentarios en inglés de estudiantes de la Universidad de *Portsmouth*, los cuales posteriormente fueron etiquetados en tres clases (positivo, negativo y neutral) por expertos del área. Los algoritmos de aprendizaje automático aplicados en el estudio son: Naïve Bayes, Naïve Bayes complementario, entropía máxima y SVM con núcleo radial. Asimismo el método de validación cruzada de 10 iteraciones es utilizado para realizar el entrenamiento de los algoritmos mencionados.

En los resultados el algoritmo SVM obtiene la más alta precisión con un valor de 94%, seguido por el clasificador bayesiano complementario, el cual obtiene un 84%.

Trabajo 2. SentBuk: Sentiment analysis for e-learning environments

Martín [84] propone una aplicación en Facebook denominada SentBuk para extraer automáticamente los comentarios que los estudiantes publican y clasificarlos como positivos, negativos y neutrales. La finalidad de este estudio es brindar al profesor una retroalimentación sobre la emoción que predomina en cada uno de los estudiantes.

Para este estudio se extraen 1,000 publicaciones en español, las cuales son clasificadas en positivo, negativo y neutral por expertos del área. Se utiliza un diccionario de emoticonos y términos etiquetados con base a una orientación semántica.

En los resultados realizan una comparación entre las publicaciones etiquetadas por la aplicación *SentBuk* contra las publicaciones etiquetadas manualmente. Los resultados obtenidos por *SentBuk* son un 96.78% de precisión para comentarios positivos y un 94.59% de precisión para comentarios negativos.

Trabajo 3. Teaching Senti-Lexicon for Automated Sentiment Polarity Definition in Teaching Evaluation

Chakrit Pong-inwong [11] construye un léxico para el idioma tailandés y una herramienta automatizada para la detección de polaridades en un conjunto de comentarios de evaluación docente. Extrae comentarios de estudiantes y realiza un proceso de limpieza de los mismos, asimismo establece un peso para cada término del conjunto de comentarios. Además define 5 atributos distintivos positivos y 5 negativos, del conjunto de comentarios, los cuales son utilizados como *features*.

Utiliza el algoritmo de *Naïve Bayes* para clasificar los comentarios y evaluar el modelo propuesto.

La precisión de la técnica se comparó contra SVM e ID3 (algoritmo de árbol de decisión creado por J. R. Quinlan [85]), obteniendo los siguientes resultados: la técnica con mayor logro es SVM con un 97.71% de precisión, *Naïve Bayes* un 95.42% de precisión y el algoritmo *ID3* con un 92% de exactitud al efectuar clasificación.

Trabajo 4. Sentiment Analysis and E-learning: A proposal

En este trabajo, Colace [86] propone la adopción, en el ámbito de la educación a distancia de un enfoque probabilístico basado en la asignación de un modelo generativo llamado Análisis Lineal Discriminante (LDA) como capturador de sentimiento. En LDA, cada documento puede ser visualizado como una composición de varios temas. Con el uso del enfoque LDA en un conjunto de documentos que pertenecen a una misma área de conocimiento, se puede obtener de forma automática un gráfico mixto de términos. Un gráfico de éste tipo

contiene un conjunto de pares de palabras equilibradas, que son importantes para la clasificación de polaridades.

Para el enfoque propuesto se aplicó un conjunto de comentarios de películas, utilizando el 25% como entrenamiento y el 75% como prueba. El principal propósito de este experimento fue evaluar el desempeño de los métodos y hacer una comparación de los mismos. Utiliza algoritmos como SVM, *Naïve Bayes* y entropía máxima, así como el enfoque LDA.

En los resultados SVM obtiene un 82.90% de precisión, *Naïve Bayes* obtiene un 81.50% de precisión, entropía máxima obtiene un 81.00% de exactitud y por último LDA obtiene un 88.50% de precisión en el proceso de la clasificación.

Trabajo 5. Identifying e-Learner's opinion using automated Sentiment Analysis in e-Learning

Bharathisindhu [8] propone una aplicación de análisis de emociones que consiste en extraer los comentarios de 100 usuarios del sitio web funtionspace.org, así como algunos atributos (páginas visitadas, propiedades, noticias, grupos, conversaciones, mensajes). Para efectuar la clasificación utiliza el algoritmo *Naïve Bayes*. Debido a que es un proyecto aun en etapa de diseño y desarrollo, los autores concluyen en que el resultado del sistema propuesto apoyará al docente a identificar las opiniones de los estudiantes de los portales *e-Learning*, que se encuentren cursando.

2.2.2 Trabajos relacionados que aplican aprendizaje automático

Trabajo 1. M-Learning Sentiment Analysis with Data Mining Techniques

Jebaseeli [87] utiliza un corpus que contiene 300 opiniones extraídas de www.market.android.com, en el contexto de sistemas *M-learning (Mobile e-Learning)*, divididas proporcionalmente en comentarios positivos, negativos y neutrales. Se consideran las técnicas de *Naïve Bayes*, vecino más cercano (*k-NN*) y *Random Forest* para efectuar las clasificaciones y comparar los resultados. La metodología propuesta en dicho trabajo se presenta a continuación.

- Pre-procesamiento. Las opiniones se extraen en formato de texto plano y se cargan a una herramienta computacional estadística que tiene la capacidad de contabilizar la aparición de cada palabra y eliminar las palabras no relevantes, mediante el procedimiento de *stemming* y el uso de *stop words*.
- Valor singular de descomposición. Este procedimiento es utilizado para encontrar la importancia de una palabra. El procedimiento se entrena con la ayuda del conjunto de palabras previamente puntuadas (de forma manual), ayudándose también de parámetros como la frecuencia de la palabra.
- Pre-procesamiento manual. En esta etapa se eliminan los comentarios que contengan menos de 15 o más de 85 caracteres, con la finalidad de definir márgenes que mejoren la clasificación en la siguiente etapa.

El 80% de las opiniones procesadas fueron usadas para entrenamiento y el resto fueron usadas para pruebas. Los resultados arrojados por las técnicas de clasificación posicionan a *Random Forest* en primer lugar, con un 60% de precisión, mientras *Naïve Bayes* se encuentra en segundo lugar con un 55% de precisión y por último *k-NN* obtiene un 52.67% de precisión.

Trabajo 2. Comparative Study of Classification Algorithms used in Sentiment Analysis

Gupte [88] aborda una comparativa en precisión de clasificación entre las técnicas *Naïve Bayes*, entropía máxima, árboles de decisión y *Random Forest*, la seleccionan estas técnicas es debido a su conocido desempeño en la aplicación en el área de análisis de emociones.

El estudio muestra la evaluación de las técnicas respecto a diferentes características, indicando que el clasificador más adecuado depende del contexto de cada situación.

Los resultados de precisión en un periodo de tiempo establecido fueron variables para *Naïve Bayes*, consistentes para entropía máxima e incremental para árboles de decisión y *Random Forest*. De igual manera en este trabajo se concluye que el tiempo requerido para entrenar al clasificador es poco en *Naïve Bayes*, moderado en entropía máxima, alto en árboles de decisión y recurrente en *Random Forest*.

Trabajo 3. Sentiment analysis using product review data

Fang [89] propone e implementa un algoritmo para la identificación de frases negativas, así como el cálculo de la puntuación afectiva. Se ejecutan dos experimentos de categorización de polaridad en nivel sentencia y nivel opinión, aplicando los modelos de clasificación SVM, *Naïve Bayes* y *Random Forest*.

Para el cálculo de la puntuación afectiva se seleccionaron 11,478 palabras y 3,023 frases afectivas. Cada palabra se repetía al menos 30 veces en el corpus de comentarios. El valor del puntaje afectivo para palabras positivas se encontraba localizado en los comentarios con más de tres estrellas (en una escala 1 a 5) y los puntajes altos para palabras negativas se reflejaban en los comentarios con menos de tres estrellas. El sistema separa las palabras de una oración y las clasifica. Si el número de palabras positivas es mayor, entonces la oración tiene una polaridad positiva, actuando en caso contrario para las oraciones con mayor cantidad de palabras negativas.

Como resultado, el corpus se etiqueta en tres clases: positiva, negativa y neutral. De acuerdo a los experimentos realizados, los modelos SVM y *Naïve Bayes* mantienen el mismo rendimiento (0.8) al clasificar, quedando por encima del modelo *Random Forest*.

Trabajo 4. Multilayer sentiment analysis model base on Random Forest

Liu [90] realiza una comparación del efecto de la clasificación entrenada mediante modelos tradicionales de aprendizaje automático. Se obtienen las principales

características del texto a través de segmentación, *stemming* y procesamiento gramatical utilizando las herramientas de Stanford⁸ destinadas a dicho propósito.

Los datos utilizados para el experimento son estados emocionales publicados en redes sociales, reuniendo un total de 3,000 comentarios emocionales, que se dividen en conjunto de entrenamiento (64%), el conjunto de validación (16%) y el conjunto de prueba (20%).

En los resultados se obtiene que el mejor modelo de clasificación es SVM con una precisión de 0.84, mientras que *Random Forest* obtiene una precisión de 0.83 y el modelo de regresión logística obtiene solo 0.82.

Trabajo 5. Twitter Catches The Flu: Detecting Influenza Epidemics using Twitter

Aramaki [91] utiliza un corpus de comentarios extraídos de twitter, con la característica de que cada comentario tenía que incluir la palabra influenza o *flu* por lo menos una ocasión. Para el experimento recolecta 300 millones de tweets con la palabra influenza que corresponden a los años del 2008 al 2010, el 42% de los tweets que incluyen la palabra influenza son negativos, y el resto son positivos. El corpus de entrenamiento seleccionado consta de 5000 tweets, con los cuales se construye el corpus de entrenamiento y prueba.

Los autores presentan una propuesta de un clasificador basado en aprendizaje automático para filtrar los tweets negativos, mediante el siguiente procedimiento:

- Recolección de 300 millones de tweets (entre nov 2008 y jun 2010).
- Separación de corpus de entrenamiento (5,000 tweets) y de prueba.
- Construcción de un corpus de entrenamiento etiquetado (etiquetas positivas y negativas).
- Selección de técnicas de clasificación.
- Desarrollo de algoritmo de cada una de las técnicas.
- Ejecución de pruebas.
- Comparación de resultados.

⁸ <http://nlp.stanford.edu/software/lex-parser.shtml>

En los resultados se obtienen parámetros estadísticos con base a dos características que se evaluaron: la precisión y el tiempo de ejecución, la SVM utilizando un núcleo polinomial alcanza el primer lugar en clasificación con 75.6% de precisión, en tiempo de 13.25 segundos de ejecución. La SVM con núcleo RBF obtiene un 73.8% en un tiempo de 92.72 segundos y *Random Forest* consigue un 72.9% de precisión en la clasificación, en un tiempo de 38.683 segundos.

La tabla resume los trabajos relacionados presentados en esta sección, además de presentar algunos otros más. Haciendo un enfoque al objetivo del trabajo, las técnicas utilizadas y el conjunto de datos con el cual se trabaja.

Tabla 5. Resumen de trabajos relacionados

Autor	Objetivo	Técnica utilizada	Datos utilizados
Altrabsheh [83]	Aplicación para conocer la situación emocional del estudiante.	SVM, <i>Naïve Bayes</i> , <i>Naïve Bayes complementario</i> .	1036 comentarios en inglés de estudiantes de la Universidad de Portsmouth
Martín [80]	Aplicación en Facebook para brindar al profesor retroalimentación	Se utiliza un diccionario de emoticonos y términos etiquetados.	1000 comentarios en español
Chakrit Pong-inwong [11]	Herramienta para la detección de polaridades comentarios evaluación docente	Utiliza un léxico de tailandés, así como los algoritmos <i>Naïve Bayes</i> , SVM e ID3.	175 comentarios.
Colace [82]	Propone utilizar el modelo LDA para capturar el sentimiento en comentarios de cursos de educación a distancia.	Utiliza algoritmos como SVM, <i>Naïve Bayes</i> y entropía máxima, así como el enfoque LDA.	Utiliza un conjunto de comentarios de películas.
Bharathisindhu [8]	Propone una aplicación de análisis de emociones para comentarios del sitio web funtionspace.org	Utiliza el algoritmo de <i>Naïve Bayes</i> y SVM.	Recolecta comentarios de 100 usuarios del sitio web funtionspace.org

Autor	Objetivo	Técnica utilizada	Datos utilizados
Jebaseeli [87]	Propone una aplicación para el contexto de sistemas M-learning	Utiliza <i>Naïve Bayes</i> , <i>kNN</i> y <i>Random Forest</i> .	Utiliza 300 comentarios de www.market.android.com
Gupte [88]	Compara técnicas de aprendizaje automático.	Utiliza <i>Naïve Bayes</i> , entropía máxima, árboles de decisión y <i>Random Forest</i> .	No especifica los datos utilizados.
Fang [89]	Identifica frases negativas, y calcula la puntuación afectiva.	Utiliza SVM, <i>Naïve Bayes</i> y <i>Random Forest</i> .	3,023 frases afectivas
Liu [90]	Compara el desempeño de los algoritmos SVM y <i>Random Forest</i> en procesos de clasificación.	Utiliza SVM y <i>Random Forest</i> .	3000 comentarios de redes sociales.
Aramaki [91]	Propone un clasificador basado en aprendizaje automático para filtrar tweets negativos.	Utiliza SVM con núcleo polinomial y RBF.	5000 comentarios de Twitter

En esta sección se presentaron algunos trabajos considerados importantes debido a su relación con el Modelo SocialMining, como se pudo observar cada uno de éstos se enfoca en un contexto particular, razón por la cual se considera necesario realizar propios conjuntos de datos para el proceso de entrenamiento. Los datos son obtenidos por distintos medios como las redes sociales y páginas de internet. De igual manera en los trabajos relacionados se definen o adecuan distintas metodologías para llevar a cabo los estudios y experimentos. Asimismo utilizan diferentes herramientas y algoritmos, por lo cual los resultados obtenidos varían de un estudio a otro, sin embargo se puede observar que la mayoría de las investigaciones consideran factible alcanzar un 80% en precisión para procesos de clasificación.

En el Modelo SocialMining, se utilizan los algoritmos de *Naïve Bayes*, SVM y *Random Forest*, además se crean propios corpus, debido a que en la literatura

aún no se reporta un corpus o léxico enfocado a la evaluación del desempeño docente, razón por la cual es necesario la creación de éste con la finalidad de mejorar los resultados en la clasificación de comentarios. En la siguiente sección se presenta el Modelo SocialMining, el cual posee dos enfoques, SocialMining aplicando *Naïve Bayes* y subjetividad, y SocialMining aplicando SVM y *Random Forest*.

Universidad Juárez Autónoma de Tabasco.
México.

Capítulo III

Modelo SocialMining propuesto

3.1 Modelo SocialMining aplicando *Naïve Bayes* y Subjetividad

3.1.1 Descripción del Modelo

El Modelo SocialMining aplicando *Naïve Bayes* y Subjetividad, propuesto en este trabajo, se compone de tres fases que se ilustran en la figura 10.

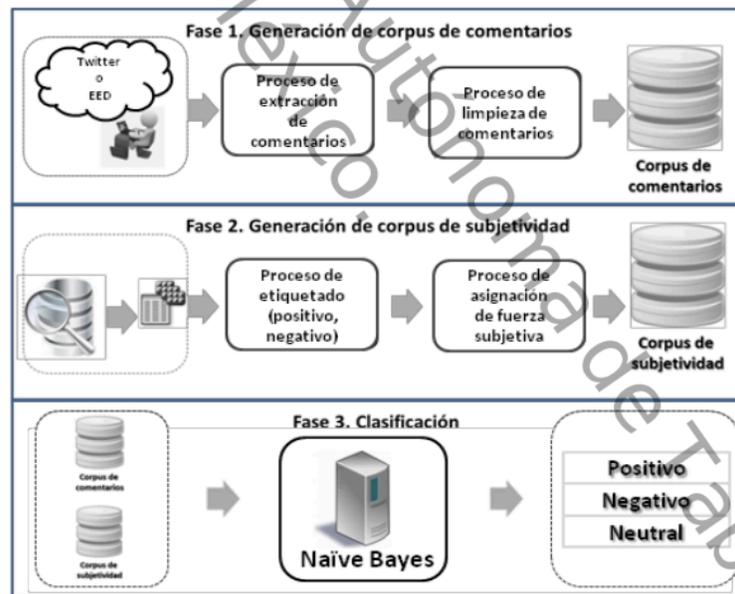


Figura 10. Arquitectura general del Modelo de SocialMining

1 Fase 1. Generación de corpus de comentarios.

En esta fase se extraen los comentarios de los alumnos realizados en Twitter y la EED. La descarga de los comentarios de Twitter se realiza mediante el uso de la API⁹ de la misma red social. En cambio los comentarios de la EED son obtenidos por medio de un archivo de la base de datos del sistema.

Una diferencia que se observa en los comentarios que se obtienen de ambos medios es que la EED no tiene restricciones en el número de caracteres, por lo cual es común encontrar comentarios con ruido a diferencia de Twitter, es decir comentarios con faltas de ortografía, mala redacción y uso de palabras o caracteres que no necesariamente aportan a la evaluación docente, por ejemplo:

“Creo que la maestra a veces se equivoca en reglas que son básicas en matemáticas. Además no me parece que su manera de explicar es excelente ni de 10 porque SIEMPRE SE ENOJA MUCHO tambien muchas veces le da muchas vueltas al asunto de lo que debería ;(y no siempre se va por el camino más fácil y practico. Sin embargo reconosco su disponibilidad y su motivación que tiene puesto que jamas nos ha tratado de manera descortés y es accesible.”

Debido a que Twitter cuenta con la restricción de usar solo 140 caracteres para realizar una publicación y además cuenta con un corrector ortográfico, se observa que la mayoría de las veces los estudiantes comentan de manera concreta y puntual. No obstante también es posible encontrar el mismo tipo de ruido de la EED en los comentarios de Twitter. Con el fin de minimizar el ruido y conformar el corpus de comentarios, se realiza un proceso de limpieza. Las actividades del proceso de limpieza se muestran en la figura 11.



Figura 11. Actividades del proceso de limpieza

⁹ <https://apps.twitter.com/>

- **Revisión de ortografía.** En este proceso se hace una revisión de cada uno de los comentarios para corregir faltas de ortografía de acuerdo con el Diccionario de la Real Academia Española, en caso de que hubiera una palabra mal escrita.
- **Proceso de tokenización.** Consiste en dividir los componentes (palabras) de un comentario en piezas, los cuales pasan a ser denominados como *tokens*. Por ejemplo el comentario: “Me gustaría que el profesor explicara más ejemplos en clase”, al pasar por el proceso de tokenización se representa de la siguiente manera: “Me”, “gustaría”, “que”, “el”, “profesor”, “explicara”, “más”, “ejemplos”, “en”, “clase”.
- **Eliminación de números.** Debido a que en este trabajo los números no son considerados relevantes en un comentario, son eliminados.
- **Eliminación de *stop words*.** Las palabras vacías comúnmente conocidas en el idioma inglés como *stop words*, son palabras sin significado como artículos, pronombres, preposiciones, entre otras que son filtradas para evitar el problema de la dimensionalidad y reducir el tiempo de procesamiento del texto. Algunas de las palabras consideradas *stop words* en este trabajo son: la, los, el, materia, profesor, escuela, entre otros.
- **Conversión a minúsculas.** En este proceso cada uno de los términos del comentario es convertido a minúsculas para facilitar el proceso de comparación.
- **El proceso de *stemming*** es un proceso de normalización lingüística, útil para reducir una palabra a su lema o raíz (*stem* en inglés), después de eliminar sus afijos, no es necesario que tenga un significado y se usa para facilitar el proceso de búsqueda y comparación de texto, en este proceso se aplica la herramienta *Snowball*, la cual según Vilares [92] es una de las mejores cuando se trata de textos en español. Un ejemplo de aplicar *stem* es *perr*, para las palabras: perros, perrito, perro.

Una vez terminado el proceso de limpieza se procede a conformar el corpus de comentarios. Con lo cual se completa la fase 1 del Modelo SocialMining ilustrado en la figura 9.

Fase 2. Generación de corpus de subjetividad

En esta fase se conforma el corpus de subjetividad basado en la metodología usada por Riloff [93], quien realizó un corpus de este tipo para el idioma inglés constituido por 6518 términos. En esta investigación todos los comentarios están escritos en el idioma español, asimismo el contexto de aplicación cambia, ya que el Modelo SocialMining se enfoca en la evaluación docente, de igual manera se incluyen términos usuales por la región donde se aplica, denominados coloquialismos. Por esta razón es importante realizar un análisis de los comentarios para identificar las palabras o términos distintivos más usados que apoyen a la clasificación de un comentario en positivo o negativo. Cabe resaltar que los términos usados para el corpus de subjetividad no necesariamente son aquellos que se repiten en una mayor cantidad de veces, sino aquellos que denotan una emoción, por ejemplo: excelente, fomenta, facilita, creativa, aburrida, impaciente, tolerante, entre otros términos.

Fase 3. Clasificación

En esta fase los datos de entrada como puede observarse en la figura 9 son los corpus obtenidos en las anteriores fases, el corpus de comentarios y el corpus de subjetividad. Para el proceso de clasificación en el Modelo SocialMining se aplica el algoritmo *Naïve Bayes* adaptado con base al método propuesto por Jurka [94], quien adapta el algoritmo e incluye el cálculo de un score de cada uno de los términos del comentario que se encuentren en el corpus de subjetividad, con base en la ecuación 15.

Ecuación 15

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

Donde: *scorefs* toma el valor de la fuerza subjetiva que contiene el término, 0.5 para los términos etiquetados con *strongsubj* y 1.0 para los términos etiquetados como *weaksubj*, la variable *prior* es valor numérico de la probabilidad con valor de 1.0 y por último la variable *count* representa la cantidad de veces que

aparecen los términos negativos o los términos positivos, dependiendo del término.

En el proceso de clasificación existe una función de comparación denominada *match*, realizada mediante la función de *match* en R [95], el cual es un lenguaje de programación dedicado al análisis de datos científicos o estadísticos. La función *match* cual realiza una comparación entre los términos del corpus de subjetividad y los términos del corpus de comentarios, el cual es previamente convertido en una matriz para la ejecución de este proceso. El algoritmo para llevar a cabo el proceso de clasificación se describe en la figura 11.

```

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

```

Figura 12. Algoritmo del proceso de clasificación

Los datos usados y el diseño experimental se describen en la siguiente sección.

3.1.2 Datos y diseño experimental

3.1.2.1 Datos

Los datos analizados en este estudio se recolectaron utilizando dos instrumentos: Twitter y la encuesta de evaluación docente (EED).

En Twitter se analizaron alrededor de 800 comentarios de un grupo de 34 estudiantes de la carrera de Ingeniería en Sistemas que evaluó a sus docentes durante un ciclo escolar. Los estudiantes evaluaron a sus profesores desde

cuentas personales o anónimas en Twitter. Utilizaron como *hashtag* (ver tabla 6) la clave de la asignatura para identificar correctamente al profesor a evaluar.

La EED es un sistema informático de la UPA que se aplica dos veces en un periodo escolar. Los datos de la EED usada en este trabajo corresponden a la evaluación realizada por 1,505 estudiantes de la UPA a 229 profesores que impartieron clases desde el segundo periodo escolar del 2015 al primer periodo del 2016, obteniendo 9,054 comentarios.

Tabla 6. Relación de hashtag para evaluación docente en Twitter

Asignatura	Hashtag	Asignatura	Hashtag
Inglés	#Englishsei	Informática estratégica	#InfoEsei
Programación Lineal	#ProgLsei	Administración del conocimiento	#Aknowsei
Calidad de software	#CSWsei	Planeación de Recursos Materiales	#MRPsei
Finanzas para Ingenieros	#Finanzassei	EGEL	#EGIS #EGDB

Mediante Twitter y la EED se recolectaron 9854 comentarios. En total participaron 21 grupos de estudiantes de seis carreras de ingeniería. Cabe resaltar que todos los estudiantes pertenecen a la misma institución, el perfil de carrera es Ingeniería (pregrado), las edades oscilan entre 18 y 22 años y además poseen preferencias hacia el uso de medios y herramientas tecnológicas.

De los 9854 comentarios recolectados de Twitter y la EED se consideró solo aquellos comentarios libres de ruido o spam (caracterizado en este estudio como textos con caracteres raros, espacios vacíos sin opinión o comentarios sin relación a la evaluación docente), quedando un total de 5002 comentarios a analizar, los cuales conformaron el denominado corpus de comentarios. Una vez obtenido el corpus de comentarios fue necesario realizar un proceso de etiquetado manual con la participación profesores de la institución, quienes identificaron 2,696 comentarios negativos y 2,306 comentarios positivos, cabe resaltar que en este proceso de etiquetado manual no se consideraron comentarios clasificados como neutrales. En este trabajo un comentario positivo es aquel que denota gusto o interés del alumno por la clase o el profesor, un comentario negativo denota disgusto o aburrimiento y un comentario neutral no denota ninguna emoción.

Por otro lado, con el objetivo de evaluar el desempeño del Modelo SocialMining con datos desconocidos por éste, se conformó un segundo corpus de comentarios. Para ello, se aplica el muestreo por conveniencia y se eligen dos grupos de 5° cuatrimestre de la carrera de Ingeniería en Sistemas de la misma institución, a quienes se les pidió realizar la evaluación docente de sus 11 profesores actuales. El nuevo corpus de comentarios para el caso de estudio, se conformó por un total de 623 comentarios, de los cuales 304 se etiquetaron manualmente como comentarios negativos, 305 como comentarios positivos y 14 comentarios como neutrales.

3.1.2.2 Diseño experimental

De acuerdo a los 5002 comentarios obtenidos en Twitter y la EED explicado en la sección de Materiales y Métodos, una vez etiquetado el corpus de comentarios se procede a realizar un análisis mediante técnicas como la nube de palabras en el programa R [95] con el fin de determinar los términos relevantes o *features* a considerar para la conformación del corpus de subjetividad.

La figura 13 presenta dos nubes de palabras, en donde la primera (a) se obtiene del corpus de comentarios original y la segunda (b) se obtiene de corpus de comentarios que ha pasado por el proceso de limpieza y omite las *stop words*, sustantivos y términos considerados sin importancia, de esta manera la nube (b) denota términos distintivos que pueden apoyar al proceso de clasificación como: debería, interesantes, amable, permite, apoya, paciencia, conoce entre otras.



Figura 13. Nube de palabras general (a), nube de palabras limpia (b).

Una vez identificados los términos distintivos y agregarlos al corpus de subjetividad, se etiqueta a cada uno con la polaridad (positivo, negativo, neutral) que le corresponde, así como la fuerza subjetiva (*weaksubj* para los términos débiles y *strongsubj* para los términos fuertes), el proceso de etiquetado de este corpus así como el de comentarios también es manual y realizado por profesores de la UPA. La conformación del corpus de subjetividad queda de la manera siguiente, de 2,704 términos, 1,690 son términos etiquetados como negativos, de los cuales 1,239 son etiquetados como *strongsubj* y 451 son etiquetados como *weaksubj*. En cuanto a los términos distintivos positivos se obtienen 1,014, de los cuales 598 son etiquetados como *strongsubj* y 416 como *weaksubj*. Un extracto del corpus de subjetividad se muestra en la tabla 7.

Tabla 7. Extracto de corpus de subjetividad

término	fuerza subjetiva	clase
apenado	weaksubj	negative
aporta	strongsubj	positive
aportar	weaksubj	positive
apreciamos	strongsubj	positive
aprecia	strongsubj	positive
apreciar	strongsubj	positive
aprende	weaksubj	positive
aprendemos	strongsubj	positive
aprobamos	weaksubj	positive
aprueba	weaksubj	positive
apruebas	weaksubj	positive

Para el proceso de clasificación se aplica el algoritmo de *Naïve Bayes* apoyado del corpus de subjetividad, para determinar el *score* de los términos que influyen en la clasificación del comentario dependiendo de su fuerza subjetiva. Por lo tanto el algoritmo de *Naïve Bayes* además de considerar la cantidad de términos negativos y positivos en el corpus de subjetividad también considera el *score* de cada término calculado. Una vez realizado el proceso de clasificación se calculan cada una de las métricas descritas en el apartado Métricas para evaluar el desempeño de una clasificación del capítulo II. A lo largo de la investigación el modelo ha evolucionado, se considera que ha pasado por tres versiones, las cuales han variado en el proceso de clasificación dependiendo de las funciones y

la actualización del corpus de subjetividad, esto de igual manera ha provocado variación en los resultados de sus métricas.

3.1.3 Resultados

Los resultados así como las características principales de cada una de las versiones del Modelo SocialMining se describen enseguida.

Versión 1

En esta versión del Modelo SocialMining, el corpus de subjetividad se integra por 1300 términos, sin embargo debido a los bajos resultados en el proceso de clasificación (ver figura 14), se realiza una revisión del corpus de subjetividad, en el cual se encuentran términos repetidos con diferente fuerza subjetiva. Por esta razón se opta por formar tres equipos de pares de profesores para revisión de comentarios por etapas, es decir, primero revisa un equipo de profesores que indica cambios a realizar, de igual forma el segundo equipo vuelve a revisar y por último el tercer equipo hace la última revisión en la que además considera las sugerencias y adaptaciones de los equipos previos.

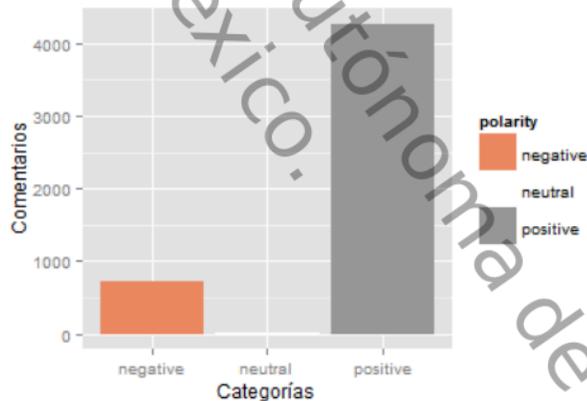


Figura 14. Gráfico de resultados de clasificación de la versión 1

Versión 2

Uno de los cambios en la versión 2, es la omisión del proceso de *stemming*, ya que en el idioma español a diferencia del inglés existen numerosos cambios morfológicos [96], Esto provoca que el cambio de la raíz en determinadas formas de una palabra o términos afecten el proceso de clasificación al realizar un

proceso de *matching* entre los términos del corpus de subjetividad y los términos del corpus de comentarios. Por ejemplo al realizar *stemming* en “oportunidad” y “oportunista”, la raíz de ambos términos es igual “oportun”, sin embargo la polaridad de las mismas es diferente, oportunidad da entender un sentido positivo mientras que oportunista se considera negativo. Los resultados de clasificación de la versión 2 se presentan en la figura 15.

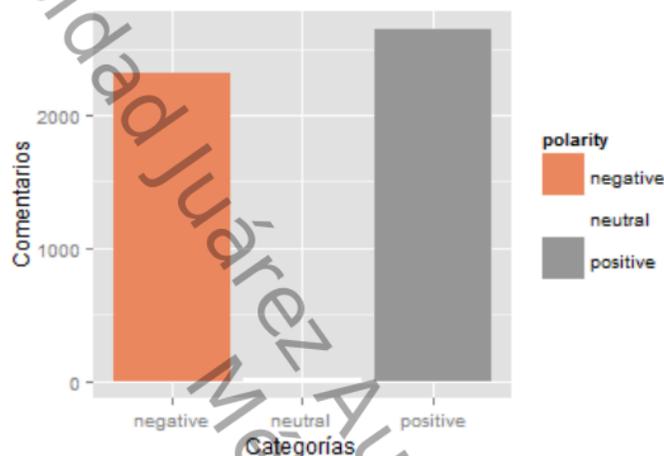


Figura 15. Gráfico de resultados de clasificación de la versión 2

Versión 3

El cambio principal en la versión 3 además de actualizar el corpus de subjetividad con 2704 términos, es el uso de la función *match*, en lugar de la función *pmatch* de R [95]. La función *match* a diferencia de *pmatch* busca los términos idénticos y no aquellos términos similares en sus prefijos. Con este cambio se logró aumentar la identificación de términos negativos en el corpus de comentarios (ver figura 16).

¹ La versión 3 del Modelo SocialMining es la más actual, en la cual se ha mejorado el proceso de clasificación.

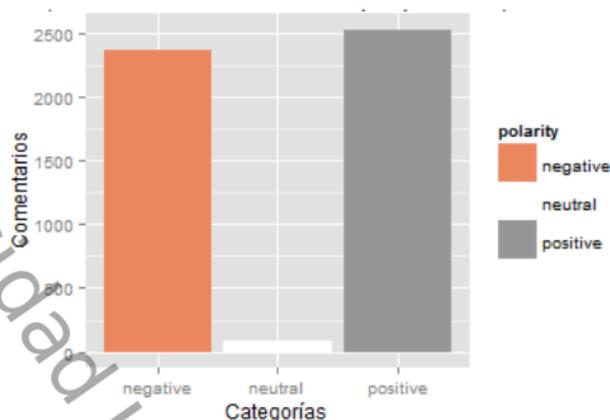


Figura 16. Gráfico de resultados de clasificación de la versión 3

La tabla 8 presenta las matrices de confusión del proceso de clasificación de los comentarios, correspondiente a cada una de las versiones. Donde la versión 1 del modelo obtiene 2,270 comentarios positivos verdaderos (VP) y 691 comentarios verdaderos negativos (VN), en cambio la versión 2 obtiene 2,197 comentarios positivos verdaderos (VP) y 2,226 comentarios verdaderos negativos (VN), finalmente la versión 3 obtiene 2,173 comentarios positivos verdaderos (VP) y 2,260 comentarios verdaderos negativos.

Tabla 8. Matices de confusión de las versiones de SocialMining

	Versión 1		Versión 2		Versión 3	
	Valor Actual					
Valor predictivo	2270	2000	2197	458	2173	363
	33	691	95	2226	118	2260

Es necesario indicar que en las matrices de confusión de cada una de las versiones no se completan los 5,002 comentarios del corpus, esto es porque el Modelo SocialMining también etiquetó comentarios como neutrales. En la versión 1, son 26 comentarios etiquetados como neutrales, en la versión 2, son 8 neutrales y por último en la versión 3 hay 88 neutrales; cabe resaltar que los comentarios neutrales son falsos debido a que en el corpus de comentarios original, no hubo comentarios etiquetados manualmente como neutrales.

Una de las razones en la variación de la clasificación de los comentarios, se debe a que en el modelo propuesto no se considera la semántica, por lo que no detecta aquellos comentarios que cambian su polaridad de positiva a negativa o de negativa a positiva debido a un término de contraposición (por ejemplo: pero, sin embargo, no obstante, aunque, entre otros), esto en consecuencia provoca un error en la clasificación.

1 La tabla 9 presenta los resultados de las métricas: sensibilidad (VRP), precisión total (ACC), precisión equilibrada (WACC) y especificidad (SPC) de cada una de las versiones del Modelo SocialMining. En dicha tabla de igual manera que en la figura 15 se puede observar la mejora en resultados de clasificación de la versión 3, y que tanto en precisión total y precisión balanceada alcanza poco más de 90%, resultado que se considera aceptable.

Tabla 9. Resultados de VRP, ACC, ACCW y SPC

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

1 Considerando los resultados de VRP y SPC de la versión 3 del Modelo SocialMining, se procede a obtener una relación de valores predictivos (ver tabla 10).

Tabla 10. Extracto de la relación de valores predictivos

N°	P(+)	P(-)	VP+	FVP+	VP-	FVP-
.01	0.148	0.852	0.064	0.936	0.999	0.001
.02	0.156	0.844	0.121	0.879	0.999	0.001
.03	0.164	0.836	0.172	0.828	0.998	0.002
.04	0.172	0.828	0.219	0.781	0.997	0.003
.05	0.180	0.820	0.261	0.739	0.996	0.004
.06	0.188	0.812	0.300	0.700	0.996	0.004
.07	0.196	0.804	0.336	0.664	0.995	0.005
.08	0.204	0.796	0.369	0.631	0.994	0.006
.09	0.212	0.788	0.399	0.601	0.993	0.007

N°	P(+)	P(-)	VP+	FVP+	VP-	FVP-
.10	0.220	0.780	0.427	0.573	0.992	0.008
.11	0.228	0.772	0.454	0.546	0.991	0.009
.12	0.236	0.764	0.478	0.522	0.991	0.009
.13	0.244	0.756	0.501	0.499	0.990	0.010
.14	0.252	0.748	0.522	0.478	0.989	0.011

Para el cálculo de la probabilidad de que el comentario sea positivo (P+) se utiliza el teorema de la probabilidad total. Para calcular la probabilidad de que el comentario sea negativo (P-) se aplica el método de la probabilidad del suceso contrario en un suceso condicionado.

Para el cálculo del valor predictivo positivo (VP+) se aplica el teorema de Bayes. De igual manera para el cálculo del falso valor predictivo positivo se aplica el método de la probabilidad del suceso contrario en un suceso condicionado.

El valor predictivo negativo (VP-) es calculado mediante el teorema de Bayes y de nuevo el falso valor predictivo es calculado utilizando el método de la probabilidad del suceso contrario en un suceso condicionado.

Con base en los resultados de los valores predictivos mostrados en la tabla 8 se presenta la figura 17, en la cual se observa que los valores predictivos positivos y los valores predictivos negativos tienden a mejorar.

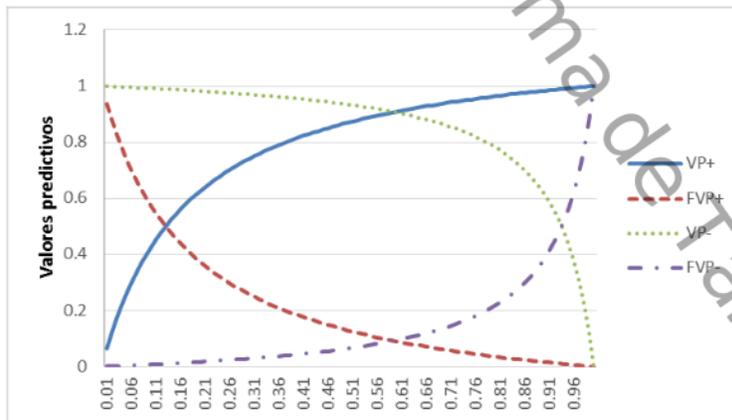


Figura 17. Gráfico de valores predictivos

La curva ROC obtenida con base a los valores predictivos mostrados en la tabla anterior, se presenta en la figura 18, en la cual puede observarse el incremento del desempeño en la clasificación de comentarios del Modelo SocialMining presentado mediante la curva.

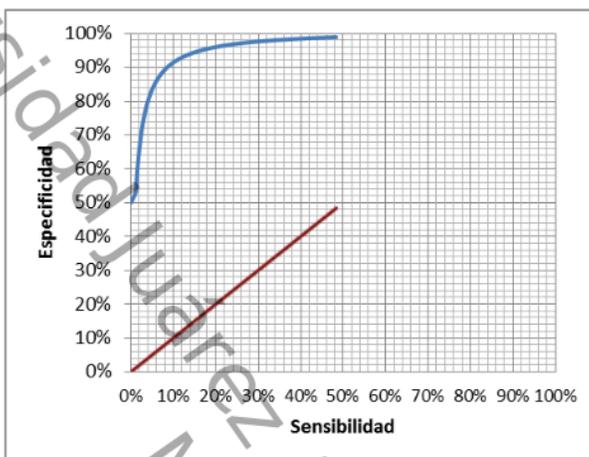


Figura 18. Curva ROC en SocialMining con Naïve Bayes y el corpus de subjetividad

Algunos autores como Mejova [97], Prasad [98] y otros [18, 88] indican que la técnica de *Naïve Bayes* en el área de análisis de emociones logra buenos resultados en procesos de clasificación, en este caso el resultado de la precisión total y de la precisión balanceada del modelo SocialMining fue de un 90%, este resultado se considera muy bueno y se atribuye al uso de *Naïve Bayes* en combinación con el corpus de subjetividad.

3.1.4 Caso de estudio

Una vez obtenidas las métricas del Modelo SocialMining, se realizó un caso de estudio con el fin de medir el desempeño del mismo modelo con un corpus de comentarios nuevo, el cual se conformó por estudiantes del 5° cuatrimestre de la carrera de Ingeniería en Sistemas que evaluaron a 11 profesores..

Una vez realizado el proceso de clasificación por el Modelo SocialMining se procedió a comparar contra el nuevo corpus de comentarios generado. La figura 19 muestra los resultados de esta comparación, en la parte horizontal de la gráfica se ubican los 11 profesores, la parte vertical presenta el número de comentarios positivos o negativos obtenidos. Los comentarios etiquetados manualmente se identifican por la palabra "manual", mientras que los comentarios etiquetados por

el Modelo SocialMining se identifican por la palabra “SM”. El nuevo corpus cómo se explicó en la sección de Materiales y Métodos está conformado por un total de 623 comentarios, de los cuales 304 son comentarios negativos, 305 son comentarios positivos y 14 son comentarios neutrales.

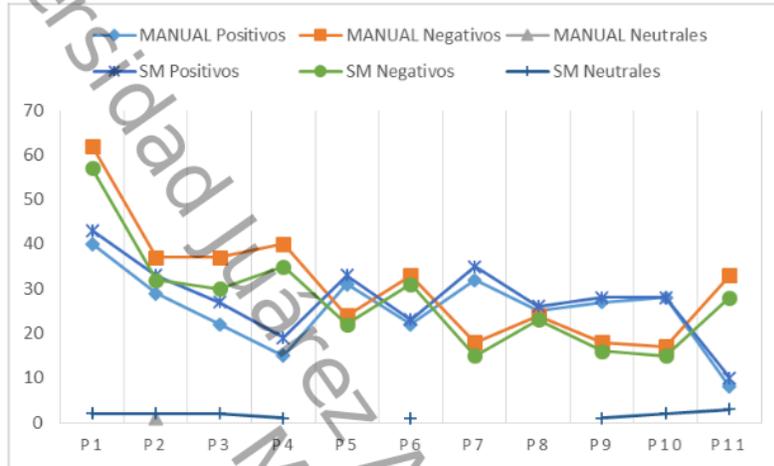


Figura 19. Gráfico de valores predictivos

Los resultados de la clasificación por el Modelo SocialMining se presentan en la tabla 11 mediante una matriz de confusión, así como los resultados de las métricas en la tabla 12.

Tabla 11. Matriz de confusión

		Valor actual	
Valor actual	73	3	
	6	02	

Tabla 12. Resultados de la precisión y la sensibilidad

ACC	ACCW	VPR	SPC
0.936482085	0.936338172	0.978494624	0.901492537

¹ De igual manera en este caso de estudio se clasificaron 9 comentarios como neutrales falsos, sin embargo como puede observarse los valores de ACC y ACCW subieron de 90% (ver tabla 2) a un 93% (ver tabla 5), por lo cual el Modelo SocialMining denota nuevamente que el desempeño en el proceso de clasificación

tiende a mejorar con el uso del corpus de subjetividad y el algoritmo de *Naïve Bayes*.

3.1.5 Discusión de resultados

El análisis de los comentarios recolectados en Twitter y la EED favorecen la adecuación del corpus de subjetividad mediante la identificación de términos distintivos. El corpus de subjetividad en combinación con el algoritmo de *Naïve Bayes* mejora el proceso de clasificación. Sin embargo, para la mejora continua en los resultados de clasificación se considera importante mantener actualizado el corpus de subjetividad, ya que por ejemplo los coloquialismos varían con el tiempo y la región.

En las ciencias de la computación resulta atractivo utilizar este tipo de Modelos de aprendizaje para automatizar procesos, ahorrar tiempo y contribuir adecuadamente a la toma de decisiones. El desarrollo del Modelo SocialMining apoya en generalizar los comportamientos a partir de los datos no estructurados aportados por estudiantes. El análisis de emociones se basa en el análisis de textos y el modelo presentado ofrece una solución factible a la problemática del análisis de comentarios de la evaluación docente.

La aplicación de *Naïve Bayes* en la evaluación docente resulta factible para diferentes aspectos a mejorar en el proceso de enseñanza, ya que no existe un corpus en español dedicado a la educación y específicamente enfocado a la evaluación docente. La ventaja de la adecuación del corpus de subjetividad en combinación con *Naïve Bayes* puede derivar incluso en una implementación de un sistema recomendador, para en el caso que sea detectada una debilidad en el desempeño del profesor, ofrecer a éste un curso de formación de competencias docentes para contribuir a la mejora de sus clases.

La aplicación del Modelo SocialMining resulta factible ya que incluso se detectan observaciones del comportamiento de los estudiantes en sus comentarios. Por ejemplo una de las observaciones percibidas mediante el análisis de comentarios, es que los estudiantes por lo regular usan la evaluación docente como un medio para quejarse de las clases si existe un caso o situación que no les agrada, muy pocos aplauden el esfuerzo del profesor y sugieren estrategias o cambios para mejorar la clase, incluso este comportamiento puede observarse por grupos, ya que grupos donde el índice de aprovechamiento es alto

sugieren al profesor cambios constructivos con el fin de aprender mejor o bien elogian la manera de impartir clases, contrario a lo que sucede en algunos otros grupos donde no muestran interés por evaluar a sus profesores y se observa su bajo rendimiento académico.

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

Los resultados obtenidos en esta investigación permiten evaluar la precisión del Modelo SocialMining mediante el corpus de subjetividad y la técnica de Naïve Bayes, suponiendo factible la implementación de éste en instituciones de educación superior, sin embargo es importante considerar la región donde pueda ser implementado debido a la variedad de coloquialismos con la que se pueden expresar los estudiantes de las instituciones, por esta razón se cree que es relevante mantener actualizado el corpus de subjetividad con términos de la región, de igual manera se puede aplicar el Modelo a otro contexto siempre y cuando se adecúe el corpus de subjetividad al área correspondiente.

Con el fin de comparar el desempeño del Modelo SocialMining aplicando Naïve Bayes y Subjetividad, se aplican dos algoritmos más, la SVM y *Random Forest*. En la siguiente sección se describe el desarrollo del Modelo SocialMining aplicando los dos algoritmos mencionados, de igual manera se presentan los datos utilizados y el diseño experimental realizado.

3.2 Modelo SocialMining aplicando SVM y *Random Forest*

3.2.1 Descripción del Modelo

La arquitectura del Modelo SocialMining considerando la implementación de los algoritmos *SVM* y *Random Forest* se presenta en la figura 20.

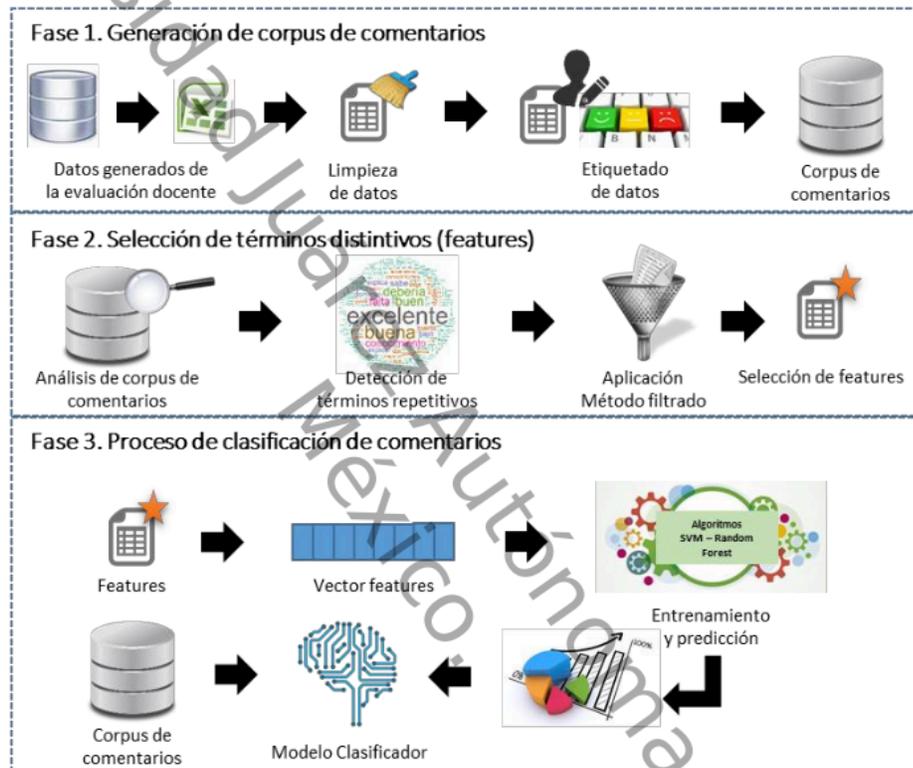


Figura 20. Arquitectura de SocialMining aplicando aprendizaje automático

Fase 1. Generación de corpus de comentarios

En esta fase se genera el corpus de comentarios, los cuales pueden ser extraídos de diferentes medios (por ejemplo: redes sociales, encuestas, herramientas de software para evaluación docente), en el caso de este estudio, los comentarios corresponden a una encuesta de evaluación de profesores aplicada a tres grupos de estudiantes de Ingeniería en Sistemas de la UPA. Cabe resaltar que solo se consideran aquellos comentarios libres de spam.

Una vez obtenidos los comentarios son sometidos a un proceso etiquetado manual realizado por profesores de la institución, quienes consideran un rango de valores de -2 a +2. Los comentarios negativos son etiquetados dentro de un rango de -2 a -0.2, usando el valor -2 para aquellos comentarios muy negativos. Los comentarios etiquetados dentro del rango de +0.2 a +2 son considerados positivos, el valor 2 es usado para etiquetar comentarios muy positivos. Por otra parte los comentarios etiquetados con valor de 1 son considerados como neutrales.

Asimismo una vez finalizado el proceso de etiquetado en los comentarios, pasan a un proceso de limpieza, en el cual se eliminan las *stop words* y sustantivos que aparecen en la mayoría de los comentarios como: profesor, universidad, clase, materia, escuela, entre otros. Además, cada uno de los términos del comentario son convertidos a minúsculas y se eliminan los signos de puntuación y números.

Fase 2. Selección de términos distintivos (*features*)

En esta fase se realiza una selección de los términos distintivos o *features*. Para la selección de *features* se utiliza el método de Modelo de Espacio Vectorial (*Vector Space Model* en inglés), en la cual el texto de los comentarios es tratado como una Bolsa de Palabras (técnica conocida en inglés como *Bag of Words* [99]), donde se ignora la estructura gramática del texto [100]. Este proceso es realizado mediante la función *findFreqTerms()* de R [95]. Posteriormente se aplica la función *removeSparseTerms()* también de R para depurar la dispersión de los términos del corpus de comentarios. El resultado de esta fase es el corpus de comentarios contenido en una matriz, en donde los comentarios son almacenados en filas y las *features* en columnas.

Fase 3. Proceso de Clasificación

En esta fase el corpus de comentarios y *features* es particionado en dos conjuntos de datos independientes (*datasets*). Uno de los *dataset* es dedicado al proceso de entrenamiento (*train*) y se usa en algoritmos de clasificación para encontrar patrones o relaciones entre los datos; el segundo *dataset* es considerado para el proceso de prueba (*test*) con el fin de ajustar el desempeño del modelo (una vez que el algoritmo es entrenado, recibe el nombre de modelo). En el Modelo

SocialMining se utilizan dos tercios (2/3) de los comentarios para el *dataset* de entrenamiento y un tercio (1/3) para el *dataset* de prueba.

Los algoritmos de aprendizaje automático implementados en el Modelo SocialMining son: SVM (considerando los tres núcleos lineal, radial y polinomial) y Random Forest, a los cuales se aplica un proceso de configuración de parámetros y un proceso de entrenamiento. El proceso de configuración de parámetros es realizado antes del entrenamiento para ajustar los parámetros de cada uno de los algoritmos.

Para el proceso de configuración de parámetros y entrenamiento se aplica el método de validación cruzada de k iteraciones. En el método de validación cruzada los datos del corpus de comentarios se dividen en k subconjuntos. Uno de los subconjuntos se utiliza como datos de prueba y el resto ($k-1$) como datos de entrenamiento. El proceso de validación cruzada es repetido durante k iteraciones con cada uno de los posibles subconjuntos de datos de prueba, dando como resultado una matriz de confusión con los valores promedio. Una vez que las iteraciones k han terminado se obtiene el resultado de la validación cruzada. En el Modelo SocialMining se configura k con valor de diez. Mediante el proceso de entrenamiento es posible identificar si el valor de los parámetros varía o permanecen constantes.

El resultado de esta fase es el modelo clasificador de aprendizaje automático previamente configurado y entrenado, posteriormente se obtiene la matriz de confusión, los valores de sensibilidad, especificidad y precisión, así como un análisis ROC, para determinar el desempeño del modelo clasificador.

3.2.2 Datos y diseño experimental

3.2.2.1 Datos

Los datos usados en este estudio son conformados mediante la evaluación docente (EED), realizada por tres grupos de la carrera de Ingeniería en Sistemas, en donde evaluaron a 21 profesores en el primer periodo escolar del 2016. Mediante esta evaluación docente se reúne un total de 1,040 comentarios, los cuales conforman el corpus de comentarios.

Posteriormente se realiza un proceso de etiquetado, donde se identifican 515 comentarios negativos y 515 comentarios positivos, así como 10 comentarios

neutrales, razón por la cual se considera un corpus balanceado, ya que la prioridad es realizar una clasificación binaria, es decir, clasificar comentarios en positivo o negativo. Asimismo en la conformación del corpus de comentarios se cuida que el promedio de los valores de etiquetado en las dos polaridades (positivo, negativo) sea equitativo.

En este estudio se identifica un conjunto de 99 *features*. Un extracto de las *features* consideradas se muestra en la tabla 13.

Tabla 13. Extracto de features utilizadas en SocialMining

Feature	Polaridad	Feature	Polaridad
Aburrida	Positive	Debería	Negative
Agrada	Positive	Debe	Negative
Apoya	Positive	Impuntual	Negative
Aprender	Positive	Elitista	Negative
Bien	Positive	Impaciente	Negative
Bipolar	Positive	Problemático	Negative
Buena	Positive	Lento	Negative

3.2.2.2 Diseño experimental

En esta sección se describe el proceso de experimentación, para lo cual se utiliza el corpus descrito en la sección anterior, utilizando dos tercios para conformar el *dataset* de entrenamiento y un tercio para el *dataset* de prueba. Se realizan 30 ejecuciones aplicando el método de validación cruzada de 10 iteraciones.

En cada una de las ejecuciones de la validación cruzada de 10 iteraciones se controla y configura una semilla diferente, cuidando que cada una de estas semillas sean iguales para cada algoritmo. El uso de una semilla diferente en cada ejecución asegura que la partición del corpus en *datasets* de entrenamiento y prueba será diferente.

3.2.3 Configuración de parámetros

Algunos de los núcleos de la SVM, así como el algoritmo *Random Forest* requieren la configuración de los parámetros, para esto se considera el método propuesto por Hsu [101] denominado *grid search*. En este método se define un conjunto de valores de los parámetros que desea aplicar en un determinado modelo o algoritmo. Este método es importante debido a que algunos algoritmos

como SVM con núcleo polinomial o sigmoideal requieren la configuración de sus parámetros, de lo contrario el entrenamiento puede ser largo y provocar errores.

En SVM los valores usados en el proceso de tuneo para el parámetro C van de 0.1 a 2, el valor de sigma (σ) va de 0.01 a 2, para el parámetro *degree* se usan diferentes valores que van de 2 a 10, y valores de 0 a 1 para el parámetro de coef utilizado en el núcleo polinomial.

En Random Forest es necesario la configuración de dos parámetros: *mtry* y *ntree*, los cuáles se describen en la sección 2.1.6.3 Bosques aleatorios (*Random Forest*). En primera instancia no se configura el parámetro *mtry*, dejándolo al azar, sin embargo para el parámetro *ntree* se consideran valores entre un rango de 100 a 1000, utilizando el método de Hsu [101] y la validación cruzada de 10 iteraciones.

Una vez obtenidos los valores de los parámetros se procede a realizar el proceso de predicción, mediante 30 ejecuciones usando 30 semillas diferentes para calcular las métricas de sensibilidad, especificidad, precisión, precisión equilibrada y el análisis ROC.

3.2.4 Resultados de optimización de parámetros

En esta sección se presentan los resultados del proceso de optimización de los parámetros de cada uno de los algoritmos implementados, así como un análisis de las *features* y por último los resultados obtenidos en la predicción considerando las métricas de desempeño.

Para el proceso de optimización de parámetros se establece una semilla con valor igual a uno y acorde con la metodología propuesta por Hothorn [9] se realiza una comparación entre los algoritmos utilizando la técnica de remuestreo. La tabla muestra el resumen de los resultados del remuestreo aplicado en los núcleos lineal, radial y polinomial de SVM, así como en *Random Forest*.

El remuestreo se aplica mediante la función *resamples* en R, las métricas de desempeño utilizadas son ROC, sensibilidad y especificidad (ver figura 21).

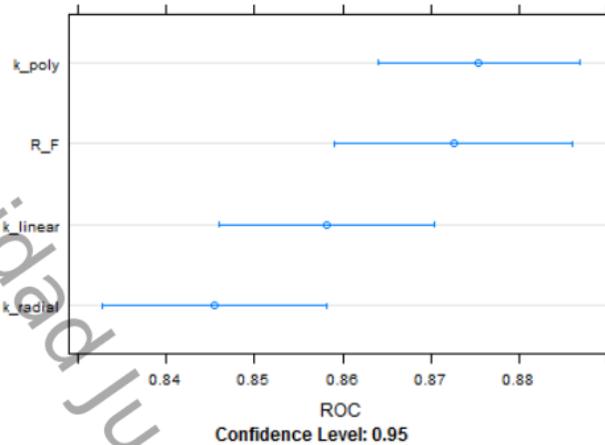


Figura 21. Resultados de remuestreo utilizando ROC

En la figura 20 se presenta el desempeño del promedio de resultados del remuestreo utilizando la métrica ROC, en donde se puede observar que el núcleo polinomial (*k_poly*) de SVM, así como el algoritmo *Random Forest* (*R_F*) muestran un mejor desempeño a comparación del núcleo lineal (*k_linear*) y kernel radial (*k_radial*).

El resumen de los resultados en el proceso de optimización de parámetros se muestra en la tabla 14, donde es posible observar que efectivamente el núcleo polinomial y *Random Forest* presentan un buen desempeño en este proceso.

Tabla 14. Resumen de resultados de optimización de parámetros

ROC	3	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	<i>k_linear</i>		0.7786	0.8366	0.8626	0.8581	0.879
<i>k_radial</i>		0.7856	0.8207	0.8473	0.8455	0.8611	0.9161
<i>k_poly</i>		0.8169	0.8563	0.8797	0.8755	0.8965	0.924
<i>R_F</i>		0.7881	0.8627	0.8805	0.8726	0.8885	0.9422
Sensibility	3	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	<i>k_linear</i>		0.5857	0.6571	0.7	0.7006	0.7571
<i>k_radial</i>		0.6571	0.7286	0.7571	0.7646	0.8	0.9
<i>k_poly</i>		0.6143	0.7143	0.7429	0.7451	0.7857	0.8571
<i>R_F</i>		0.1857	0.4571	0.4714	0.4971	0.6143	0.8

Specificity	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
<i>k_linear</i>	0.7826	0.8696	0.8971	0.8895	0.913	0.9565
<i>k_radial</i>	0.6812	0.7681	0.7826	0.7919	0.8261	0.8696
<i>k_poly</i>	0.8116	0.8551	0.8824	0.8802	0.8986	0.942
<i>R_F</i>	0.8986	0.9275	0.971	0.9623	0.9855	1

Una vez realizado el proceso de optimización de parámetros para SVM y *Random Forest*, se procede a ejecutar el proceso de predicción, así como la evaluación por métricas. La tabla 15 presenta los resultados promedio y la desviación estándar de las 30 ejecuciones de cada algoritmo.

Tabla 15. Resultados promedio de clasificación

	Total Accuracy	Weighted Accuracy	Sensibility	Specificity
<i>k_linear</i>	0.8038	0.8149	0.8936	0.7160
	0.0153	0.0146	0.0277	0.0364
<i>k_radial</i>	0.7850	0.7893	0.8242	0.7467
	0.0190	0.0159	0.0422	0.0561
<i>k_poly</i>	0.6779	0.7014	0.8661	0.4941
	0.0363	0.1336	0.1649	0.1009
<i>R_F</i>	0.7126	0.7905	0.9661	0.4650
	0.0561	0.0228	0.0216	0.1270

En la tabla 15 es posible observar que el núcleo lineal obtiene un valor de precisión equilibrada de poco más de 0.80 (80%). Los valores obtenidos en la métrica de sensibilidad son más altos que los obtenidos en la métrica de especificidad. Por otra parte, el núcleo polinomial obtiene el valor más bajo en todas las métricas excepto en sensibilidad, métrica en la cual sobresale *Random Forest*. En resumen los tres núcleos de SVM y *Random Forest* resultan más sensibles que específicos. Sin embargo estos resultados no indican que los modelos clasificadores presenten un buen desempeño en la clasificación ya que un valor alto en sensibilidad significa que tiene capacidad para identificar los comentarios verdaderos positivos, y un valor bajo en especificidad indica un bajo desempeño en la identificación de comentarios verdaderos negativos. Razón por la cual se realiza una optimización de *features*.

3.2.5 Optimización de *features*

Debido a los bajos resultados de precisión total y precisión balanceada mostrados en la tabla 13, se considera realizar una optimización de *features* utilizando el método de filtrado (explicado en la sección 2.1.7 Métodos de selección de términos distintivos), para esto se usa específicamente el algoritmo *Random Forest Importance Measure* (VIM) como método para valoración de *features*.

Una de las ventajas de trabajar en aprendizaje automático es que existe la posibilidad de mejorar el desempeño de los clasificadores [102]. El método de VIM utiliza el método *Out Of Bag* (OOB) para obtener diferentes muestras de *features* y medir la precisión de la predicción [103, 104]. La metodología propuesta para la optimización de *features* se presenta en la figura 22.

1. Obtener nube de palabras de corpus de comentarios.
2. Identificar los términos más frecuentes del corpus de comentarios.
3. Particionar corpus de comentarios en test de prueba y entrenamiento.
 - a. Definir conjuntos de valores de los términos del modelo para evaluar
 - b. **Para** cada conjunto de términos **hacer**
 - i. **Para** cada iteración de remuestreo **hacer**
 1. Mantener muestras específicas
 2. Pre procesar los datos
 3. Ajustar el modelo
 4. Predecir las muestras (*features*) a mantener
 - ii. **Fin (para)**
 - iii. Calcular el desempeño promedio utilizando la predicción de muestras
 - c. **Fin (para)**
4. Determinar el conjunto óptimo de *features*.
5. Comparación entre términos más frecuentes y términos ponderados.

Figura 22. Metodología propuesta para la optimización de *features*

Una vez que el corpus de comentarios ha pasado por un proceso de limpieza, donde se eliminan las *stop words* y sustantivos o términos repetitivos sin importancia, se corrige ortografía y se revisa que los términos mantenidos puedan contribuir al proceso de clasificación. Para esto se genera una nube de palabras para visualizar los términos repetitivos en el corpus.

no.	feature	score
10	debería	29.387665
11	falta	26.2034223
14	mejores	23.7956642

A pesar de que el método VIM también considera algunos términos irrelevantes, estos son relativamente pocos, a diferencia de los términos que apoyan al proceso de clasificación. Una vez que el proceso de optimización de *features* realizado mediante VIM es terminado, se realiza nuevamente el proceso de optimización de parámetros empleando el método de validación cruzada de 10 iteraciones.

Los resultados obtenidos en el proceso de optimización de parámetros en SVM con núcleo lineal se presentan en la tabla 17, donde puede observarse que los valores de C sugeridos para entrenar el algoritmo son: 0.55, 0.85, 1.00, 1.30 y 1.60 con base en los resultados obtenidos en la métrica de *Accuracy* (precisión).

Tabla 17. Resultados de ajuste de SVM con núcleo Lineal

C	Accuracy	Kappa	Accuracy SD	Kappa SD
0.10	0.8516	0.7036	0.0458	0.0913
0.55	0.8530	0.7064	0.0457	0.0909
0.85	0.8530	0.7064	0.0457	0.0909
1.00	0.8530	0.7064	0.0457	0.0909
1.30	0.8530	0.7064	0.0457	0.0909
1.60	0.8530	0.7064	0.0457	0.0909
2.00	0.8501	0.7007	0.0472	0.0941

Los resultados obtenidos en el proceso de optimización de parámetros en SVM con núcleo radial se presentan en la tabla 18, donde puede observarse que el valor sugerido para entrenar el algoritmo en el parámetro C es de 1, con base en los resultados obtenidos en la métrica de *Accuracy* (precisión) y *Kappa*.

Tabla 18. Resultados de ajuste de SVM con núcleo Radial

C	Accuracy	Kappa	Accuracy SD	Kappa SD
0.25	0.7926	0.5846	0.0580	0.1163
0.5	0.8675	0.7350	0.0512	0.1023

C	Accuracy	Kappa	Accuracy SD	Kappa SD
1	0.8804	0.7609	0.0481	0.0961
2	0.8746	0.7493	0.0427	0.0852
4	0.8803	0.7608	0.0468	0.0934
8	0.8803	0.7608	0.0468	0.0934
16	0.8803	0.7608	0.0468	0.0934
32	0.8803	0.7608	0.0468	0.0934
64	0.8803	0.7608	0.0468	0.0934

De igual manera en SVM con núcleo radial, el parámetro sugerido para sigma es 0.01928, como se puede observar en la figura 24.

```

Support Vector Machines with Radial Basis Function Kernel

694 samples
1142 predictors
2 classes: 'X2', 'X1'

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 625, 625, 624, 624, 624, 625, ...
Resampling results across tuning parameters:

C          Accuracy  Kappa   Accuracy SD  Kappa SD
0.25      0.7925673  0.5846458  0.05804889   0.11628853
0.50      0.8674741  0.7349926  0.05117539   0.10233037
1.00      0.8803727  0.7609027  0.04814651   0.09609801
2.00      0.8745756  0.7492890  0.04267475   0.08515536
4.00      0.8803313  0.7607846  0.04679272   0.09337932
8.00      0.8803313  0.7607846  0.04679272   0.09337932
16.00     0.8803313  0.7607846  0.04679272   0.09337932
32.00     0.8803313  0.7607846  0.04679272   0.09337932
64.00     0.8803313  0.7607846  0.04679272   0.09337932
128.00    0.8803313  0.7607846  0.04679272   0.09337932

Tuning parameter 'sigma' was held constant at a value of 0.01928653
Accuracy was used to select the optimal model using the largest value.
The final values used for the model were sigma = 0.01928653 and C = 1.
elapsed time is 10.690000 seconds
    
```

Figura 24. Resultados de la ejecución en la optimización de parámetros en SVM - radial

Los resultados obtenidos en el proceso de optimización de parámetros en SVM con núcleo polinomial se presentan en la tabla 19, donde puede observarse que el valor sugerido para entrenar el algoritmo en el parámetro C es de 0.01 y

para el parámetro *degree* se sugiere el valor de 1, con base en los resultados obtenidos en la métrica de *Accuracy* (precisión) y *Kappa*.

Tabla 19. Resultados de ajuste de SVM con núcleo Polinomial

degree	C	Accuracy	Kappa	Accuracy SD	Kappa SD
1	0.01	0.8919	0.7841	0.0444	0.0884
1	0.08	0.8530	0.7066	0.0401	0.0799
1	0.1	0.8559	0.7122	0.0360	0.0718
1	0.5	0.8487	0.6978	0.0468	0.0932
1	1	0.8501	0.7007	0.0472	0.0941
1	1.3	0.8530	0.7064	0.0457	0.0909
2	0.01	0.8501	0.7006	0.0411	0.0819
2	0.08	0.8529	0.7063	0.0427	0.0850
2	0.1	0.8529	0.7063	0.0427	0.0850
2	0.5	0.8529	0.7063	0.0427	0.0850
2	1	0.8529	0.7063	0.0427	0.0850
2	1.3	0.8529	0.7063	0.0427	0.0850
3	0.01	0.8486	0.6977	0.0461	0.0918
3	0.08	0.8486	0.6977	0.0461	0.0918

Los resultados obtenidos en el proceso de optimización de parámetros en *Random Forest* se presentan en la tabla, donde puede observarse que el valor sugerido para entrenar el algoritmo en el parámetro *mtry* es de 77, con base en los resultados obtenidos en la métrica de *Accuracy* (precisión) y *Kappa*.

Tabla 20. Resultados de ajuste de Random Forest

mtry	Accuracy	Kappa	Accuracy SD	Kappa SD
77	0.8515	0.7034	0.0481	0.0957
88	0.8515	0.7035	0.0461	0.0917

El resultado promedio del proceso de optimización de los parámetros de SVM y *Random Forest* se presenta en la tabla 21, en donde puede observarse que los núcleos radial (*k_radial*) y polinomial (*k_poly*) presentan un mejor desempeño en la métrica de precisión y *Kappa*.

Tabla 21. Promedio de los resultados de ajuste de las SVM y Random Forest

Accuracy						
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
<i>k_linear</i>	0.7826	0.8214	0.8696	0.853	0.8918	0.9
<i>k_radial</i>	0.7971	0.8514	0.8921	0.8804	0.9104	0.942
<i>k_poly</i>	0.8116	0.8643	0.8986	0.8919	0.9247	0.9565
<i>R_F</i>	0.7536	0.8194	0.8634	0.8515	0.8809	0.913
Kappa						
	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
<i>k_linear</i>	0.566	0.6429	0.7397	0.7064	0.7836	0.8
<i>k_radial</i>	0.5951	0.7032	0.7842	0.7609	0.8208	0.8841
<i>k_poly</i>	0.6245	0.7286	0.7974	0.7841	0.8494	0.9131
<i>R_F</i>	0.5086	0.6398	0.727	0.7034	0.7621	0.8261

La figura 25 presenta la gráfica de los resultados promedio del proceso de optimización de los parámetros mostrado en la tabla anterior.

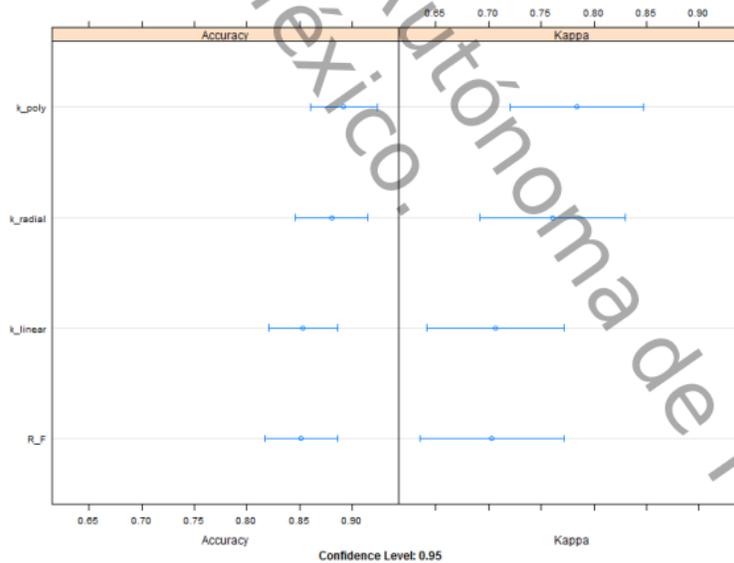


Figura 25. Promedio de los resultados de ajuste de las SVM y Random Forest

Como se puede observar en las métricas de precisión y *Kappa* [105], el núcleo polinomial (*k_poly*) presenta un mejor desempeño en la optimización de parámetros seguido por el núcleo radial, y lineal, caso contrario para el algoritmo

Random Forest, el cual presenta el desempeño más bajo en optimización de parámetros en ambas métricas. Dado que SVM y *Random Forest* poseen los mismos datos de entrenamiento tiene sentido hacer inferencias sobre las diferencias entre éstos. De esta manera es posible reducir la correlación que pueda existir. El cálculo de las diferencias es se presenta en la tabla y figura 22.

Tabla 22. Resultados de diferencia en SVMs y Random Forest

Accuracy				
	linear	k_radial	k_poly	R_F
k_linear		0.0274	0.0389	0015
k_radial	0.1211		0.0115	0.0289
k_poly	0.0043	1.0000		0.0404
R_F	1.0000	0.5935	0.0301	

Kappa				
	k_linear	k_radial	k_poly	R_F
k_linear		0.0545	0.0776	0.0030
k_radial	0.1234		0.0232	0.0575
k_poly	0.0043	1		0.0806
R_F	1	0.5974	0.0299	

De acuerdo a los datos de la tabla 22 es posible observar que entre los núcleos no existe gran diferencia en el desempeño, a diferencia del algoritmo de *Random Forest*, visualmente las diferencias pueden observarse en la figura 26, en donde sobresalen los núcleos de SVM comparados con *Random Forest*. Sin embargo esto no debe dar por hecho que en el proceso de predicción el algoritmo de *Random Forest* obtenga un bajo desempeño.

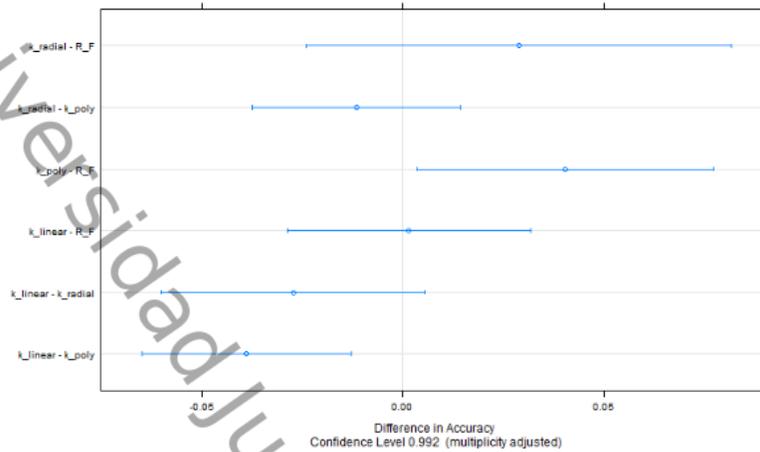


Figura 26. Resultados de diferencia en SVMs y Random Forest

3.2.6 Resultados de clasificación con base a optimización de *features*

Considerando las nuevas *features* obtenidas, así como la optimización de parámetros de los algoritmos SVM y *Random Forest*, se procede a medir el desempeño de estos en predicción. Los resultados de desempeño con nuevas *features* son presentados en la tabla 23.

A diferencia de los resultados en la optimización de parámetros, donde el núcleo polinomial de SVM tuvo un mejor desempeño, ahora en la evaluación de desempeño con las *features* optimizadas para el proceso de predicción, el algoritmo de *Random Forest* resulta poco más alto que los demás algoritmos. Sin embargo en el caso de la métrica de precisión equilibrada el algoritmo de SVM con núcleo radial obtiene un mejor desempeño.

Tabla 23. Resultados promedios de métricas

	Precisión total	Precisión equilibrada	Sensibilidad	Especificidad
K_linear	0.84624277	0.8583468	0.9380117	0.75657143
	0.01345802	0.01287573	0.01593038	0.02417855
k_radial	0.8517341	0.86167796	0.93235867	0.77295238
	0.01485677	0.01383666	0.02250189	0.03280663
k_poly	0.72418113	0.80467569	0.52923977	0.91466667
	0.07601224	0.03229329	0.26285412	0.11522699

	Precisión total	Precisión equilibrada	Sensibilidad	Especificidad
R_F	0.84865125	0.85777845	0.92787524	0.7712381
	0.01511305	0.01505062	0.02016065	0.02417312

Por otra parte, en el caso de la métrica de sensibilidad el núcleo lineal de SVM supera al núcleo radial. El núcleo polinomial de SVM obtiene más bajo en las métricas de precisión total, precisión equilibrada y sensibilidad, no obstante logra el desempeño más alto en la métrica de especificidad.

Las figuras 27, 28, 29 y 30 muestran una comparación del desempeño de cada uno de los algoritmos clasificadores, en cada una de las métricas (precisión total, precisión equilibrada, sensibilidad o especificidad), considerando los resultados del experimento antes de la optimización de *features* y los resultados con las *features* optimizadas.

En la figura 27 se presenta el desempeño del núcleo lineal con los resultados del primero (k_lineal v1) y segundo experimento (k_lineal v2). Como se puede observar hay una mejora en el desempeño en cada de las métricas debido a la optimización de *features*.

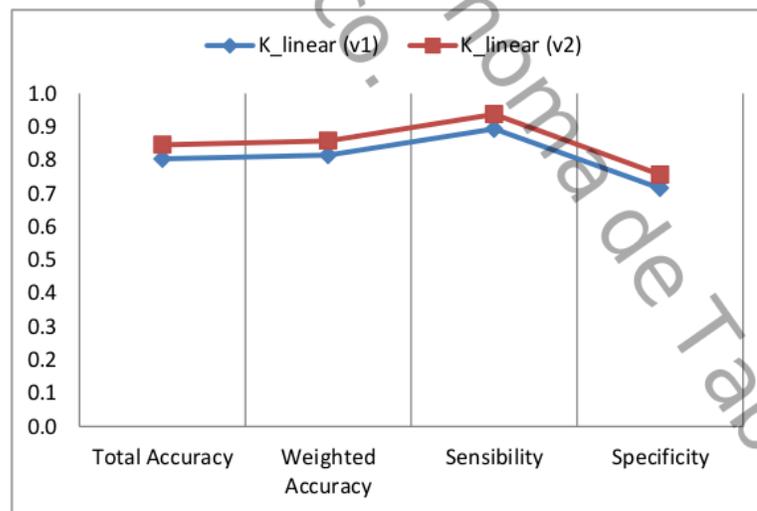


Figura 27. Comparación de desempeño del núcleo lineal

La comparación del desempeño de SVM con núcleo radial se presenta en la figura 28. El desempeño del núcleo radial crece en cada una de las métricas, incluso sobresale en la métrica de sensibilidad, sin embargo en la métrica de especificidad no se aprecia gran cambio.

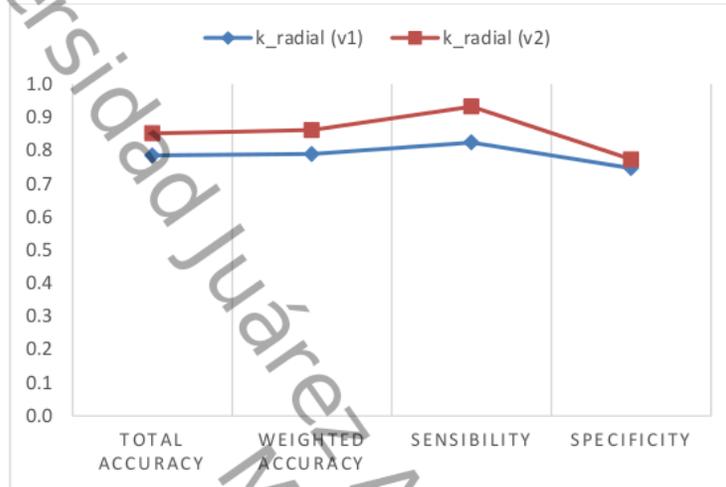


Figura 28. Comparación de desempeño del núcleo radial

En la figura 29 se indica el desempeño de la SVM con núcleo polinomial, en donde se presenta una mejora en las métricas de precisión total, precisión equilibrada y especificidad, sin embargo muestra un desempeño más bajo en la métrica de sensibilidad.

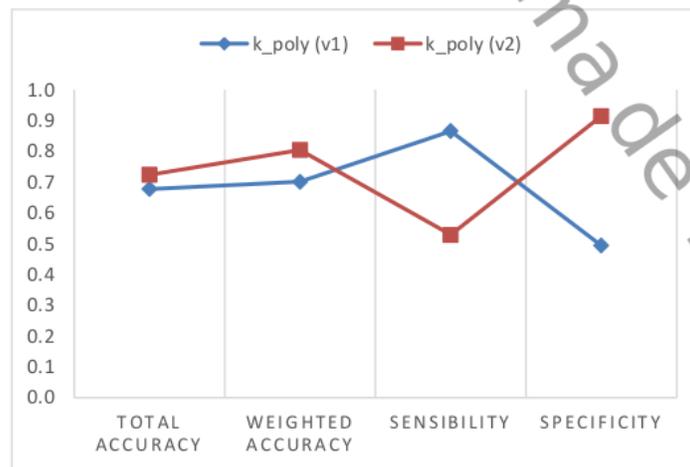


Figura 29. Comparación de desempeño del núcleo polinomial

La figura 30 muestra el desempeño de *Random Forest* con los resultados del primero (R_F v1) y segundo experimento (R_F v2). Como se puede observar hay una mejora en el desempeño en las métricas de precisión total y equilibrada, no obstante el desempeño en la métrica de sensibilidad baja a diferencia del resultado en el primer experimento, en la métrica de especificidad registra un cambio de mejora, aunque no es precisamente alto.

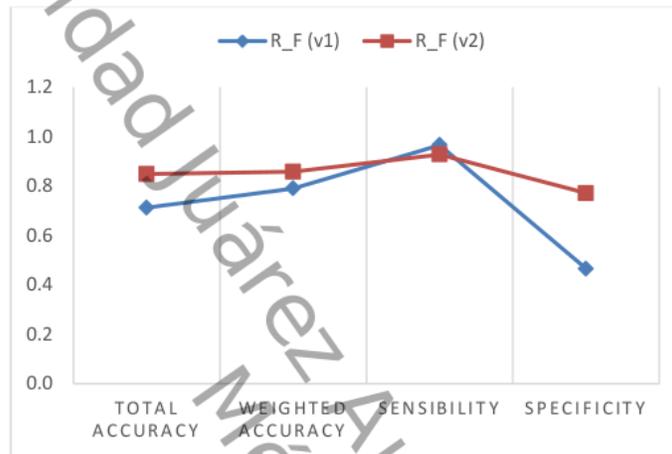


Figura 30. Comparación de desempeño del núcleo radial

Como se pudo observar en las anteriores figuras (27, 28, 29 y 30), los núcleos lineal y radial mejoran el desempeño mediante la optimización de *features*. SVM con núcleo polinomial y *Random Forest* presentan una mejora en las métricas de precisión total y precisión equilibrada, a pesar de que el núcleo polinomial decrece en la métrica de sensibilidad, en la métrica de especificidad alcanza el mejor desempeño en comparación con los demás algoritmos, por su parte el algoritmo de *Random Forest* decrece fuertemente en la métrica de especificidad.

Considerando que los resultados del núcleo radial de SVM en las métricas de precisión total y precisión equilibrada superan a lo demás núcleos y a *Random Forest*, se realiza el análisis ROC más para determinar por medio de los resultados obtenidos en las métricas de especificidad y sensibilidad (ver tabla 22), los valores predictivos, positivos y negativos, así como la curva ROC. La tabla presenta un extracto de los resultados obtenidos en este análisis.

Con base a los valores de la tabla 24, la figura 31 presenta la curva de ROC obtenida.

Tabla 24. Extracto de valores predictivos en núcleo Radial de SVM

P(+)	P(-)	VP+	FVP+	VP-	FVP-
0.2341	0.7659	0.0398	0.9602	0.9991	0.0009
0.2412	0.7588	0.0773	0.9227	0.9982	0.0018
0.2482	0.7518	0.1127	0.8873	0.9973	0.0027
0.2553	0.7447	0.1461	0.8539	0.9964	0.0036
0.2623	0.7377	0.1777	0.8223	0.9954	0.0046
0.2694	0.7306	0.2077	0.7923	0.9944	0.0056
0.2764	0.7236	0.2361	0.7639	0.9935	0.0065
0.2835	0.7165	0.2631	0.7369	0.9924	0.0076
0.2905	0.7095	0.2888	0.7112	0.9914	0.0086
0.2976	0.7024	0.3133	0.6867	0.9904	0.0096
0.3046	0.6954	0.3367	0.6633	0.9893	0.0107
0.3117	0.6883	0.3590	0.6410	0.9882	0.0118
0.3187	0.6813	0.3803	0.6197	0.9871	0.0129
0.3258	0.6742	0.4007	0.5993	0.9860	0.0140
0.3328	0.6672	0.4202	0.5798	0.9848	0.0152

Como se puede observar el comportamiento de la curva es muy similar a la curva de ROC obtenida en el Modelo SocialMining aplicando *Naïve Bayes* en combinación con el corpus de subjetividad (la gráfica de la curva ROC del Modelo SocialMining aplicando *Naïve Bayes* en combinación con el corpus de subjetividad, se mostró en la figura 18).

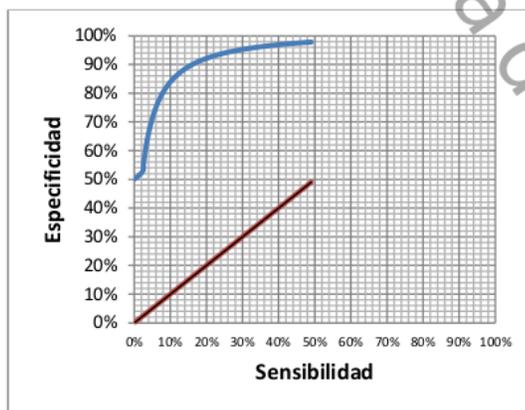


Figura 31. Curva de ROC en SocialMining aplicando SVM y Random Forest

Los resultados obtenidos en las métricas de sensibilidad y especificidad son muy parecidos, sin embargo lo más importante que se puede concluir, es que los resultados obtenidos con algoritmos de aprendizaje automático pueden mejorar aun sin la implementación de un corpus de subjetividad, no obstante es sumamente importante la selección de *features*, para prever un buen desempeño en el proceso de predicción y clasificación.

Universidad Juárez Autónoma de Tabasco.
México.

3.2.7 Discusión de Resultados

En la literatura se reportan numerosos clasificadores que pertenecen a distintas familias, como los estadísticos (análisis de discriminación lineal), los de inteligencia artificial (clasificadores basados en reglas), los de enfoques de conexión (redes neuronales) y otros (regresión, clúster) [106]. Estas familias de algoritmos clasificadores pueden ser aplicadas en diferentes contextos, sin embargo esto no significa que el desempeño obtenido de cada uno será el mismo. Por esta razón se considera importante realizar una revisión de la literatura con el fin de conocer el dominio de los clasificadores a utilizar. Para la selección de los algoritmos clasificadores, algunos investigadores [106] sugieren considerar el tipo y la cantidad de datos con los cuales se trabajará, el método de entrenamiento a realizar e incluso el equipo de cómputo con el que se cuenta. La figura 32 presenta un panorama para la elección de algunos algoritmos considerando características como la cantidad de datos, etiquetado de datos, categorías definidas o estructura.

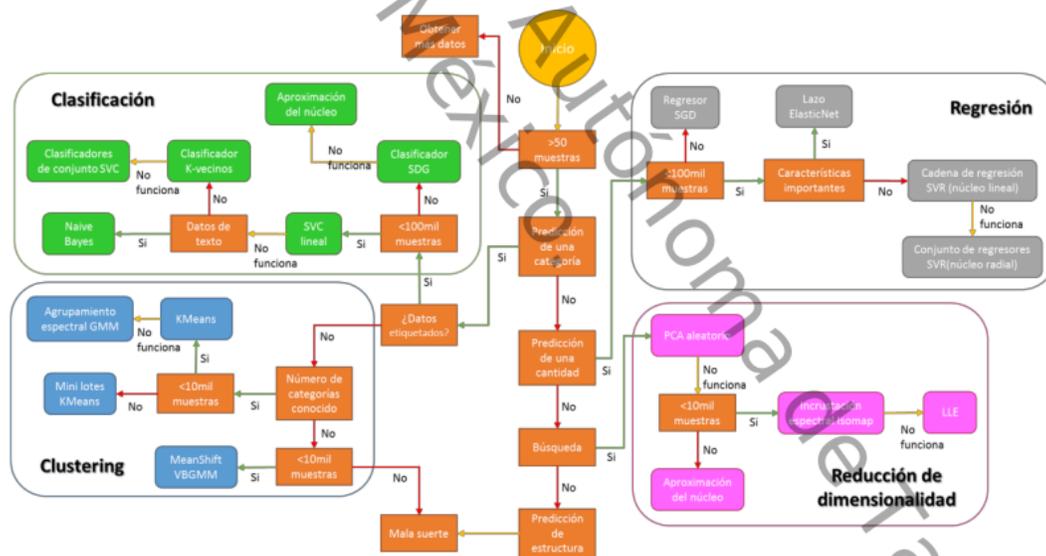


Figura 32. Enfoques diversos de aprendizaje automático según los materiales a utilizar [107]

Según Fernández-Delgado [106] la mayoría de los investigadores buscan comparar el desempeño de precisión con clasificadores de la misma familia, comúnmente familias del clasificador favorito, esto provoca que el investigador no conozca que otros clasificadores funcionan mejor que los que ya conoce y limita

su conocimiento a unos cuantos algoritmos clasificadores. Por esta razón en este trabajo se escogieron algoritmos clasificadores basados en estadística y en aprendizaje automático, con el fin de comparar el desempeño en precisión. Asimismo uno de los objetivos del presente estudio fue determinar cuál es el algoritmo que logra obtener el más alto desempeño en clasificación de comentarios de la evaluación docente. Sin embargo los resultados obtenidos indican que el desempeño de los algoritmos depende de la naturaleza de los datos a utiliza, la selección de *features*, la optimización de parámetros y recursos como los léxicos o corpus de subjetividad.

En el Modelo SocialMining la SVM con núcleos lineal y radial obtiene los resultados más viables en las métricas de precisión total y equilibrada así como en la sensibilidad. Por lo cual la implantación de estos dos núcleos puede resultar ser de beneficio para el desempeño de clasificación del Modelo de SocialMining. No obstante, se resalta nuevamente que la selección de *features* es importante para la construcción de cualquier modelo de aprendizaje.

La presencia de *features* irrelevantes generaliza bajos resultados en el desempeño de la predicción. Por lo tanto la selección de *features* es el proceso que se considera más importante para permitir al Modelo SocialMining tener flexibilidad y mejores resultados en el proceso de clasificación de comentarios. Por lo cual se concluye que una selección adecuada de *features* permitirá una representación efectiva de los datos, dando por resultado un mejor desempeño en procesos de clasificación y predicción.

En la siguiente sección se presenta la contribución de este trabajo, el diagrama del algoritmo general del Modelo SocialMining aplicando aprendizaje automático.

3.2.8 Algoritmo general del Modelo SocialMining

La figura 33 presenta el diagrama general del Modelo SocialMining.

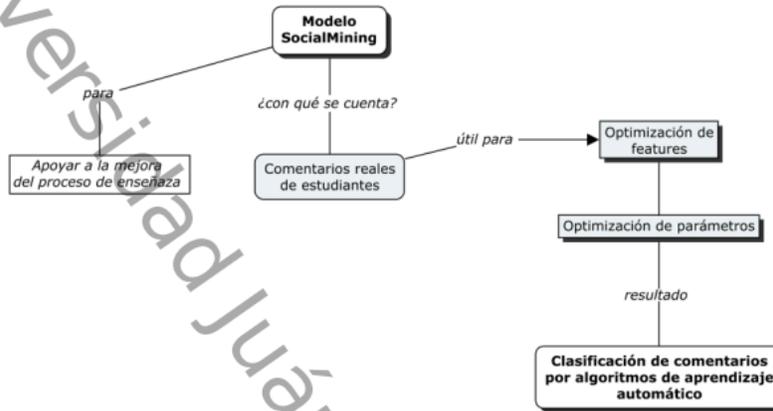


Figura 33. Diagrama general del Modelo SocialMining aplicando aprendizaje automático

Por último los algoritmos del proceso de optimización de *features* (ver figura 34) y del proceso de optimización de parámetros (ver figura 35), se presentan enseguida.

- Algoritmo para optimización de features**
1. Carga de corpus de comentarios
 2. Ejecución de proceso de limpieza en corpus de comentarios
 3. Transformación de los datos (matriz de comentarios)
 4. Obtención de términos frecuentes
 5. Limpieza de términos repetitivos irrelevantes
 6. Partición del conjunto de datos resultante
 - 6.1 Usar 2/3 del conjunto de datos para entrenamiento
 - 6.2 Usar 1/3 del conjunto de datos para prueba
 7. Ejecución del método *Random Forest Importance Measure* (VIM)
 - 7.1 Cálculo del peso de los términos del conjunto de entrenamiento
 - 7.2 Selección de los términos con peso más alto
 8. Configuración de método de validación cruzada de 10 iteraciones para realizar entrenamiento
 - 8.1 Configuración de vector con 30 semillas para trabajar con diferentes conjuntos de comentarios
 9. Obtención de predicción de clases de comentarios positivo, negativo, neutral)
 10. Obtención de matriz de confusión
 11. Cálculo de las métricas de precisión total y equilibrada.

Figura 34. Algoritmo para optimización de features

Algoritmo para optimización de parámetros

1. Cargar conjunto de datos de features
2. Configuración de método de validación cruzada de 10 iteraciones para realizar entrenamiento
 - 2.1 Selección de parámetros temporales empleando el método de Hsu
 - 2.2 Entrenamiento de algoritmos de aprendizaje automático(SVM y *Random Forest*)
 - 2.3 Obtención de parámetros sugeridos con base al entrenamiento
3. Aplicar nuevamente el método de Hsu para entrenamiento de los algoritmos con los parámetros sugeridos
4. Validación de parámetros
5. Cálculo del desempeño de los algoritmos en la validación de parámetros

Figura 35. Algoritmo para optimización de parámetros

Capítulo IV

Conclusiones y trabajo futuro

4.1 Conclusiones

En este trabajo se diseñó y desarrolló un Modelo de Análisis de Emociones denominado SocialMining enfocado a apoyar la evaluación del desempeño docente mediante diferentes técnicas reportadas como efectivas en la literatura. A pesar de que el Modelo SocialMining está enfocado hacia la evaluación del desempeño docente, es posible orientarlo a diferentes contextos como la política o mercadotécnica. Esto debido a que el Modelo muestra los procesos necesarios para analizar emociones en textos, los cuales son obtenidos mediante twitter o bases de datos, además incluye algoritmos reconocidos por su alto desempeño en procesos de predicción y clasificación. Principalmente la inclusión de algoritmos de aprendizaje automático para apoyar a la mejora de la educación de estudiantes de nivel superior en México es la motivación de este trabajo.

La realización de esta tesis doctoral acerca del análisis de emociones ha traído diversos logros y retos, ya que ha sido un trabajo que a pesar de que ha cumplido satisfacciones y generado diversas publicaciones en congresos nacionales e internacionales, así como revistas científicas importantes, también ha dado mucho de qué hablar, enfrentando preguntas como: *¿qué ocurre si el Modelo SocialMining clasifica comentarios sarcásticos en clases que no corresponden?*, o preguntas como: *¿Si Facebook ya implementó las emociones mediante emoticonos (ver figura 36), entonces cuál es la innovación del Modelo SocialMining?*. Es común que surjan este tipo de preguntas debido a que el área de análisis de emociones es relativamente nueva y aún se desconocen los alcances o niveles de esta área. En la actualidad el área del análisis de emociones se ha extendido a diferentes contextos y se ha combinado con grandes áreas de

conocimiento como: el procesamiento de lenguaje natural, la minería de datos, minería de textos, semántica e inteligencia artificial con aprendizaje automático y aprendizaje profundo (conocido en inglés como *Deep Learning* [108]), incluso también se relaciona con el área de macrodatos (conocido en inglés como *Big Data* [109]), por lo cual es recomendable definir un alcance y un objetivo. Debido a los diferentes enfoques del análisis de emociones no es posible abarcar en una investigación todos los problemas que conlleva esta área, como por ejemplo: el sarcasmo, la desambiguación semántica, análisis sintáctico o lingüístico, recuperación de emociones, recuperación de sentimientos, análisis de subjetividad, creación o adecuación de léxicos semánticos afectivos, entre otros.

Asimismo se debe reconocer el enfoque de ciertas aplicaciones, por ejemplo volviendo al caso de que Facebook ha implementado emoticonos para reconocer emociones, tal aplicación funciona siempre y cuando el usuario elija cierto emoticón para evaluar la publicación y aunque es posible que esta red social implemente un sistema para evaluar emociones como lo ha hecho Google, son enfoques diferentes y utilizan recursos diferentes.



Figura 36. Emociones en Facebook

A pesar de que este trabajo está dirigido al contexto de la educación buscando apoyar la evaluación del desempeño docente, es posible aplicarlo a otro contextos, ya que lo más importante es que se diseñó un modelo (SocialMining), del cual se conoce la metodología, técnicas y recursos que pueden contribuir a la

mejora del proceso de clasificación, por lo cual bien podría ser aplicado en la política, la psicología, la medicina, entre otros.

Considerando los objetivos específicos del presente trabajo doctoral se presentan enseguida las conclusiones de los mismos.

Objetivo 1. Mediante el análisis de comentarios de la evaluación docente se logró realizar la conformación del corpus de subjetividad en español, el cuál incrementó el desempeño del algoritmo clasificador *Naïve Bayes*, ya que incluyó diferentes términos distintivos de la evaluación docente, asimismo la inclusión de coloquialismos fue necesaria debido al lenguaje con el cual se expresan los jóvenes de hoy en día. El etiquetado en positivo, negativo y neutral, así como la asignación de fuerza subjetiva en los términos distintivos permitió al Modelo SocialMining calcular un *score* mediante *Naïve Bayes* y obtener la clasificación del comentario según los términos que encontrará en el mismo y en el corpus de subjetividad. Cabe resaltar que para la creación del corpus de subjetividad se tomaron en consideración diversas investigaciones sobre léxicos semánticos afectivos [45, 110]. La investigación de Carrillo [111] y Jourka [94] resultaron ser de gran ayuda debido a la metodología que los autores proponen, la cual se adecuó en la creación del corpus de subjetividad en idioma español considerando coloquialismos del Estado de Aguascalientes México.

Objetivo 2. La aplicación del algoritmo *Naïve Bayes* para clasificar comentarios de la evaluación docente se considera factible debido a que permite ser modificado, incluyendo variables como un *score* para aumentar el desempeño en clasificaciones y no solo clasificar con base a probabilidades.

Objetivo 3. La aplicación de técnicas de aprendizaje automático en el Modelo SocialMining también se consideró factible, ya que con base a los experimentos realizados se pudo observar que los algoritmos SVM y *Random Forest* obtuvieron resultados casi iguales a los obtenidos con *Naïve Bayes* combinado con el corpus de subjetividad, sobrepasando un 80% de precisión en clasificación. A pesar de no alcanzar exactamente el resultado obtenido con *Naïve Bayes* y el corpus de subjetividad (poco más de 90%), los resultados obtenidos con SVM con núcleo radial y lineal se consideran viables según investigaciones de la literatura. Asimismo se observó que con el experimento de optimización de *features*, hay un incremento en la mejora del desempeño de los clasificadores, por lo que se puede

concluir que la optimización de *features*, influye en el entrenamiento de los algoritmos para obtener los parámetros adecuados. Cabe resaltar que para la optimización de *features* es importante que el corpus se encuentre balanceado, es decir, que se componga de cantidades similares de comentarios positivos y negativos, ya que son las clases que el Modelo SocialMining busca clasificar.

Es importante resaltar que gracias al estudio del análisis de emociones y algoritmos de aprendizaje automático, este trabajo concluye con un modelo que se puede adaptar a otras áreas, además el que sea un modelo para el idioma español, prevé oportunidades de aplicación en México y América Latina. El área del análisis de emociones en México aún se considera es un tema innovador ya que existen pocas investigaciones y trabajos desarrollados al respecto.

Considerando las hipótesis planteadas en este trabajo se presentan enseguida las respuestas de las mismas.

Hipótesis 1 – cumplida. La combinación del corpus de subjetividad con el algoritmo de Naïve Bayes contribuyó al Modelo SocialMining, para alcanzar un 90% en precisión total y equilibrada, en la clasificación de comentarios de la evaluación docente.

Hipótesis 2 – cumplida. La optimización de *features* permitió a los algoritmos SVM y *Random Forest* disminuir el margen de error en la clasificación de comentarios.

Considerando las preguntas de investigación planteadas en este trabajo se presentan enseguida las respuestas de las mismas.

Pregunta de investigación 1. Sobre cuáles son los términos distintivos que deben incluirse en el corpus de subjetividad y que fuerza subjetiva se debe asignar, no es posible responder e indicar aquellos que deben ser incluidos, ya que puede variar con respecto a la región donde se aplique por aquello de los coloquialismos, además los términos distintivos pueden variar con el tiempo. Por lo cual se sugiere mantener en continua actualización el corpus de subjetividad para contribuir al buen desempeño de la clasificación.

Pregunta de investigación 2. Sobre la técnica de aprendizaje automático que obtiene un menor margen de error, es SVM con núcleo lineal y radial, sin embargo se considera que una optimización más exhaustiva en la selección de

features puede aumentar el desempeño en clasificación del algoritmo *Random Forest*.

Debido a que el Modelo SocialMining no es exhaustivo en cuanto a la optimización de *features*, se considera que es un área de mejora en el mismo con el fin de mejorar aún más el desempeño en clasificación, la siguiente sección presenta el trabajo futuro propuesto para continuar esta investigación doctoral.

4.2 Trabajo futuro

Fase de optimización de *features*. Como trabajo futuro se plantea incluir una fase al Modelo SocialMining enfocada en el proceso de selección y optimización de *features* considerando algoritmos de aprendizaje correspondientes a: métodos de filtrado (*Filter Methods*), métodos de contenedor (*Wrapper Methods*) o métodos embebidos (*Embedded Methods*). La parte fundamental de esta investigación sería para determinar qué tipo algoritmos influyen mejor en la optimización de *features* considerando la naturaleza de los datos (longitud del texto, idioma, contexto, enfoque) con los cuales se trabajaría.

Interfaz de usuario. Del mismo modo se planea la creación de una interface para el usuario considerando áreas como ingeniería de software e interacción humano computador, con el fin de distribuir la aplicación o bien conectarla a otras aplicaciones, por medio de las cuales se pudiera obtener acceso a datos (por ejemplo el EED) y realizar el proceso de clasificación de manera casi automática, simplificando al usuario del Modelo SocialMining el manejo de reportes con base al resultado de análisis de comentarios.

Análisis de sentimientos. La inclusión de análisis de sentimientos es también parte del trabajo futuro en SocialMining, con el fin de poder detectar más allá de emociones positivas, negativas y neutrales. Identificar el sentimiento del usuario en textos escritos, puede contribuir a proyectos aplicados para el bienestar de la comunidad, como el proyecto de RedSiam [112], enfocado en prevenir la violencia contra la mujer en una relación afectiva.

Anexo A

Glosario de términos

A

Algoritmo de clasificación (binaria o multiclase): Algoritmo que tiene la tarea de asignar una clase, es decir, predecir a que clase pertenece un conjunto de datos.

Aprendizaje automático: en inglés *Machine Learning*, es la rama de la Inteligencia Artificial que tiene como objetivo desarrollar técnicas que permitan soluciones inteligentes para tareas complejas.

Aprendizaje supervisado: es una técnica de aprendizaje automático; su objetivo es crear una función capaz de predecir el valor de salida correspondiente a cualquier objeto de entrada válida, después de haber visto una serie de ejemplos, los datos de entrenamiento.

Área bajo la Curva ROC: en inglés *Area Under ROC Curve (AUC)*, es un indicador que permite interpretar la evaluación arrojada por la curva ROC con respecto a algoritmos de clasificación.

C

Corpus de comentarios: Conjunto de comentarios que son recabados y agrupados para su procesamiento.

Curva ROC: en inglés *ROC Curve*, es una representación gráfica de la sensibilidad frente a la especificidad en un sistema clasificador binario.

E

Especificidad: en inglés *Specificity*, la especificidad de un clasificador representa la probabilidad de que un dato negativo sea clasificado como tal.

I

Inteligencia Artificial: en inglés *Artificial Intelligence*, la IA agrupa un conjunto de técnicas que, mediante circuitos electrónicos y programas avanzados de computadora, busca imitar procedimientos similares a los procesos inductivos y deductivos del cerebro humano.

M

Máquina de soporte vectorial: en inglés *Support Vector Machine (SVM)*, es un algoritmo de aprendizaje supervisado relacionado con problemas de clasificación y regresión. Mediante datos de entrenamiento es posible asignar clases y entrenar una SVM para construir un modelo que prediga la clase de una nueva muestra.

Matriz de confusión: en inglés *Confusion Matrix*, es una herramienta que permite la visualización del desempeño de un algoritmo que se emplea en aprendizaje supervisado. Cada columna representa el número de predicciones de cada clase, mientras que cada fila representa los valores de clasificación real.

Método train-test: es un método que divide los datos disponibles en train y test, manteniendo los conjuntos anteriores independientes, con la finalidad de evaluar modelos de predicción.

Minería de datos: en inglés *Data Mining*, es el proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos.

Minería de textos: en inglés *Text Mining*, se refiere al proceso de derivar información nueva de textos.

N

Núcleo: en inglés *Kernel*, es una función de similitud. Esta función toma dos entradas y devuelve las similitudes encontradas en las mismas. Este

procedimiento se realiza mediante un algoritmo de aprendizaje automático, en esta trabajo mediante la SVM.

P

Precisión ponderada: en inglés *Balanced Accuracy*, permite una mejor estimación del desempeño de un clasificador cuando en un conjunto de datos existe una distribución desigual de dos clases mediante la ecuación de precisión.

Precisión total: en inglés *Accuracy*, es la proporción de elementos predichos por un modelo, que son elementos reales y clasificados correctamente.

S

Sensibilidad: en inglés *Sensitivity*, la sensibilidad de un clasificador representa la probabilidad de que un dato positivo sea clasificado como tal.

T

Términos distintivos: en inglés *features*, son características distintivas de un elemento, en esta trabajo corresponden a términos distintivos de la evaluación docente que apoyan el proceso de clasificación de comentarios. La elección de características independientes, discriminativas e informativas es fundamental para la eficiencia de los algoritmos de clasificación.

V

Validación cruzada de 10 iteraciones: en inglés *10-cross-validation*, es una técnica utilizada para evaluar los resultados de un análisis estadístico, en este trabajo se usa para el entrenamiento de los algoritmos de aprendizaje automático utilizados en el Modelo para el proceso de clasificación.

Anexo B

Preguntas de la Evaluación

Docente de la UPA

Las preguntas realizadas son las que se muestran a continuación, éstas se evalúan del 1 al 10, siendo la máxima 10 (Muy satisfecho) y la mínima 1 (Muy insatisfecho).

1. Conocimientos - Considero que el profesor demuestra en todas las sesiones el dominio de los contenidos de la asignatura.
2. Habilidades - Planeación - Considero que el profesor prepara prácticas, problemas o casos de estudio, en donde puedo desarrollar mis conocimientos, habilidades y actitudes.
3. Habilidades - Planeación - Considero que los materiales didácticos adicionales al pizarrón tales como diapositivas, programas, simulaciones, prototipos, etc. son útiles para mi aprendizaje de los contenidos de la materia.
4. Habilidades - Planeación - Considero que el profesor utiliza de manera efectiva tecnologías de la información y comunicación (plataforma educativa, WEB, objetos de aprendizaje, medios interactivos, redes sociales, etc.) para favorecer mi aprendizaje
5. Habilidades - Mediación - Considero que el profesor se dirige a mí de manera respetuosa, me permite expresar y acepta mis ideas, fomentando positivamente la participación del grupo.
6. Habilidades - Mediación - Considero que el profesor logra aclarar las dudas en forma efectiva.
7. Habilidades - Mediación - Considero que el profesor realiza preguntas que favorecen mi reflexión y mi aprendizaje.
8. Habilidades - Mediación - Considero que el profesor vincula la práctica profesional con el contenido de sus clases.
9. Habilidades - Evaluación - El profesor me dio a conocer al inicio del cuatrimestre los objetivos a lograr y los contenidos a desarrollar en cada unidad, los instrumentos de

evaluación (rúbricas, listas de cotejo o guías de observación) y el porcentaje en que cada una sería considerada para obtener mi calificación total de la materia.

10. Habilidades - Evaluación - Considero que los instrumentos de evaluación utilizados por el profesor (listas de cotejo, rúbricas o guías de observación) se apegan a los temas desarrollados en clase y consideran mis conocimientos, habilidades y actitudes para determinar mi calificación.

11. Habilidades - Evaluación - Considero que el profesor realiza las evaluaciones en las fechas acordadas, retroalimenta oportunamente y se apega a la forma de evaluación de la materia dada a conocer al inicio del curso.

12. Actitudes - Considero que el profesor propicia un ambiente de respeto y de comunicación armónica entre los compañeros del grupo durante las sesiones.

13. Actitudes - Considero que el profesor tiene disposición para reponer las sesiones cuando por situaciones administrativas o de fuerza mayor no se impartió la clase.

14. Actitudes - Considero que el profesor asiste a todas sus sesiones y respeta las horas de entrada y salida de su clase.

15. Actitudes - Considero que el profesor acepta retroalimentación respetuosa y fundamentada respecto a su curso.

16. Actitudes - Considero que el profesor tiene disposición para aclarar dudas y ofrecer asesorías fuera de clases.

17. Actitudes - Considero que el profesor actúa en congruencia con los reglamentos y valores de la universidad.

18. Actitudes - Considero que el profesor se expresa respetuosamente de la institución y de los diferentes miembros de la comunidad universitaria.

19. Evaluación General - - Considero que el desempeño general del profesor ha sido excelente.

20. Evaluación General - - Me gustaría tomar otro curso con este maestro.

Anexo C

Evidencias de Evaluación docente en Twitter

A continuación se muestran algunos de los comentarios hechos por los alumnos correspondientes a las materias de programación lineal (#ProgLsei), calidad de software (#CSWsei), administración del conocimiento (#Aknowsei), planeación de recursos materiales (#MRPsei), informática estratégica (#InfoEsei) y finanzas para ingenieros (#Finanzassei).

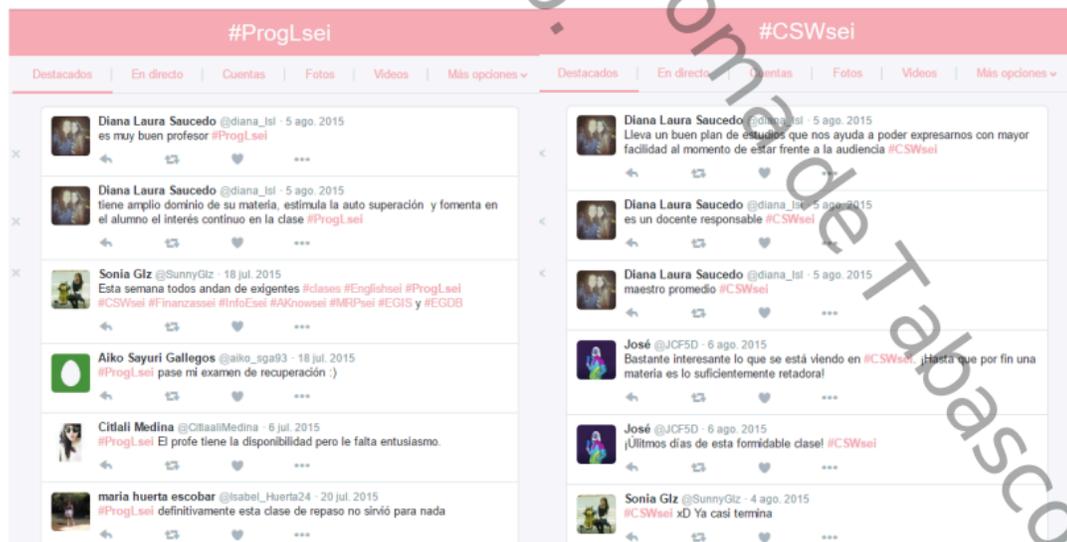


Figura 37. Comentarios en Twitter #ProgLsei

Figura 38. Comentarios en Twitter #CSWsei



Figura 39. Comentarios en Twitter #Aknowsei

Figura 40. Comentarios en Twitter #MRPsei



Figura 41. Comentarios en Twitter #InfoEsei



Figura 42. Comentarios en Twitter #Finanzassei

Bibliografía

1. Wiebe, J.R., E., *Creating Subjective and Objective Sentence Classifiers from Unannotated Texts*. In Proceedings of the 6th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing-05), 73-99., 2005.
2. Pang, B., & Lee, L., *Opinion mining and sentiment analysis*. Foundations and Trends in Information Retrieval, Volume 2 Issue 1-2, 1 - 135, doi: <http://dx.doi.org/10.1561/15000000011>, 2008.
3. Turney, D.P., P., *From frequency to meaning: Vector space models of semantics*. Journal of Artificial Intelligence Research, 37:141–188., 2010.
4. Turney, P., *Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews*. ACL '02 Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Pages 417-424 2002.
5. Wilson, T., Wiebe, J., and Hoffmann, P., *Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis*. Computational Linguistics, vol. 35(3), pp. 399–433., 2009.
6. Brooke, J., *A semantic approach to automated text sentiment analysis*. Master's thesis, Simon Fraser University., 2009.
7. Carrillo de Albornoz, J., Chugur, I., & Amigó, E., *Using an Emotion-based Model and Sentiment Analysis Techniques to Classify Polarity for Reputation*. In proceedings CLEF 2012 Labs and Workshop Notebook Paper. (<http://goo.gl/UltHhg>), 2012.
8. Bharathisindhu, P., & Brunda, S., *Identifying e-learner's opinion using automated sentiment analysis in e-learning*. IJRET: International Journal of Research in Engineering and Technology. Volume: 03 Issue: 01. doi: <http://dx.doi.org/10.15623/ijret.2014.0319086>, 2014.
9. Balahur, A., Hermida, J.M., & Montoyo, A., *EmotiNet: a Knowledge Base for Emotion Detection in Text Build on the Appraisal Theories*. In Lecture Notes in Computer Science, proceedings of NLDB 2011., 2011.
10. Iwashita, M., Shimogawa, S., & Nishimatsu, K., . *Semantic Analysis and Classification Method for Customer Enquiries in Telecommunication Services*. . Engineering Applications of Artificial Intelligence, 24(8), 1521-1531., 2011.

11. Pong-inwong, C.R., W., *TeachingSenti-Lexicon for Automated Sentiment Polarity Definition in Teaching Evaluation*. In Semantics, Knowledge and Grids (SKG), 10th International Conference, 84 – 91, Beijing: IEEE., 2014.
12. Altrabsheh, N., Gaber, M. and Cocea, Mihaela, *SA-E: Sentiment Analysis for Education*. In: 5th KES International Conference on Intelligent Decision Technologies. (<http://goo.gl/5oTSSl>), 2013.
13. Gupte, A., Joshi, S., Gadgul, P., & Kadam, A., *Comparative Study of Classification Algorithms used in Sentiment Analysis*. International Journal of Computer Science and Information Technologies; Vol. 5 Issue 5, 6261 - 6264. (<http://goo.gl/tiilHBT>), 2014.
14. Baldoni, M.B., Cristina; Patti, Viviana; Rena, Paolo., *From tags to emotions: Ontology-driven sentiment analysis in the social semantic web*. *Intelligenza Artificiale* 6(1): 41-54 2012.
15. Nielsen, *State of the Media: The Social Media Report 2012*. NM Incite, 2012.
16. Pak, A.a.P., P., *Twitter as a Corpus for Sentiment Analysis and Opinion Mining*. Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23 May 2010, Valletta, Malta, 2010.
17. Kououmpis, E., Wilson, T. and Moore J., *Twitter Sentiment Analysis: The Good the Bad and the OMG!* Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, 2011.
18. Singh, R.a.K., Rajdeep, *Sentiment Analysis on Social Media and Online Review* International Journal of Computer Applications (0975 – 8887) Vol. 121 – No.20., 2015.
19. I.Hemalatha, G.P.S.V., and A.Govardhan, *Sentiment Analysis Tool using Machine Learning Algorithms*. International Journal of Emerging Trends & Technology in Computer Science. Volume 2, Issue 2., 2013.
20. Pang, B., & Lee, L., *Thumbs up? Sentiment Classification using Machine Learning Techniques*. EMNLP '02 Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10. Pages 79-86. doi>10.3115/1118693.1118704, 2002.
21. Das, S., & Chen, M., *Yahoo! for Amazon: Extracting market sentiment from stock message boards*. In *Proceedings of the Asia Pacific Finance Association annual conference*. 2001.
22. Liu, B., *Sentiment Analysis and Subjectivity*. In N. Indurkhia & F. J. Damerau (Eds.). *Handbook of natural language processing*. pp. 627-666. Chapman and Hall: CRC Press. 2010.
23. Bing, L., *Sentiment Analysis and Opinion Mining (Synthesis Lectures on Human Language Technologies)*. Morgan & Claypool Publishers., 2012.
24. Balahur, A., *Methods and Resources for Sentiment Analysis in Multilingual Documents of Different Text Types*. Tesis de Doctorado. Departamento de Lenguajes y Sistemas Informáticos. Universidad de Alicante. , 2011.

25. Ortony, A., Clore, G., & Foss, M. A. , *The Psychological Foundations of the Affective Lexicon*. . Journal of Personality and Social Psychology, 53, 751-766, 1987.
26. Picard, R.W., *Affective computing. Technical Report 321, MIT*. 1995.
27. Wang, G.S., J.; Ma, J.; Xu, K. & Gu, J., *Sentiment classification: The contribution of ensemble learning*. Decision Support Systems (2013), <http://dx.doi.org/10.1016/j.dss.2013.08.002>, 2013.
28. Feldman, R., *Techniques and applications for sentiment analysis*. Communications of the ACM. 56(4), 82-89, 2013.
29. Tsur, O., Davidov, D. & Rappoport, A., *A great catchy name: Semi-supervised recognition of sarcastic sentences in online product reviews*. In Fourth International AAAI Conference on Weblogs and Social Media., 2010.
30. Ferran, P., & Hurtado, L., *Political Tendency Identification in Twitter using Sentiment Analysis Techniques*. The 25th International Conference on Computational Linguistics 2014.
31. Kaur, G., & Singla, A., *Sentimental Analysis of Flipkart reviews using Naïve Bayes and Decision Tree algorithm*. International Journal of Advanced Research in Computer Engineering & Technology. Vol. 5, issue 1, pp. 148-153, 2016.
32. Wiebe, J., *Tracking point of view in narrative*. Computational Linguistics, 20(2), 233-287., 1994.
33. Banfield, A., *Unspeakable sentences: Narration and Representation in the Language of Fiction*. Routledge and Kegan Paul., 1982.
34. Quirk, R., *A Comprehensive Grammar of the English Language*. Longman, London, UK., 1985.
35. Wiebe, J., Wilson, T., Bruce, R., Bell, M., & Martin, M., *Learning subjective language*. Computational Linguistics, 30(3), 277-308.. 2004.
36. Stone, P., Dumphy, D., Smith, M., & Ogilvie, D., *The General Inquirer: A Computer Approach to Content Analysis*. The MIT Press, Cambridge, USA., 1966.
37. Macleod, C., Grishman, R., & Meyers, A., *Creating a common syntactic dictionary of english*. In Proceedings of the International Workshop on Sharable Natural Language Resources., 1994.
38. Yu, D.H., V., *Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences*. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2003: 129-136, Shapporo, Japan., 2003.
39. Wiebe, J., Wilson, T., & Cardie, C., *Annotating expressions of opinions and emotions in language*. In Language Resources and Evaluation, vol. 39(2-3),119-122., 2005.
40. Banea, C., Mihalcea, R. & Wiebe, J., *Multilingual subjectivity: are more languages better?* In Proceedings of the International Conference on Computational Linguistics (COLING 2010), p. 28-36, Beijing, China., 2010.

41. Pang, B.L., L., *A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts*. In Proceedings of the ACL 2004, pp.271-278, Barcelona, Spain., 2004.
42. Strapparava, C.V., A., *Wordnet-affect: an affective extension of wordnet*. In Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004), pp.1083-1086, Lisbon, Portugal., 2004.
43. Balahur, A.M., A., *An incremental multilingual approach to forming a culture dependent emotion triggers lexical database*. In Proceedings of the Conference of Terminology and Knowledge Engineering (TKE 2008). 2008.
44. Cerini, S., Compagnoni, V., Demontis, A., Formentelli, M., & Gandini, G., *Micrownop: A gold standard for the evaluation of auto-matically compiled lexical resources for opinion mining*. Milano, IT., 2007.
45. Carrillo de Albornoz, J., Plaza, L., & Gervás, P., *SentiSense: An easily scalable concept-based affective lexicon for Sentiment Analysis*. The 8th International Conference on Language Resources and Evaluation. (<http://goo.gl/RAcRO9>), 2012.
46. Maks, I., and Vossen, P., *A lexicon model for deep sentiment analysis and opinion mining applications*. *Decision Support Systems*, 53, pp. 680-688. 2012.
47. Jindal, N.L., B., *Identifying comparative sentences in text documents*. SIGIR '06 Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval Pages 244 - 251. ISBN:1-59593-369-7 doi>10.1145/1148170.1148215, 2006.
48. I.H. Witten, E.F., and M. A. Hall., *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, third edition., 2011.
49. Kumar, L., & Bhatia, P. K., *Text Mining: Concepts, Process and Applications*. *Journal of Global Research in Computer Science*, vol. 4(3), pp. 36-39., 2013.
50. Witten, I.H., Don, K. J, Dewsnip, M. and Tablan, V., *International Journal on Digital Libraries*, vol. 4, no. 1, pp. 56-59. 2003.
51. Surjandari, I., Naffisah, M. S., & Prawiradinata, M. I. , *Text Mining of Twitter Data for Public Sentiment Analysis of Staple Foods Price Changes*. *Journal of Industrial and Intelligent Information*, vol, 3(3). 2015.
52. Contreras, M., *Minería de texto: una visión actual*. *Biblioteca Universitaria*, vol. 17(2), pp. 129-138. 2014.
53. Cai, Y., & Sun, J. T., *Text Mining*. In *Encyclopedia of Database Systems*, pp. 3061-3065. Springer US., 2009.
54. Sebastiani, F., *Machine learning in automated text categorization*. *ACM computing surveys (CSUR)*, vol 34(1), pp. 1-47., 2002.
55. Andrieu, C., De Freitas, N., Doucet, A., & Jordan, M. I. , *An introduction to MCMC for machine learning*. *Machine learning*, vol. 50(1-2), pp. 5-43. 2003.
56. Bratko, I., Michalski, R. S., & Kubat, M., *Machine learning and data mining: methods and applications*. 1999.

57. Esuli, A.S., F., *SENTIWORDNET: a publicly available lexical resource for opinion mining*. In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC 2006). pages 417-422, Genoa, Italy., 2006.
58. Vapnick, V., *Statistical Learning Theory*. Wiley, New York, 1998.
59. Blum, A.L., & Langley, P., *Selection of relevant features and examples in machine learning*. *Artificial intelligence*, vol. 97(1), pp. 245-271. 1997.
60. Nan, L.a.D., D., *Using text mining and sentiment analysis for online forums hotspot detection and forecast*. Decision Support Systems archive. Volume 48 Issue 2, January, 2010. Pages 354-368. Elsevier Science Publishers B. V. Amsterdam, The Netherlands, The Netherlands, 2010.
61. Bayes, T.a.P., M. , *An essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, frs communicated by mr. price, in a letter to john canton, amfrs*. *Philosophical Transactions (1683-1775)*, pp. 370-418. 1763.
62. Breiman, L., *Random Forest*. *Machine Learning*, 45, pp. 5–32., 2001.
63. Breiman, L., *Bagging predictors*. *Machine Learning*, 24(2), pp. 123 -140, 1996.
64. Schapire, Y.F.a.R., *Experiments with a new boosting algorithm*. In Proceedings of the Thirteenth International Conference on Machine Learning, pp 148 -156, Bari, Italy, 1996.
65. John, R.K.a.G., *Wrappers for feature subset selection*. *Artificial intelligence*. Vol. 97, no. 1-2, pp. 273–324., 1997.
66. Canul-Reich, J., *An iterative feature perturbation method for gene selection from microarray data*. University of South Florida. Graduate Theses and Dissertations. <http://scholarcommons.usf.edu/etd/1588>, 2010.
67. Liu., M.D.a.H., *Feature selection for clustering*. In T. Terano, H. Liu, and A. L. P. Chen, editors, *Knowledge Discovery and Data Mining. Current Issues and New Applications*, vol. 1805. *Lecture Notes in Computer Science*, pp. 110 -121., 2000.
68. I. Inza, P.L., R. Etxeberria, and B. Sierra, *Feature subset selection by bayesian network-based optimization*. *Artificial Intelligence*, vol.123(1-2). pp.157 -184, 2000.
69. I. Guyon, S.G., M. Nikravesh, and L. Zadeh, *Feature Extraction: Foundations and Applications (Studies in Fuzziness and Soft Computing)*. Springer-Verlag New York, Inc. Secaucus, NJ, USA, 2006.
70. Y. Saeys, I.I., and P. Larranaga, *A review of feature selection techniques in bioinformatics*. *Bioinformatics*, vol. 23, no. 19, 2007.
71. L. Breiman, J.F., R. Olshen, and C. Stone., *Classification and Regression Trees*. Wadsworth Inc., 1984.
72. H. Fu, Z.X., E. Dellandrea, W. Dou, and L. Chen, *Image categorization using esfs: A new embedded feature selection method based on sfs*. In J. Blanc-Talon, W. Philips, D. Popescu, and P. Scheunders, editors, *Advanced Concepts for Intelligent Vision Systems*. Vol 5807. *Lecture Notes in Computer Science*, pp. 288 -299., 2009.
73. Romanski, P.a.K., L., *Package 'FSelector'*. <https://cran.r-project.org/web/packages/FSelector/FSelector.pdf>, 2016.

74. G. Casella, S.F.a.I.O., *An Introduction to Statistical Learning with Applications in R*. Springer New York Heidelberg Dordrecht London. DOI 10.1007/978-1-4614-7138-7, 2013.
75. Verikas, A., Gelzinis, A. and Bacauskiene, M., *Mining data with random forests: A survey and results of new tests*. *Pattern Recognition*, vol. 44(2), pp. 330–349. 2011.
76. Hjerpe, A., *Computing Random Forests Variable Importance Measures (VIM) on Mixed Numerical and Categorical Data*. 2016.
77. Díaz-Urarte, R.a.A., S., *Gene selection and classification of microarray data using random forest*. *BMC bioinformatics*, vol. 7(1). 2006.
78. Bradley, A.P., *The use of the area under the ROC curve in the evaluation of machine learning algorithms*. *Pattern recognition*, vol. 30(7), pp. 1145-1159. 1997.
79. Packman, K.A., *Signal detection theory: Valuable tools for evaluating inductive learning*. *Proceedings of the Sixth International Workshop on Machine Learning*. San Mateo, CA: Morgan Kaufman, pp. 160–163., 1989.
80. Och, F.J., *Minimum error rate training in statistical machine translation*. *ACL '03 Proceedings of the 41st Annual Meeting on Association for Computational Linguistics*. Vol. 1. pp. 160-167, 2003.
81. Kohavi, R., *A study of cross-validation and bootstrap for accuracy estimation and model selection*. *Ijcai*. Vol. 14. No. 2. pp. 1137-1145, 1995.
82. Refaeilzadeh, P., Tang, L., & Liu, H., *Cross-validation*. In *Encyclopedia of database systems*, pp. 532-538. Springer US., 2009.
83. Altrabsheh, N., Cocea, M., & Fallahkhair, S., *Learning Sentiment from Students' Feedback for Real-Time Interventions in Classrooms*. *Adaptive and Intelligent Systems*. Volume 8779 of the series *Lecture Notes in Computer Science*, 40-49. doi: http://dx.doi.org/10.1007/978-3-319-11298-5_5, 2014.
84. Ortigosa, A., Ortigosa, A. & Carro, R., *SentBuk: Sentiment analysis for e-learning environments*. *Computers in Education (SIIE)*, 2012 International Symposium on, 1-6., 2012.
85. Quinlan, J.R., *Induction of decision trees*. *Machine learning*, 1(1), 81-106. 1986.
86. F., C., M., de Santo and L., Greco, *SAFE: A Sentiment Analysis Framework for E-Learning*. *International Journal of Emerging Technologies in Learning*, Vol. 9, Issue 6, pp. 37, 2014.
87. Kirubakaran, A.J.a.E., *M-Learning Sentiment Analysis with Data Mining Techniques*. *International Journal of Computer Science and Telecommunications*. Trichy-India., 2012.
88. A. Gupte, S.J., P. Gadgul and A. Kadam, *Comparative Study of Classification Algorithms used in Sentiment Analysis* *International Journal of Computer Science and Information Technologies*, Vol. 5 (5) , 2014, 6261-6264, 2014.
89. X. Fang and J., Z., *Sentiment analysis using product review data*. *Department of Computer Science, North Carolina A&T State University, Greensboro, NC, USA.*, 2015.

90. W., L.a.J., Zhang, *The multilayer sentiment analysis model based on Random Forest*. School of Automation Beijing University of Posts and Telecommunications Beijing, China. AMEII 2016, 2016.
91. Aramaki, M.a.M., *Twitter catches the flu: Detecting influenza epidemics using Twitter*. Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, pagues 1568-1576, Edinburgh, Scotland, UK., 2011.
92. Vilares, J., *Aplicaciones del procesamiento de lenguaje natural en la recuperación de información*. Tesis de Doctorado. Universidad de A Coruña. (<http://goo.gl/q07ns1>), 2005.
93. Riloff, E., & Wiebe, J., *Learning extraction patterns for subjective expressions*. Conference on Empirical Methods in Natural Language Processing, 105-112. (<https://goo.gl/se0alg>), 2003.
94. Jurka, T., *Sentiment: Tools for Sentiment Analysis*. R package version 0.1 (<http://goo.gl/oxASCV>), 2012.
95. R-Core-Team, *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. (<http://goo.gl/e40yiU>), 2013.
96. Bordigo, F.a.P., W., *Procesamiento de Variables Morfológicas en Búsquedas de Textos en Castellano*. Rev. Interam. Bibliot. Medellín (Colombia) Vol. 24 No. 1, pp. 69 - 88, 2001.
97. Mejova, Y., *Sentiment Analysis: An Overview. Comprehensive Exam Paper*. (<https://goo.gl/xsFTV9>). 2009.
98. Prasad, S., *Micro-blogging Sentiment Analysis Using Bayesian Classification Methods*. CS224N Project Report, Stanford. (<http://goo.gl/W2koQT>), 2010.
99. Zhang, Y., Jin, R., and Zhou, Z. H., *Understanding bag-of-words model: a statistical framework*. *International Journal of Machine Learning and Cybernetics*, 1(1-4), pp. 43-52. 2010.
100. Man, Y., *Feature Extension for Short Text Categorization Using Frequent Terms*. *Procedia Computer Science* 31, pp. 663 – 670 2014.
101. C.C. Hsu, C.W.C.a.C.J.L., *A practical guide to support vector classification*. Technical report, Department of Computer Science, National Taiwan University, 2003.
102. Mladenic, D., & Grobelnik, M., *Word sequences as features in text-learning*. In Proceedings of the 17th Electrotechnical and Computer Science Conference., 1998.
103. Bartosz, M., *Robustness of Random Forest-based gene selection methods*. *BMC Bioinformatics*. DOI: 10.1186/1471-2105-15-8, 2014.
104. Biau, G.a.S., E., *A Random Forest Guided Tour*. *TEST*. Vol. 25, Issue 2 , pp. 197 - 227 2015.
105. Benitez, R., Escudero, G., Kanaan, S., & Rodó, D. M., *Inteligencia artificial avanzada*. Editorial UOC. 2014.

106. Fernández-Degado, M., Cernadas, E. and Barro, S., *Do we Need Hundreds of Classifiers to Solve Real World Classification Problems?* Journal of Machine Learning Research. Vol. 15, pp. 3133-3181, 2014.
107. learn, S. *Choosing the right estimator.* 2014.
108. LeCun, Y., Bengio, Y., and Hinton, G., *Deep learning.* Nature, vol. 521(7553), pp. 436-444. 2015.
109. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Byers, A. H. , *Big data: The next frontier for innovation, competition, and productivity.* 2011.
110. Socher, R., Perelygin, A., Wu, J., Chuang, C., Ng, A. & Potts, C., *Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank.* Conference on Empirical Methods in Natural Language Processing (EMNLP 2013), 2013.
111. Carrillo de Albornoz, J., *Un Modelo Lingüístico-Semántico Basado en Emociones para la Clasificación de Textos según su Polaridad e Intensidad.* Tesis doctoral. Departamento de Ingeniería del Software e Inteligencia Artificial. Facultad de Informática. Universidad Complutense de Madrid, 2011.
112. Margain, L., Gutiérrez, G., García, B., Méndez, A., Ochoa, A., De Luna, A. & Hernández G., *Ubiquitous Computing to Lower Domestic Violence Rate Based on Emotional Support Social Network (Redsiam).* Distributed Computing and Artificial Intelligence. Vol 217 of the series Advances in Intelligent Systems and Computing pp. 163-170, 2013.

Modelo de Análisis de Emociones para evaluar el desempeño docente aplicando algoritmos de Aprendizaje Automático

ORIGINALITY REPORT

10%

SIMILARITY INDEX

PRIMARY SOURCES

1	revistas.uned.es Internet	1154 words — 6%
2	repository.unab.edu.co Internet	245 words — 1%
3	ijcopi.org Internet	107 words — 1%
4	repositorio.uancv.edu.pe Internet	96 words — < 1%
5	docplayer.es Internet	62 words — < 1%
6	hdl.handle.net Internet	54 words — < 1%
7	link.springer.com Internet	47 words — < 1%
8	www.slideshare.net Internet	40 words — < 1%
9	fdocuments.ec Internet	30 words — < 1%

10	aprenderly.com Internet	29 words — < 1%
11	libraetd.lib.virginia.edu Internet	27 words — < 1%
12	www.researchgate.net Internet	26 words — < 1%
13	www.ijcert.org Internet	22 words — < 1%
14	bdigital.unal.edu.co Internet	19 words — < 1%
15	www.acarindex.com Internet	19 words — < 1%
16	repositorio.umsa.bo Internet	18 words — < 1%
17	www.springerprofessional.de Internet	18 words — < 1%
18	www.clouds-news.com Internet	17 words — < 1%
19	Anders Mejlbjerg. "Research-based knowledge to corporate businesses: the Danish experience", Biblioteca Universitaria, 2014 Crossref	15 words — < 1%

EXCLUDE BIBLIOGRAPHY ON

EXCLUDE MATCHES

< 15 WORDS