



**FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA DE SISTEMAS
BOGOTÁ D.C.**

LICENCIA CREATIVE COMMONS: Atribución-No-Comercial 2.5 Colombia (CC BY-NC 2.5).

AÑO DE ELABORACIÓN: 2020.

TÍTULO: Implementación de un modelo de análisis de sentimientos con respecto a la JEP basado en minería de datos en twitter.

AUTOR (ES): Erika Paez, Felipe Monroy

DIRECTOR(ES)/ASESOR(ES): Roger Guzman.

MODALIDAD: Trabajo de investigación tecnológica.

PÁGINAS:	98	TABLAS:	12	CUADROS:	0	FIGURAS:	35	ANEXOS:	4
-----------------	-----------	----------------	-----------	-----------------	----------	-----------------	-----------	----------------	----------

CONTENIDO:

INTRODUCCIÓN

1. GENERALIDADES

1.1. LÍNEA DE INVESTIGACIÓN

1.2 PLANTEAMIENTO DEL PROBLEMA

1.2.1 Descripción del problema

1.2.2 Formulación del problema

1.3 OBJETIVOS

1.3.1 Objetivo general

1.3.2 Objetivos específicos

1.4 JUSTIFICACIÓN

1.5 Delimitaciones



- 1.5.1 Limitaciones
- 1.5.2 Alcances
- 2 MARCO REFERENCIAL
 - 2.2 MARCO TEÓRICO
 - 2.2.1 Minería de datos predictiva
 - 2.2.2 Regresión
 - 2.2.3 Métodos bayesianos
 - 2.2.4 Discriminante
 - 2.2.5 Árboles de decisión
 - 2.2.6 Redes neuronales
 - 2.2.7 Minería de datos de descriptiva
 - 2.2.8 Clustering
 - 2.2.9 Segmentación
 - 2.2.10 Asociación
 - 2.2.11 Análisis exploratorio
 - 2.2.12 Análisis discriminante lineal
 - 2.2.13 Análisis discriminante cuadrático
 - 2.3 MARCO CONCEPTUAL
 - 2.3.1 Minería de datos
 - 2.3.2 Minería de texto
 - 2.3.3 Análisis de sentimientos
 - 2.3.4 Conjunto de datos
 - 2.3.5 Modelos de predicción
 - 2.3.6 Modelo supervisado
 - 2.3.7 Jurisdicción especial para la paz (JEP)
 - 2.4 ESTADO DEL ARTE



3 METODOLOGÍA

4 DISEÑO METODOLÓGICO

Conjunto de datos.

Procesamiento de lenguaje natural.

Extracción de características

Muestreo.

Entrenamiento:

Evaluación del desempeño:

4.2 Instalaciones y equipo requerido

4.3 Estrategias de comunicación y divulgación

5 RESULTADOS

6 DISCUSIÓN DE RESULTADOS

7 CONCLUSIONES

8 RECOMENDACIONES

9 ANEXOS

10 BIBLIOGRAFÍA

DESCRIPCIÓN: Este documento presenta un experimento de métodos basados en aprendizaje de máquina y minería de datos con el fin que mediante el tema de la Jurisdicción Especial para la Paz (JEP), implementar tres algoritmos los cuales fueron seleccionados mediante el estado del arte, en los que se realizó una comparación con las características de unigramas y bigramas de esta forma generando un modelo de análisis de sentimientos de las personas (positivo, negativo y neutro) y la opinión en este tema político, donde según estadísticas, las personas no se encuentran de acuerdo con los rubros asignados y los gastos que la JEP ha llegado a generar. Por parte de los distintos partidos políticos se han implementado diferentes acuerdos para una paz duradera, pero las personas mediante la red social Twitter han llegado a manifestar sus distintas opiniones.



METODOLOGÍA: Para realizar el proceso de análisis de sentimientos se implementaron seis etapas donde cada una dependerá de la anterior, para completar cada etapa se tiene que realizar una serie de tareas específicas, las cuales son extracción de datos, limpieza del conjunto de datos, procesamiento del lenguaje natural, muestreo, entrenamiento y clasificación.

Etapa 1 Conjunto de datos: En la primera etapa se recopiló la máxima cantidad de tweets, este conjunto de datos se obtiene desde el comienzo de la implementación de la metodología, a partir de la creación de una cuenta de twitter, la cual se debe convertir en una cuenta de desarrollador, con el fin de tener acceso al API de la aplicación. El acceso a esto es necesario para introducir las claves proporcionadas por el API de twitter en un aplicativo cuyo propósito es extraer tweets, ya que de esta manera se logró realizar la creación de un conjunto de datos. El siguiente paso es realizar el etiquetado de los tweets según su sentimiento (positivo, negativo o neutro).

Etapa 2 Procesamiento del lenguaje natural: En esta etapa se realiza el procesamiento inicial del texto, en lo que se ejecuta es la tokenización en la cual se divide la oración en palabras para facilitar la limpieza del texto, después se procede a eliminar las stopwords que son las palabras comunes y poco informativas del léxico, el siguiente paso es lematización y Stemming esto consiste en reducir cada palabra a su raíz, eliminando cualquier tipo de derivación. Con el fin de reducir el texto que se quiere procesar y utilizar solo lo más relevante, esta etapa es importante para obtener resultados más exactos.

Etapa 3 Extracción de características: Después de realizar el procesamiento de lenguaje natural, se procede a iniciar la extracción de características que consiste en implementar TF-IDF, el cual representa de manera numérica el texto del conjunto de datos. La extracción de características es convertida en un conjunto de datos, el cual se divide en dos partes, la primera es el conjunto de unigramas y el segundo de bigramas.

Etapa 4 Muestreo: Esta etapa se separan el conjunto de datos con el fin de obtener la información para el entrenamiento y el testeo de cada algoritmo en el cual se ejecutaron de manera diferente, para el caso de Máquina de Soporte Vectorial se le aplicara KFold, para el caso de Naive Bayes y Random Forest se dividirán los datos en porcentajes de la siguiente manera 70, 75 y 80% que corresponde al entrenamiento, por otra parte 30, 25 y 20% que es la prueba.



Etapa 5 Entrenamiento y clasificación: Se ejecuta el entrenamiento de los algoritmos de Máquina de Soporte Vectorial, Random Forest y Naive Bayes. Los cuales realizaron los pasos de entrenamiento y prueba, de acuerdo con los porcentajes mencionados.

Etapa 6 Desempeño: En esta etapa final en la Figura 23 de la sección F, Se evalúa el porcentaje del desempeño de los algoritmos, el cual se obtiene mediante las métricas de precisión, recall y F_1 – Score. Estas se lograron obtener con la matriz de confusión, basadas en la clasificación.

PALABRAS CLAVE: ANÁLISIS DE SENTIMIENTOS, JEP, MACHINE LEARNING, MÁQUINAS DE SOPORTE VECTORIAL, MINERÍA DE DATOS, NAIVE BAYES, RANDOM FOREST, TWITTER.

CONCLUSIONES:

- Se logró identificar los sentimientos de los usuarios mediante aprendizaje de máquina, implementando los algoritmos de Random Forest, Naive Bayes y Máquina de Soporte Vectorial midiendo el rendimiento de cada uno con las métricas de F_1 – Score, exactitud, recall y precisión.
- Con los resultados de la **¡Error! No se encuentra el origen de la referencia.**, donde se evidencia la cantidad de tweets y el tiempo de ejecución de las herramientas Python y R, se determinó que la mejor herramienta para la construcción del conjunto de datos es R.
- Se realizó un análisis frente a los experimentos que se han implementado, para la realización e implementación de la metodología, teniendo como base aprendizaje automático junto a los algoritmos de Máquina de Soporte vectorial, Naive Bayes y Random Forest, ejecutando una comparación entre los tres algoritmos para evidenciar cuál de estos es óptimo para la solución de este tipo de problemáticas, ya que son los más implementados.
- Se implementó un modelo de aprendizaje automático junto con minería de datos, en el cual se utilizó procesos de lenguaje natural, TF-IDF, KFold, kernel, particionamiento y matrices de confusión, para los diferentes



algoritmos, con los cuales se clasificaron los sentimientos de los usuarios para lograr analizar su posición frente a la JEP.

- Se evidencio que con la implementación del conjunto de datos que se obtuvo para la creación del proyecto, la cantidad de tweets positivos por parte de los usuarios fueron disminuyendo a través del tiempo.
- Se evaluó el rendimiento de las diferentes métricas, al momento de medir el desempeño de los clasificadores Naive Bayes, Random Forest y Máquina de Soporte Vectorial, se observa que el algoritmo de clasificación Random Forest con las características de unigramas obtuvo un mejor resultado, ya que este tiene una predicción mayor para el conjunto de datos que se encuentran desbalanceados en comparación con los otros clasificadores. De esta manera los resultados de las métricas en este algoritmo obtuvieron un valor de 68.10% en F_1 – Score, una precisión de 74.56%, un resultado óptimo en recall de 70.15% y una exactitud de 70.15%.

FUENTES:

jep.gov.co [en línea] ¿Que es la jurisdicción especial para la paz? <<https://www.jep.gov.co/Infografas/conozcalajep.pdf>>

elmundo.com [en línea] Acuerdo de paz: errores en serie <<https://www.elmundo.com/noticia/Acuerdo-de-pazerrores-en-serie/377466>>

eltiempo.com [en línea] El 11% de la población urbana en Colombia usa Twitter Disponible en internet <<https://www.eltiempo.com/archivo/documento/CMS-12406882>>.

eltiempo.com [en línea] Lo mejor del 2018 en Twitter: las cuentas y los hashtags más populares <<https://www.eltiempo.com/tecnosfera/novedades-tecnologia/las-cuentas-y-las-etiquetas-mas-populares-en-twitter-durante-2018-301996>>.

diarioinformacion.com [en línea] El éxito de Obama y la minería de datos <<https://www.diarioinformacion.com/opinion/2012/11/17/exito-obama-mineria-datos/1315856.html>>

**RESUMEN ANALÍTICO EN EDUCACIÓN
- RAE -**



UNIVERSIDAD CATÓLICA
de Colombia
Vigilada Mineducación

RIUCaC

bbc.com [en línea] Elecciones en Estados Unidos ¿Fue facebook la clave para el triunfo de Donald Trump? <<https://www.bbc.com/mundo/noticias-internacional-37946548>>

javeriana.edu.co [en línea]
<<https://repository.javeriana.edu.co/bitstream/handle/10554/20516/CaicedoOrtizLuisEduardo2016.pdf?sequence=1&isAllowed=y>>

elespectador.com [en línea] El 47% de los colombianos tiene una opinión favorable de la jep <<https://www.elespectador.com/noticias/politica/47-de-colombianos-tienen-una-opinion-favorable-de-la-jep-gallup-poll-articulo-861085>>

lafm.com [en línea] Las polémicas cuentas de la jep <<https://www.lafm.com.co/judicial/las-polemicas-cuentas-de-la-jep>>

portafolio.com [en línea] El plebiscito por la paz cuesta \$350.000 millones, ¿qué se puede hacer con ese mismo dinero? <<https://www.portafolio.co/tendencias/cuanto-cuesta-el-plebiscito-por-la-paz-499066>>

portafolio.com [en línea] El plebiscito por la paz cuesta \$350.000 millones, ¿qué se puede hacer con ese mismo dinero? <<https://www.portafolio.co/tendencias/cuanto-cuesta-el-plebiscito-por-la-paz-499066>>

mintic.gov.co [en línea] Colombia es uno de los países con más usuarios en redes sociales en la región <https://mintic.gov.co/portal/604/w3-article-2713.html?_noredirect=1>

bbc.com [en línea] Elecciones en Estados Unidos ¿Fue facebook la clave para el triunfo de Donald Trump? <<https://www.bbc.com/mundo/noticias-internacional-37946548>>

openaccess.uoc.edu [en línea] Análisis de sentimientos en twitter <<http://openaccess.uoc.edu/webapps/o2/bitstream/10609/81435/6/jsobrinosTFM0618memoria.pdf>>

repository.javeriana.edu.co [en línea] análisis del sentimiento político mediante la aplicación de herramientas de minería de datos a través del uso de redes

**RESUMEN ANALÍTICO EN EDUCACIÓN
- RAE -**



UNIVERSIDAD CATÓLICA
de Colombia
Vigilada Mineducación

RIUCaC

SOCIALES<<https://repository.javeriana.edu.co/bitstream/handle/10554/20516/CaicedoOrtizLuisEduardo2016.pdf?sequence=1&isAllowed=y>>

las2orillas.co/ [en línea] Aumenta la polarización política tras la decisión de la JEP sobre Santrich <<https://www.las2orillas.co/polarizacion-politica-decision-jep/>>

elespectador.com [en línea] 47 % de colombianos tienen una opinión favorable de la JEP: Gallup Poll <<https://www.elespectador.com/noticias/politica/47-de-colombianos-tienen-una-opinion-favorable-de-la-jep-gallup-poll-articulo-861085>>

bigdata-social.com [en línea] Análisis predictivo<<http://www.bigdata-social.com/que-es-el-analisis-predictivo/>>

Christopher M. Bishop. Pattern recognition and machine learning: Linear models for regression En: Information Science and Statics. Pages 45 – 46.

docs.microsoft.com [en línea] Algoritmos de minería de datos<<https://docs.microsoft.com/es-es/analysis-services/data-mining/data-mining-algorithms-analysis-services-data-mining?view=sql-server-2017>>

docs.microsoft.com [en línea] Naive Bayes Algorithm <<https://docs.microsoft.com/es-es/analysis-services/data-mining/microsoft-naive-bayes-algorithm?view=sql-server-2017>>

Christopher M. Bishop. Pattern recognition and machine learning: Linear models for regression En: Information Science and Statics. Contents 3. Pages 149 – 150.

dataprix.com [en línea] análisis discriminante < <https://www.dataprix.com/blog-it/mineria-datos/data-mining-analisis-discriminante-caso-sas>>

Christopher M. Bishop. Pattern recognition and machine learning: Linear models for regression En: Information Science and Statics. Contents 2. Pages 90 – 97.

docs.microsoft.com [en línea] Árboles de decisión <<https://docs.microsoft.com/es-es/analysis-services/data-mining/microsoft-decision-trees-algorithm?view=sql-server-2017>>

**RESUMEN ANALÍTICO EN EDUCACIÓN
- RAE -**



UNIVERSIDAD CATÓLICA
de Colombia
Vigilada Mineducación

RIUCaC

towarddatascience.com [en línea] Árboles de decisión <
<https://towardsdatascience.com/decision-trees-in-machine-learning-641b9c4e8052>>

docs.microsoft.com [en línea] Neural network algorithm
<<https://docs.microsoft.com/es-es/analysis-services/data-mining/microsoft-neural-network-algorithm?view=sql-server-2017>>

Christopher M. Bishop. Pattern recognition and machine learning: Linear models for regression En: Information Science and Statistics. Contents 2. Pages 110.

Tamps.cinvestav.mx [en línea] Minería de datos descriptiva
<<https://www.tamps.cinvestav.mx/~hmarin/Mineria/EC2.pdf>>

Techdifferences.com [en línea] Minería de datos descriptiva
<<https://techdifferences.com/difference-between-descriptive-and-predictive-data-mining.html>>

cs.us.es [en línea] técnicas de clustering <
https://www.cs.us.es/~fran/curso_unia/clustering.html>

es.coursera.org [en línea] que es clustering <<https://es.coursera.org/lecture/mineria-de-datos-introduccion/que-es-clustering-TMSYv>>

educba.com [en línea] Clustering < <https://www.educba.com/what-is-clustering-in-data-mining/>>

docs.microsoft.com [en línea] Algoritmos de minería de datos
<<https://docs.microsoft.com/es-es/analysis-services/data-mining/data-mining-algorithms-analysis-services-data-mining?view=sql-server-2017>>

ibm.com [en línea] Segmentation
<https://www.ibm.com/support/knowledgecenter/en/SSEPGG_9.7.0/com.ibm.datatools.datamining.doc/miningplan_custseg.html>

stateofdigital.com [en línea] Segmentation < <https://www.stateofdigital.com/one-size-does-not-fit-all-data-segmentation/>>

**RESUMEN ANALÍTICO EN EDUCACIÓN
- RAE -**



UNIVERSIDAD CATÓLICA
de Colombia
Vigilada Mineducación

RIUCaC

docs.microsoft.com [en línea] Algoritmos de minería de datos
<<https://docs.microsoft.com/es-es/analysis-services/data-mining/data-mining-algorithms-analysis-services-data-mining?view=sql-server-2017>>

docs.oracle.com [en línea] association <
https://docs.oracle.com/cd/B28359_01/datamine.111/b28129/market_basket.htm#DMCON009>

urg.es [en línea] analisis exploratorio <
<https://www.ugr.es/~batanero/pages/ARTICULOS/anaexplora.pdf> >

stat.cmu.edu [en línea] análisis exploratorio <
<https://www.stat.cmu.edu/~hseltman/309/Book/chapter4.pdf>>

Geeksforgeeks.org [en línea] análisis discriminante lineal <
<https://www.geeksforgeeks.org/ml-linear-discriminant-analysis/>>

Machinelearningmastery.com [en línea] análisis discriminante lineal
<<https://machinelearningmastery.com/linear-discriminant-analysis-for-machine-learning/>>

Uc-r.github.io [en línea] discriminant analysis http://uc-r.github.io/discriminant_analysis

Datascienceblog.net [en línea] linear and quadratic discriminant analysis
<https://www.datascienceblog.net/post/machine-learning/linear-discriminant-analysis/>

docs.microsoft.com [en línea] Conceptos de minería de datos
<<https://docs.microsoft.com/es-es/analysis-services/data-mining/data-mining-concepts?view=sql-server-2017>>

docs.oracle.com [en línea] Conceptos de minería de datos <
https://docs.oracle.com/cd/B28359_01/datamine.111/b28129/process.htm#DMCON046 >

ibm.com [en línea] Conceptos de minería de texto <
https://www.ibm.com/support/knowledgecenter/es/SS3RA7_18.1.1/ta_guide_ddita/textmining/shared_entities/tm_intro_tm_defined.html >

**RESUMEN ANALÍTICO EN EDUCACIÓN
- RAE -**



UNIVERSIDAD CATÓLICA
de Colombia
Vigilada Mineducación

RIUCaC

Docs.oracle.com [en línea] About text mining <
<https://docs.oracle.com/database/121/DMPRG/GUID-3E60BDD1-DE22-494F-8B6D-C73A03EDD01B.htm#DMPRG778> >

Docs.oracle.com [en línea] Text mining
<https://docs.oracle.com/cd/B28359_01/datamine.111/b28129/text.htm#BCEDHEDD >

Christopher M. Bishop. Pattern recognition and machine learning: Linear models for regression En: Information Science and Statics. Contents 5. Pages 225 - 236.

arimetrics.com [en línea] Análisis de sentimientos
<<https://www.arimetrics.com/glosario-digital/analisis-de-sentimiento>>

towardsdatascience.com [en línea] Sentiment analysis: concept, analysis and applications <
<https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17>>

lexalytics.com [en línea] Sentiment analysis explained
<<https://www.lexalytics.com/technology/sentiment-analysis>>

cleverdata.io [en línea] conjunto de datos <<https://cleverdata.io/conceptos-basicos-machine-learning/>>

ibm.com [en línea] what is a data set? <
https://www.ibm.com/support/knowledgecenter/zosbasics/com.ibm.zos.zconcepts/zconc_datasetintro.htm>

docs.microsoft.com [en línea] Modelo de minería de datos
<https://docs.microsoft.com/es-es/analysis-services/data-mining/mining-models-analysis-services-data-mining?view=sql-server-2017#bkmk_mdIDefine>

docs.oracle.com [en línea] Predictive analysis
<https://docs.oracle.com/cd/E28280_01/admin.1111/e14568/predict.htm#AAMAD5159>

cleardata.io [en línea] Conceptos básicos de Machine Learning
<<https://cleverdata.io/conceptos-basicos-machine-learning/>>



Docs.oracle.com [en línea] supervised data mining <
https://docs.oracle.com/cd/B19306_01/datamine.102/b14339/3predictive.htm#i1005885>

jep.gov.co [en línea] Jurisdicción especial para la paz
<<https://www.jep.gov.co/Paginas/JEP/Jurisdiccion-Especial-para-la-Paz.aspx> >

Ema Kušena, Mark Strembeck [año de publicación] 2018 Politics, sentiments, and misinformation: An analysis of the Twitter discussion on the 2016 Austrian Presidential Elections. Disponible en <
<https://www-sciencedirect-com.ucatolica.basesdedatosezproxy.com/science/article/pii/S2468696417301088>>

Efthymios Kouloumpis, Theresa Wilson, Johanna Moore [año de publicación] 2010 Twitter Sentiment Analysis: The Good the Bad and the OMG!. Disponible en <
<https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2857/3251>>

Trevor Hastie, Robert Tibshirani, Jerome Friedman. The elements of statistical learning, The second edition. En: chapter 12. Page 417 – 418.

Trevor Hastie, Robert Tibshirani, Jerome Friedman. The elements of statistical learning, the second edition. En: chapter 15. Page 587– 589.

Trevor Hastie, Robert Tibshirani, Jerome Friedman. The elements of statistical learning, the second edition. En: chapter 6. Page 210 – 211

Trevor Hastie, Robert Tibshirani, Jerome Friedman. The elements of statistical learning, the second edition. En: chapter 6. Page 191 – 200.

Eric S.Tellez, Sabino Miranda-Jiménez, Mario Graff, Daniela Moctezuma, Oscar S.Siordia, Elio A.Villaseñor [año de publicación] 2017 A case study of Spanish text transformations for twitter sentiment analysis. Disponible en <
<https://www-sciencedirect-com.ucatolica.basesdedatosezproxy.com/science/article/pii/S0957417417302312>>

Marcela Mayumi Mauricio Yagui, Luís Fernando Monsore Passos Maia, Jonice Oliveira, Adriana S. Vivacqua [año de publicación] 2018 Data mining of social manifestations in Twitter: Analysis and aspects of the social movement "Bela, recatada e do lar" (Beautiful, demure and housewife) Disponible en <
<http://web.a.ebscohost.com.ucatolica.basesdedatosezproxy.com/ehost/pdfviewer/p>

**RESUMEN ANALÍTICO EN EDUCACIÓN
- RAE -**



UNIVERSIDAD CATÓLICA
de Colombia
Vigilada Mineducación

RIUCaC

dfviewer?vid=1&sid=85eafd29-f950-4be5-94fd-ac673c3bef37%40sessionmgr4008
>

Ankita, Nabizath Saleenaa [año de publicación] 2018 An Ensemble Classification System for Twitter Sentiment Analysis. Disponible en < [https://www.sciencedirect-com.ucatolica.basesdedatosezproxy.com/science/article/pii/S187705091830841X](https://www.sciencedirect.com.ucatolica.basesdedatosezproxy.com/science/article/pii/S187705091830841X)>

Samah Mansour [año de publicación] 2018 Social Media Analysis of User's Responses to Terrorism Using Sentiment Analysis and Text Mining. Disponible < <https://www.sciencedirect-com.ucatolica.basesdedatosezproxy.com/science/article/pii/S1877050918319707> >

Carlos Arcila-Calderón, Félix Ortega-Mohedano, Javier Jiménez-Amores y Sofía Trullenque [año de publicación] 2017 Supervised sentiment analysis of political messages in spanish: Real-Time of tweets based on machine learning. Disponible en < <http://www.elprofesionaldelainformacion.com/contenidos/2017/sep/18.pdf> >

Lina andre torres samboni [año de publicación] 2015 análisis de sentimientos sobre el posconflicto colombiano utilizando herramientas de minería de text. Disponible en < <https://repositorio.escuelaing.edu.co/bitstream/001/403/1/Torres%20Samboni%2C%20Lina%20Andrea%20-%202016.pdf> >

Towardsdatascience.com [en línea] Natural language processing feature <<https://towardsdatascience.com/natural-language-processing-feature-engineering-using-tf-idf-e8b9d00e7e76>>

web.stanford.edu [en línea] The elements of statistical learning <<https://web.stanford.edu/~hastie/Papers/ESLII.pdf>>

Monkeylearn.com [En línea] Support vector Machine <<https://monkeylearn.com/blog/introduction-to-support-vector-machines-svm/>>

Medium.com [En línea] SVM <<https://medium.com/machine-learning-101/chapter-2-svm-support-vector-machine-theory-f0812effc72>>

Towards data science [en línea] Naive Bayes < <https://towardsdatascience.com/naive-bayes-classifier-81d512f50a7c>>

**RESUMEN ANALÍTICO EN EDUCACIÓN
- RAE -**



UNIVERSIDAD CATÓLICA
de Colombia
Vigilada Mineducación

RIUCaC

scikit-learn.org [en línea] random forest <[https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.htm](https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html)
>

Towards data science [en línea] Naive bayes <<https://towardsdatascience.com/naive-bayes-classifier-81d512f50a7c>>

Towards data science [en línea] Cross validation <<https://towardsdatascience.com/why-and-how-to-cross-validate-a-model-d6424b45261f>>

Towardsdatascience.com [en línea] KFOLDS <<https://towardsdatascience.com/natural-language-processing-feature-engineering-using-tf-idf-e8b9d00e7e76>>

Towardsdatascience.com [en línea] NLP <<https://towardsdatascience.com/natural-language-processing-feature-engineering-using-tf-idf-e8b9d00e7e76>>

Towardsdatascience.com [en línea] NGRAM <<https://towardsdatascience.com/introduction-to-language-models-n-gram-e323081503d9>>

Knime.com [en línea] <<https://www.knime.com/about>>

Semanticscholar [en línea] <<https://pdfs.semanticscholar.org/1d10/6a2730801b6210a67f7622e4d192bb309303.pdf>>

Analytics Lane [en línea] <<https://www.analyticslane.com/2019/12/16/cual-es-la-diferencia-entre-parametro-e-hiperparametro/>>

Investopedia [en línea] <<https://www.investopedia.com/terms/w/weightedaverage.asp>>

Machine learning [en línea] <<https://developers.google.com/machine-learning/crash-course/classification/accuracy>>

**RESUMEN ANALÍTICO EN EDUCACIÓN
- RAE -**



UNIVERSIDAD CATÓLICA
de Colombia
Vigilada Mineducación

RIUCaC

Investopedia [en línea]
<<https://www.investopedia.com/terms/w/weightedaverage.asp>>

Confusion matrix [en línea] <<https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62>>

SVM with kernel [en línea] <<https://towardsdatascience.com/support-vector-machines-svm-c9ef22815589>>

LISTA DE ANEXOS:

Anexo A: Resultados con Knime

Anexo B: Conjunto de datos

Anexo C: Repositorio de código

Anexo D: Repositorio Knime