Trabajo Integrador de Minería de Texto

Análisis de Papers en Idioma Español



Contexto y Proceso de Extracción

Papers



03 temas de interés

- Redes Neuronales para predicción
- Clustering alumnos de una carrera Universitaria
- Minería de Opinión para
 Comentarios en TripAdvisor



Descargados previamente en formato PDF para la elaboración del proyecto de tesis

23 Documentos

8 RNA

7 Clustering

8 Minería de Opinión

Se renombro los documentos de acuerdo al titulo planteado en el documento, país de procedencia y año de publicación.

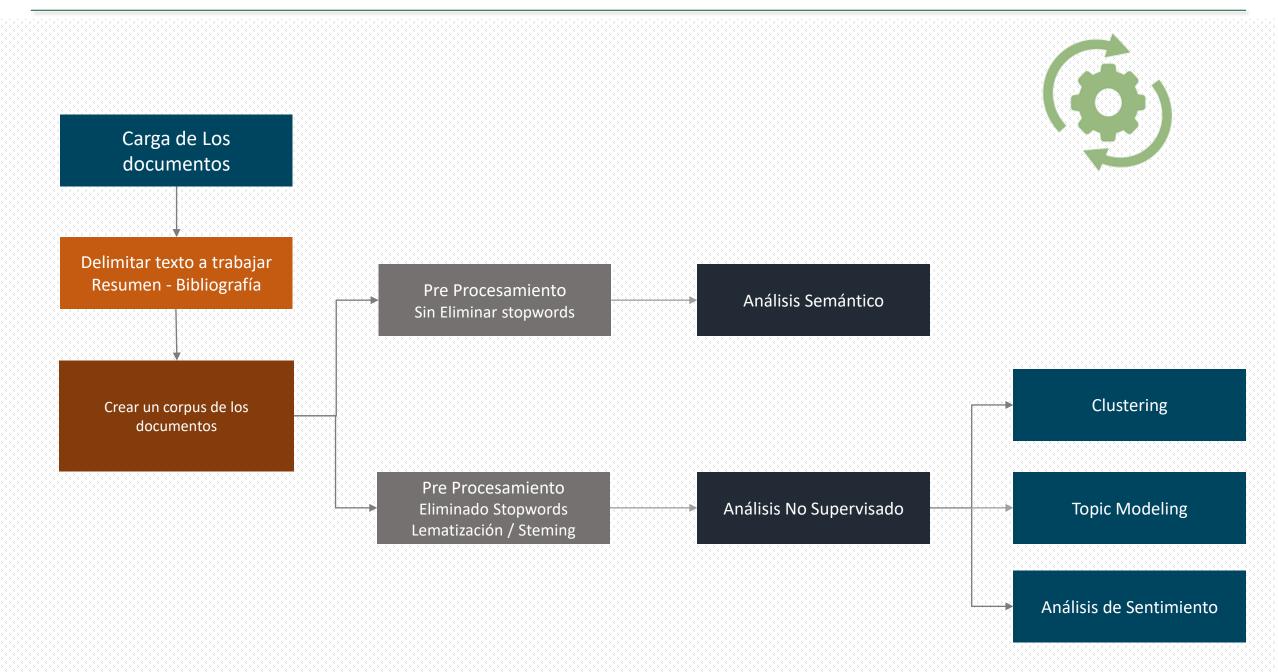




Librería Tika



Apache Tika es una biblioteca que se utiliza para la detección de tipos de documentos y la extracción de contenido de varios formatos de archivo.



Análisis Semántico

OBJETIVO DEL ANÁLISIS

Diversidad Léxica: la proporción de palabras diferentes del total de palabras de una composición.

Densidad Léxica:

relación de las palabras nocionales, es decir con contenido semántico (sustantivos, verbos, adjetivos, adverbios) con respecto al total de palabras.

PROCESO UTILIZADO

- 1 Limpieza de texto I
- Detección Tipo de Palabra Part Of Speach (POS)
- 3 Limpieza de texto II
- 4 Cálculo de Diversidad Léxica
- 5 Cálculo de Densidad Léxica

LIMPIEZA DE TEXTO

 Remover saltos de línea



- Remover Saltos de página
- Transformar a minúsculas
- Remover comillas
- Remover palabras con errores (1,2,3 letras)
- Remover signos de puntuación

Se creó un listado para exceptuar las palabras que si están correctamente escritas



Palabras Nocionales

Tipo	Total	Nro Únicas
Sustantitvos	25,911	3,150
Verbos	7,581	2,176
Adjetivos	9,459	2,471
Adverbios	2,547	261

SUSTANTIVOS

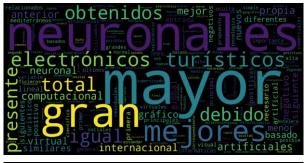




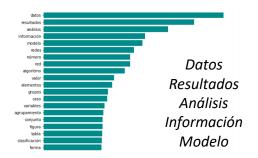


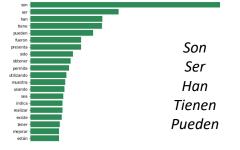
VERBOS

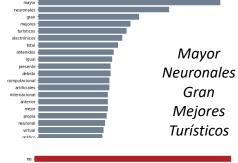
ADVERBIOS

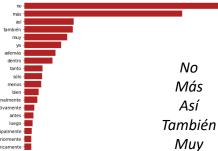












Análisis Semántico: Diversidad y Densidad Léxica

Cómo Corpus

Palabras 105,575 Totales:

Palabras

Únicas 11,415

Palabras

Nocionales:

45,498

Diversidad

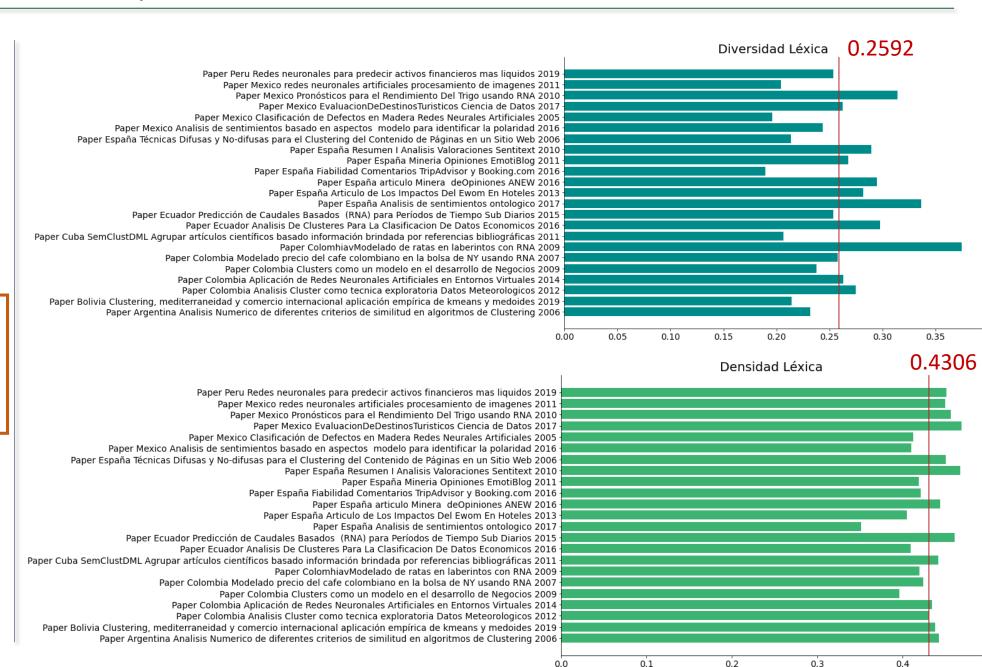
0.1081

Léxica:

Densidad 0.431

Léxica:

- Diversidad Léxica
 Promedio de Diversidad
 Léxica, los textos tienen
 una sintaxis muy similar
- El contenido semántico es similar entre los documentos



Análisis No Supervisado

OBJETIVO DEL ANÁLISIS

Clustering: Agrupar llos documentos con contenido similar.

Topic Modeling:

Identificar, los temas (tópicos) principales que encierra un texto.

Análisis de Sentimiento: Carga emocional que tienen los documentos

PROCESO UTILIZADO

- 1 Limpieza de texto
- 2 Lematización / Steming

Pre Procesa miento

- 3 Clustering
 - Matriz TF
 - Matriz TF-IDF
 - Matriz Word Embeding
- 4 Topic Modeling (LDA)
 - Librería Gensim
 - Librería Sklearn
- 5 Análisis de Sentimiento

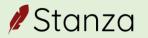
LIMPIEZA DE TEXTO

- Remover saltos de línea
- Remover Saltos de página
- Remover signos de Puntuación y comillas
- Transformar a minúsculas
- Remover Palabras que contengan números
- Remover Stopwords
- Remover palabras con errores (1,2,3 letras)

Se creó un listado para exceptuar las palabras que si están correctamente escritas

Pre Procesamiento

Lematización



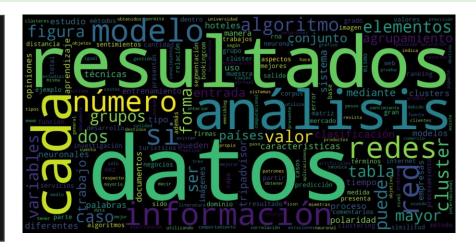
Steming:



Cómo Corpus

Tipo	Total	Nro Únicas
Total	113,057	18,365
Limpio	55,071	10,546
Lematizado	55,071	7,396
Stematizado	55,071	6,387





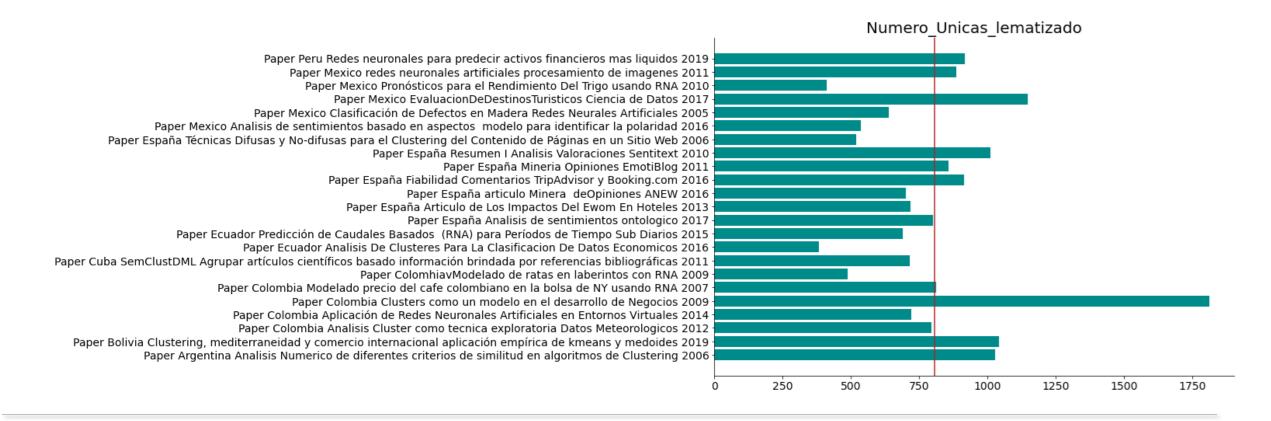
LEMATIZADO



Lematización por Documento

Palabras Únicas Lematizadas

Promedio: 806 Palabras Únicas Por Documento Desviación +/- 300 Palabras



Palabras Únicas Limpias Promedio: 2,394 Palabras Únicas Por Documento Desviación +/- 1,014 Palabras

Clustering

Con la Matriz TF

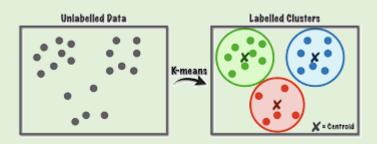


Determinación de la Matriz TF

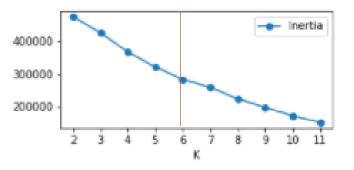
Matriz: CountVectorizer

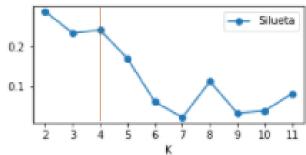
Escalar: MaxAbsScaler

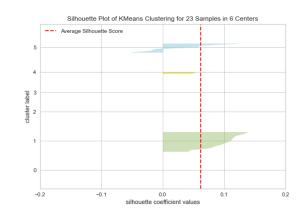
Por el Sparse (número de ceros)

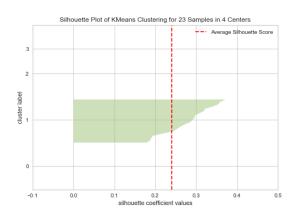


K: Óptimo

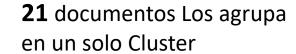








3 documentos no se pueden agrupar

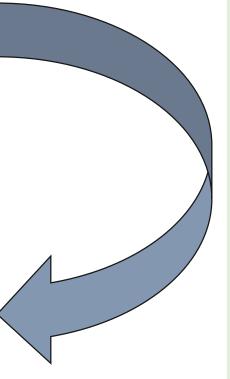


Consecuente a lo determinado en la diversidad léxica

FRECUENCIAS





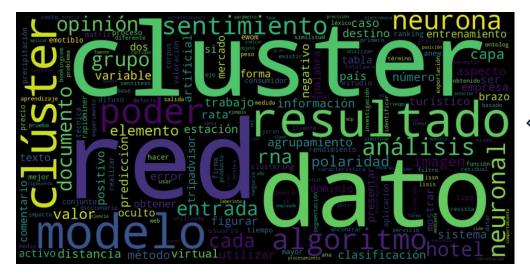


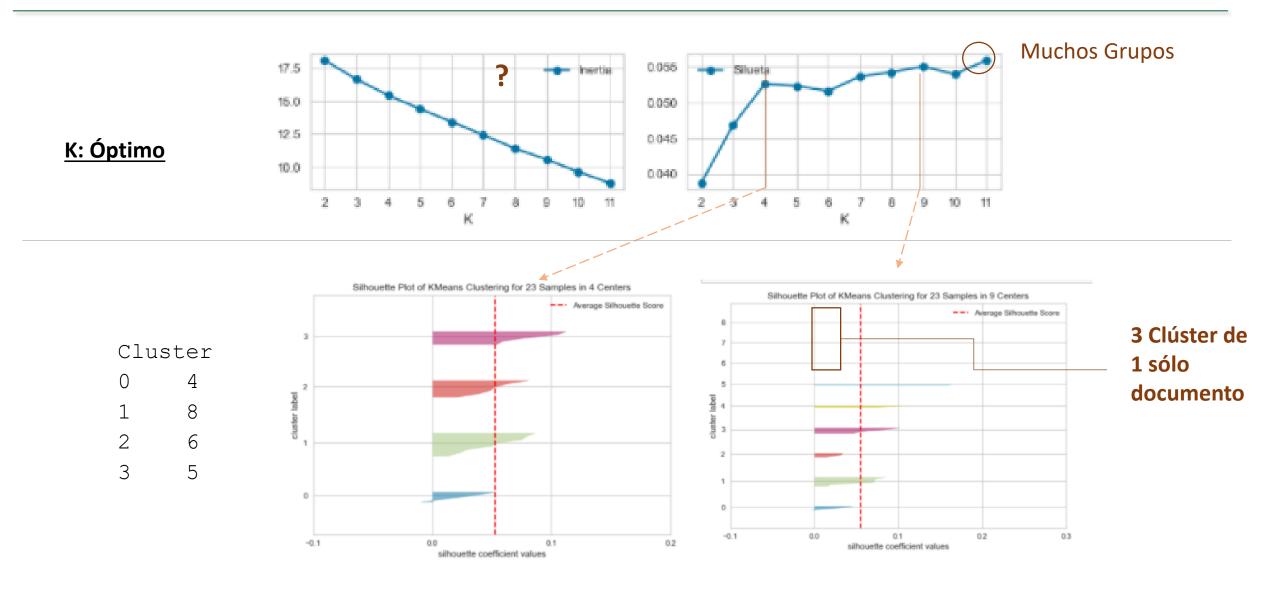


Determinación de la Matriz TF -IDF

Matriz: TfidfVectorizer

TF - IDF





Matriz con Word embedings



Creación de la matriz a partir del modelo Word2Vec

Paso1: Tokenizar (nltk)

Paso2: Definir y entrenar el

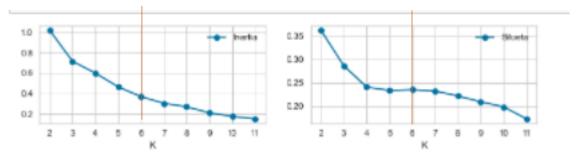
modelo

Min:5 apariciones, Windows= 35 máxima distancia entre la ocurrencia ya la palabra predicha, size = 50 dimensionalidad del vector

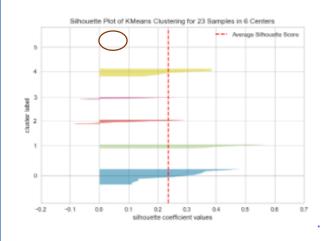
Paso3: Transformar el corpus usando el modelo Word2Vec

*Crear un array de vectores en base al modelo con las palabras que aparecen en el documento

K: Óptimo



K = 6



Titulos	
Paper España Articulo de Los Impact	os Del Ewom En Hoteles 2013
Paper Mexico EvaluacionDeDestinosTu	risticos Ciencia de Datos 2017
Paper España Resumen I Analis	is Valoraciones Sentitext 2010
Paper España artículo Mi	nera deOpiniones ANEW 2016
Paper España Analisis de	sentimientos ontologico 2017
Paper Ecuador Analisis De Clusteres Para La Clasificacio	n De Datos Economicos 2016
Paper Peru Redes neuronales para predecir activos	financieros mas liquidos 2019
Paper Colombia Clusters como un modelo en e	el desarrollo de Negocios 2009
Paper Bolivia Clustering, mediterraneidad y comercio internacional aplicación empíric	a de kmeans y medoides 2019
Paper Ecuador Predicción de Caudales Basados (RNA) para Período	s de Tiempo Sub Diarios 2015
Paper Mexico redes neuronales artificiales pro	cesamiento de imagenes 2011
Paper Mexico Pronósticos para el Rendimien	to Del Trigo usando RNA 2010
Paper Colombia Modelado precio del cafe colombiano en la b	olsa de NY usando RNA 2007
Paper España Técnicas Difusas y No-difusas para el Clustering del Contenido de	Páginas en un Sitio Web 2006
Paper Mexico Clasificación de Defectos en Madera Re	edes Neurales Artificiales 2005
Paper Colombia Aplicación de Redes Neuronales Artificia	es en Entomos Virtuales 2014
Paper ColomhiavModelado de rata	s en laberintos con RNA 2009
er Cuba SemClustDML Agrupar artículos científicos basado información brindada por	referencias bibliográficas 2011
Paper España Mine	eria Opiniones EmotiBlog 2011
Paper Colombia Analisis Cluster como tecnica explorato	ria Datos Meteorologicos 2012
Paper Mexico Analisis de sentimientos basado en aspectos modelo pa	ra identificar la polaridad 2016
Paper Argentina Analisis Numerico de diferentes criterios de similitud en	algoritmos de Clustering 2006
Paper España Fiabilidad Comentarios Tr	ipAdvisor y Booking.com 2016

Caracterización de los Grupos (k = 4, Técnica TF-IDF)

Comentarios Web + Negocios Internacionales

De Clustering

l	Titulo	Cluster
	Paper España Artículo de Los Impactos Del Ewom En Hoteles 2013	0
	Paper Mexico EvaluacionDeDestinosTuristicos Ciencia de Datos 201	0
	Paper Colombia Clusters como un modelo en el desarrollo de Negocios 2009	0
	Paper España Fiabilidad Comentarios TripAdvisor y Booking.com 2010	0
	Paper Mexico Pronósticos para el Rendimiento Del Trigo usando RNA 2010	1
	Paper Mexico Clasificación de Defectos en Madera Redes Neurales Artificiales 200	1
	Paper Mexico redes neuronales artificiales procesamiento de imagenes 201	1
	Paper Ecuador Predicción de Caudales Basados (RNA) para Períodos de Tiempo Sub Diarios 2015	1
	Paper Peru Redes neuronales para predecir activos financieros mas liquidos 2019	1
	Paper ColomhiavModelado de ratas en laberintos con RNA 2009	1
	Paper Colombia Modelado precio del cafe colombiano en la bolsa de NY usando RNA 200	1
	Paper Colombia Aplicación de Redes Neuronales Artificiales en Entornos Virtuales 2014	1
	Paper Cuba SemClustDML Agrupar artículos científicos basado información brindada por referencias bibliográficas 201	2
	Paper España Técnicas Difusas y No-difusas para el Clustering del Contenido de Páginas en un Sitio Web 200	2
	Paper Colombia Analisis Cluster como tecnica exploratoria Datos Meteorologicos 2013	2
	Paper Bolivia Clustering, mediterraneidad y comercio internacional aplicación empírica de kmeans y medoides 2019	2
	Paper Ecuador Analisis De Clusteres Para La Clasificacion De Datos Economicos 2010	2
	Paper Argentina Analisis Numerico de diferentes criterios de similitud en algoritmos de Clustering 200	2
	Paper España Analisis de sentimientos ontologico 201	3
	Paper España articulo Minera deOpiniones ANEW 2010	3
	Paper España Mineria Opiniones EmotiBlog 201	3
	Paper España Resumen I Analisis Valoraciones Sentitext 2010	3
	Paper Mexico Analisis de sentimientos basado en aspectos modelo para identificar la polaridad 2010	3

De Redes Neuronales

De Minería de Opinión

Caracterización de los Grupos (k = 4, Técnica TF-IDF)



Cluster 1

Medoides: destino, turístico, hotel, ewom, tripadvisor, firma, cluster, ranking, consumidor, empresa

Medoídes y Palabras más Frecuentes por Clúster



Cluster 2

Medoides: red, neuro modelo, neuronal, i capa, entrada, imag artificial



Cluster 3

Medoides: clúster, cluster, grupo, documento, agrupamiento, distancia, estación, dato, elemento, país



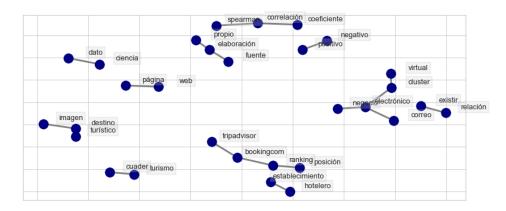
Cluster 4

Medoides: sentimient polaridad, opinió aspecto, domini texto, anew, emotible palabram corpus

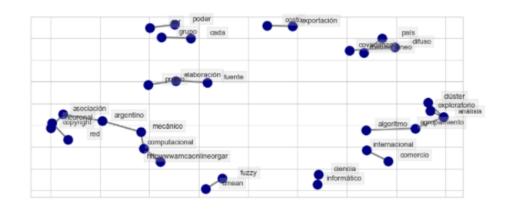
Caracterización de los Grupos (k = 4, Técnica TF-IDF)

Diagrama de Redes por Clúster

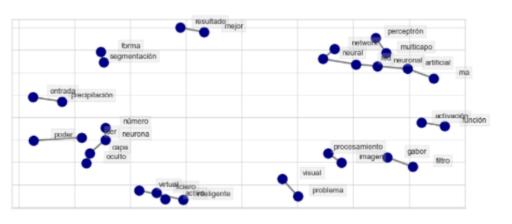
Clúster 1: Comentarios Web + Negocios Internacionales



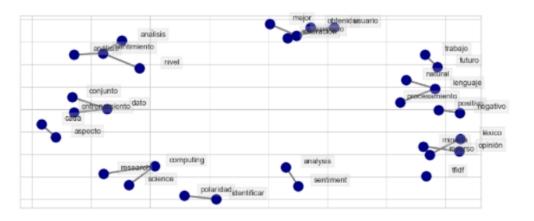
Clúster 3: Clustering



Clúster 2: Redes Neuronales



Clúster 3: Minería de Opinión



Topic Modeling

Aplicación de LDA para el topic modeling

Pasos Seguidos

- 1 Determinar el Nro. de Temas
- 2 Detección de los temas tratados

Determinación de Temas x Coherencia

Coherencia: Puntúa un solo tema midiendo el grado de similitud semántica entre las palabras de puntuación alta en el tema. Permite distinguir entre temas que son semánticamente interpretables y temas que son artefactos de inferencia estadística.

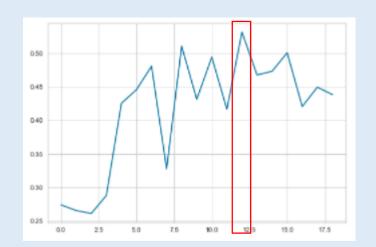
https://towardsdatascience.com/evaluate-topic-model-in-python-latent-dirichlet-allocation-lda-7d57484bb5d0

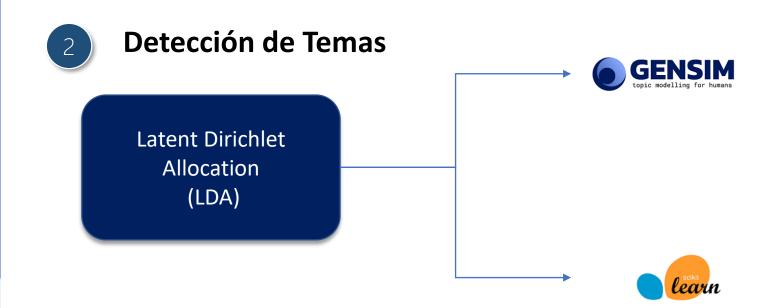


Nro. de Temas

Cálculo de la coherencia Evaluado: 02 a 20

Mayor Coherencia: 12 temas





Aplicación de LDA para el topic modeling

<u>Determinación de los Temas en el</u> <u>Corpus</u>



Topic 0:

documento cluster término difuso clustering

Topic 1:

cluster empresa destino turístico negocio

Topic 2:

tripadvisor ranking hotel posición red

Topic 3:

dato resultado poder cada trabajo

Topic 4:

documento resultado término número difuso

Topic 5:

rna virtual problema red dato

Topic 6:

evolutivo predicción productosservicios analizar utilización

Topic 7:

resultado valor aspecto cada elemento

Topic 8:

poder cluster destino valor clúster

Topic 9:

resultado red utilizar algoritmo mercado

Topic 10:

modelo red dato algoritmo poder

Topic 11:

opinión texto análisis palabra valoración

Asignación del principal Tópico a documento

Tema	Topico	Titulos	Cluster
[cluster empresa destino turístico negocio]	1	Paper España Articulo de Los Impactos Del Ewom En Hoteles 2013	0
[cluster empresa destino turístico negocio]	1	Paper Mexico EvaluacionDeDestinosTuristicos Ciencia de Datos 2017	0
[cluster empresa destino turístico negocio]	1	Paper Colombia Clusters como un modelo en el desarrollo de Negocios 2009	0
[tripadvisor ranking hotel posición red]	2	Paper España Fiabilidad Comentarios TripAdvisor y Booking.com 2016	0
[modelo red dato algoritmo poder]	10	Paper Mexico Pronósticos para el Rendimiento Del Trigo usando RNA 2010	1
[modelo red dato algoritmo poder	10	Paper Mexico Clasificación de Defectos en Madera Redes Neurales Artificiales 2005	1
[modelo red dato algoritmo poder	10	Paper Mexico redes neuronales artificiales procesamiento de imagenes 2011	1
[modelo red dato algoritmo poder	10	Paper Ecuador Predicción de Caudales Basados (RNA) para Períodos de Tiempo Sub Diarios 2015	1
[resultado valor aspecto cada elemento	7	Paper Peru Redes neuronales para predecir activos financieros mas líquidos 2019	1
[modelo red dato algoritmo poder	10	Paper ColomhiavModelado de ratas en laberintos con RNA 2009	1
[modelo red dato algoritmo poder	10	Paper Colombia Modelado precio del cafe colombiano en la bolsa de NY usando RNA 2007	1
[modelo red dato algoritmo poder	10	Paper Colombia Aplicación de Redes Neuronales Artificiales en Entomos Virtuales 2014	1
[resultado valor aspecto cada elemento	7	Paper Cuba SemClustDML Agrupar artículos científicos basado información brindada por referencias bibliográficas 2011	2
[documento cluster término difuso clustering	0	Paper España Técnicas Difusas y No-difusas para el Clustering del Contenido de Páginas en un Sitio Web 2006	2
[modelo red dato algoritmo poder	10	Paper Colombia Analisis Cluster como tecnica exploratoria Datos Meteorologicos 2012	2
[modelo red dato algoritmo poder	10	Paper Bolivia Clustering, mediterraneidad y comercio internacional aplicación emplrica de kmeans y medoides 2019	2
[modelo red dato algoritmo poder	10	Paper Ecuador Analisis De Clusteres Para La Clasificación De Datos Económicos 2016	2
[modelo red dato algoritmo poder	10	Paper Argentina Analisis Numerico de diferentes criterios de similitud en algoritmos de Clustering 2006	2
[resultado valor aspecto cada elemento	7	Paper España Analisis de sentimientos ontologico 2017	3
[opinión texto análisis palabra valoración	11	Paper España articulo Minera deOpiniones ANEW 2016	3
[resultado valor aspecto cada elemento	7	Paper España Mineria Opiniones EmotiBlog 2011	3
[opinión texto análisis palabra valoración]	11	Paper España Resumen I Analisis Valoraciones Sentitext 2010	3
[opinión texto análisis palabra valoración]	11	Paper Mexico Analisis de sentimientos basado en aspectos modelo para identificar la polaridad 2016	3

6 temas son los principales; El 10 es el tema más destacado en los documentos

Aplicación de LDA para el topic modeling

<u>Determinación de los Temas en el</u> <u>Corpus</u>



Topic 0:

'0.010*"red" + 0.010*"grupo" + 0.008*"cada" + 0.008*"resultado" + 0.008*"elemento"

Topic 1:

'0.012*"modelo" + 0.011*"tripadvisor" + 0.010*"ranking" + 0.007*"hotel" + 0.007*"posición"

Topic 2:

'0.024*"destino" + 0.022*"turístico" + 0.011*"dato" + 0.009*"servicio" + 0.009*"estudio"')

Topic 3:

'0.025*"país" + 0.012*"costo" + 0.012*"dato" + 0.011*"cluster" + 0.010*"algoritmo

Topic 4:

'0.001*"red" + 0.001*"resultado" + 0.000*"poder" + 0.000*"grupo" + 0.000*"cluster"

Topic 5:

'0.012*"opinión" + 0.011*"palabra" + 0.011*"texto" + 0.010*"valoración" + 0.010*"análisis"

Topic 6:

'0.020*"rata" + 0.017*"brazo" + 0.010*"laberinto" + 0.008*"modelo" + 0.008*"lce"

Topic 7:

'0.000*"resultado" + 0.000*"algoritmo" + 0.000*"poder" + 0.000*"cluster" + 0.000*"red"

Topic 8:

'0.015*"cluster" + 0.012*"empresa" + 0.009*"firma" + 0.009*"negocio" + 0.008*"electrónico

Topic 9:

'0.019*"documento" + 0.016*"aspecto" + 0.013*"resultado" + 0.011*"cluster" + 0.009*"término"

Topic 10:

'0.022*"entrada" + 0.018*"red" + 0.016*"algoritmo" + 0.013*"precipitación" + 0.012*"neuronal

Topic 11:

'0.020*"rna" + 0.014*"virtual" + 0.013*"ojo" + 0.011*"problema" + 0.010*"red"

Asignación del principal Tópico a documento

ler	Topico2	Intular	luster
0.015""cluster" + 0.012""empresa" + 0.008""firms" 0.009""negocio" + 0.008""electrónic	8	Paper España Articulo de Los Impactos Del Ewom En Hoteles 2013	0
0.024""dealino" + 0.022""lurfalico" + 0.011""delo" 0.009""servicio" + 0.009""estudi	2	Paper Mexico Evaluacion/DeDestinos/Turtaticos Ciencia de Datos 2017	0
0.015""duster" + 0.012""empresar" + 0.009""firma" 0.009""negods" + 0.008""electrónic	8	Paper Colombia Clusters como un modelo en el desarrollo de Negocios 2009	0
0.012"modelo" + 0.011"lnipadvisor" + 0.010"manking" 0.007"holef + 0.007"posició	1	Paper Exparis Fisblided Comentation TripAdvisor y Booking.com 2016	0
0.025""paix" + 0.012""costo" + 0.012""dato" + 0.011""ckuster 0.010""bigostim	3	Paper Mexico Pronústicos para el Rendimiento Del Trigo usando RNA 2010	1
0.010""red" + 0.010""grupo" + 0.008""cadw" + 0.008""resultado" 0.008""wiernanii	0	Paper Mexico Classificación de Defectos en Madera Hedes Neurales Artificiales 2005	- 1
0.010""red" + 0.010""grupo" + 0.008""cade" + 0.008""resultado" 0.008""element	0	Paper Micrico redes neuronales artificiales procesamiento de imagenes 2011	1
0.022""entrada" + 0.018"'red" + 0.018"'algoritmo" 0.013" precipitación" + 0.012" neurona	10	Paper Ecuador Predicción de Caudales Basados (RNA) para Períodos de Tiempo Sub Diarios 2015	- 1
0.012"modelo" + 0.011"lnipadvisor" + 0.010"mnrking" 0.00/"holef" + 0.00/"posició	1	Paper Peru Redea neuronales para predecir activos financieros mas liquidos 2019	- 1
0.020""nata" + 0.017""brazo" + 0.010""laberinto" + 0.008""model + 0.008"10	8	Paper ColombiavModelado de ratas en laberintos con HNA 2009	1
0.012"modelo" + 0.011"lnipadvisor" + 0.010"mnrking" 0.007"holef + 0.007"poseció	1	Paper Colombia Modelado precio del cafe colombiano en la bolza de NY usando RNA 2007	1
0.020"ma" + 0.014""virtual" + 0.013""ojo" + 0.011""problema" 0.010"ma	11	Paper Colombia Aplicación de Redex Neuronales Artificiales en Entomos Virtuales 2014	1
0.010""red" + 0.010""grupo" + 0.008""cadw" + 0.008""nesultado" 0.008""element	0	Paper Cuba SemClustDML Agruper ertículos científicos basado información brindada por referencias bibliográficas 2011	2
0.019""documento" + 0.016""aspecto" + 0.013"'resultado" 0.011" cluster" + 0.005" términ	9	Paper España Técnicas Difusas y No-difusas para el Clustering del Contenido de Páginas en un Sito Web 2008	2
0.010"red" + 0.010"'grupo" + 0.008"'cade" + 0.008"'resultado" 0.008"wiernanii	0	Paper Colombia Areibia Cluster como tecnica exploratoria Datos Meteorologicos 2012	2
0.025"'paix" + 0.012"'costo" + 0.012"'dato" + 0.011"'duster 0.010"'algoritm	3	Paper Bolivia Clustering, mediterransidad y comercio internacional aplicación empirica de kmeans y medoides 2019	2
0.015""duster" + 0.012""empress" + 0.009""firms" 0.009""negodo" + 0.008""electrónic	8	Paper Ecuador Areitois De Clusteres Pare La Classificación De Datos Economicos 2016	2
0.010"red" + 0.010"grupo" + 0.008"cade" + 0.008"resultado" 0.008"wiernenii	0	Paper Argentina Analisis Numerico de diferentes criterios de similitud en algoritmos de Clustering 2008	2
0.019"documento" + 0.016"aspecto" + 0.013"resultado" 0.011"duster" + 0.009"1érmin	9	Paper España Analists de sentimientos ontológico 2017	3
0.012**opinión* + 0.011**palatra* + 0.011**lexto* 0.010**veloración* + 0.010**ansitat	5	Paper España artículo Minera deOpiniones ANEW 2018	3
0.010"red" + 0.010"grupo" + 0.008"cade" + 0.008"resultado" 0.008"element	0	Paper España Minerta Opinionea EmoliBiog 2011	3
0.012**opinión* + 0.011**palatra* + 0.011**lanto* 0.010**valoración* + 0.010**ansiitai	5	Paper España Resurren I Analisis Valoraciones Sentiticol 2010	3
0.019"documento" + 0.016"aspecto" + 0.013"resultado" 0.011"cluster" + 0.005"3érmin	9	Paper Mexico Analisis de sentimientos basado en aspectos modelo para identificar la potenidad 2016	3

10 temas son los principales; El 04 es el tema más destacado en los documentos

Análisis de Sentimiento

Análisis de Sentimiento



Lexicón

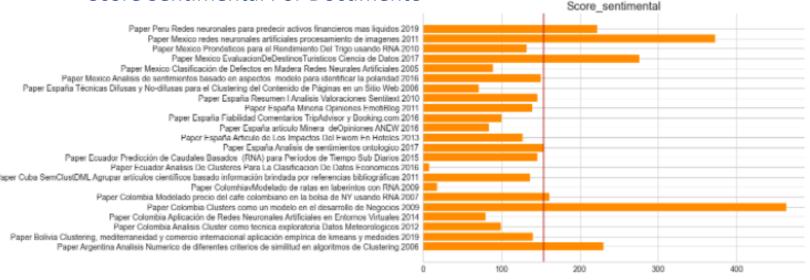
Desarrollado por la Universidad
Javeriana de Colombia

-1: Palabra Negativa; O Palabra Neutra; +1 Palabra Positiva

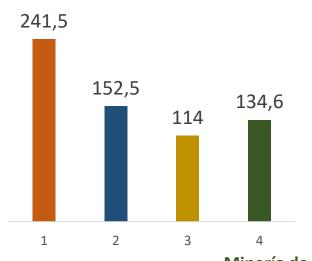


Score Sentimental

Score Sentimental Por Documento



Score Sentimental Por Clúster



Más Positivo : Clúster como un modelo en el desarrollo de Negocio

Menos Positivo: Análisis de Clúster para la clasificación de Datos Económicos (neutro)

Redes Neuronales

Minería de Opinión

Comentarios Web + Negocios Internacionales

Clustering

Conclusiones

Conclusiones

- 1. En el análisis semántico de los documentos se encontró que se tiene una densidad léxica de 0.431 que es similar al promedio por documento (0.4306), mientras que se tenía una diversidad léxica de 0.1081 que es mucho menor al promedio por documento (0.2592). Esto quiere decir que los documentos se parecen entre sí, lo cual es consecuente con que se tratan de temas de ciencia de datos y a su vez con la metodología para hacer un paper.
- En la agrupación de los documentos (clustering) se determinó que la mejor técnica para este corpus, de las utilizadas en la investigación, fue la aplicación de K-means a una matriz de TF-IDF. Obteniéndose 04 clústeres diferenciados con una silueta promedio de 0.053, y una silueta mínima de 0.02 en un grupo, los grupos los definimos como:
 - Clúster 01: Papers de Comentarios Web y Negocios Internacionales
 - Clúster 02: Papers de Redes Neuronales
 - Clúster 03: Papers de Clustering
 - Clúster 04: Papers de Minería de Opinión
- En el análisis de Topic Modeling se determinaron 12 temas en los 23 documentos aplicando el algoritmo de LDA, definidos a través del indicador de coherencia. El algoritmo aplicado de la librería GENSIM nos dio mejores resultados dado que de los 12 temas, 11 eran los más representativos de algún documento. En cambio, en la librería sklearn los temas eran menos específicos dado que sólo 6 temas eran los más representativos en algún documento.
- 4. El análisis de sentimiento se aplicó para determinar la carga emocional que tenían los documentos, utilizando un lexicón desarrollado por la universidad Javeriana. Encontrándose que el documento menos positivo (neutro) es el referido al "Análisis de Clúster para la clasificación de Datos Económicos desarrollado" en Ecuador en el año 2016, mientras que el de mayor carga positiva fue el de "Desarrollo de Clúster como un modelo en el desarrollo de Negocio" desarrollado en Colombia en el año 2007. De Manera global los papers de Comentarios Web y Negocios Internacionales tienen una carga más positiva en su narrativa.
- **5.** A través de distintas técnicas de minería de Texto cómo es clustering, Topic Modeling y Análisis de sentimiento se pudo analizar y comprender el contenido de los diferentes papers que se utilizaron como fuente de información en la elaboración de los proyectos de tesis.