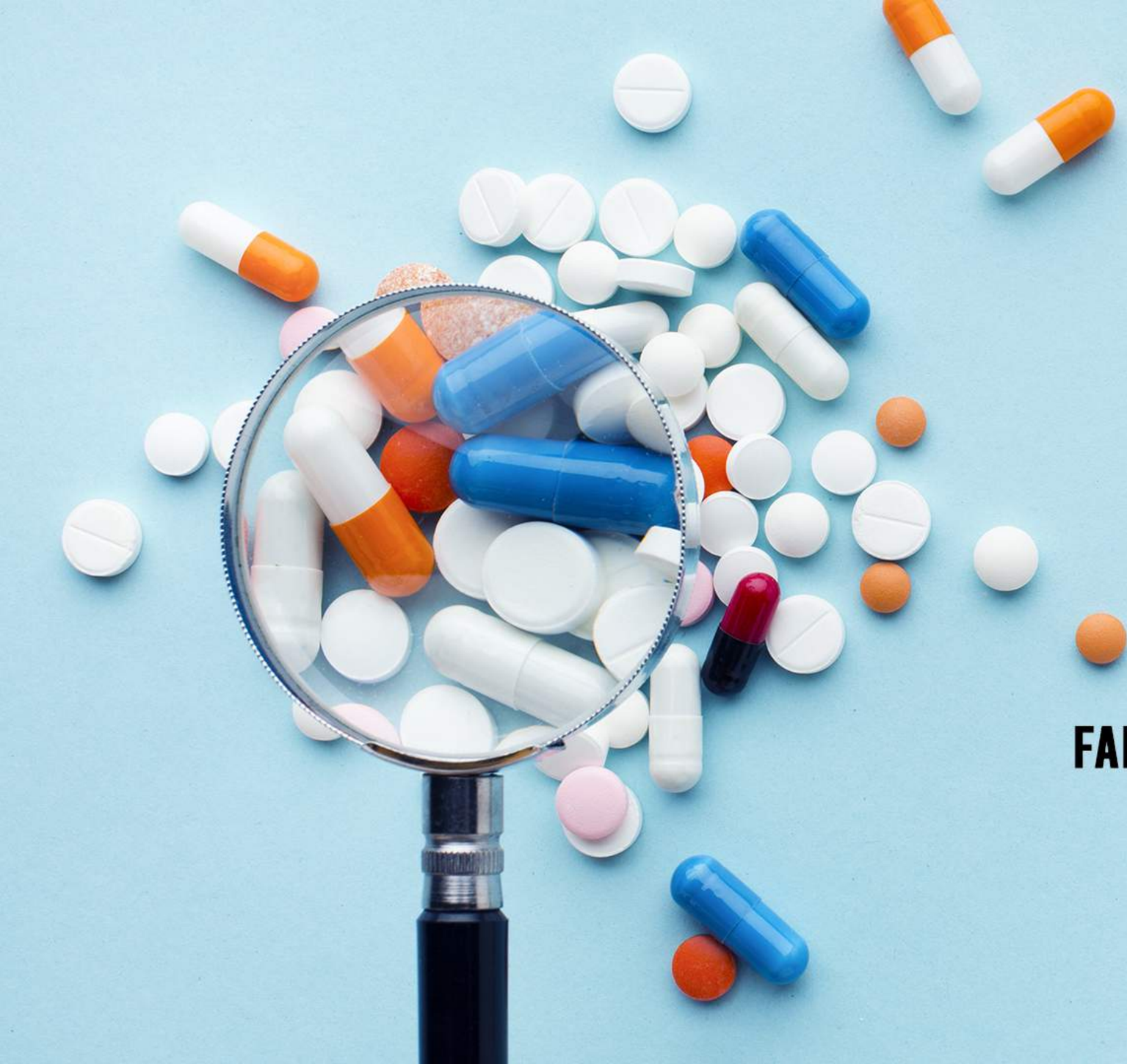


Saturdays.AI
La Paz

APLICANDO MACHINE LEARNING

PHARMASAF:
FARMACOVIGILANCIA INTELIGENTE

PHARMASAF:
FARMACOVIGILANCIA INTELIGENETE

PHARMASAF ▶

▶ NOMBRES UNICOS DE MEDICAMENTOS COMERCIALES ▶ ABYDOS ▶ OBTENCION DE MEDICAMENTOS CONFUSOS

▶ K-MEANS ▶ COEFICIENTE DE SILLHOUETTE ▶ AGRUPAMIENTO DE MEDICAMENTOS

Saturdays.AI
La Paz

CONTEXTO Y MOTIVACION

¿Y SI PUDIESEMOS PREVENIR MILES DE MUERTES ANUALES POR ERRORES DE MEDIACION?

LA FARMACOVIGILANCIA FUE DEFINIDA EN 2002 POR LA ORGANIZACIÓN MUNDIAL DE LA SALUD (OMS) COMO LA CIENCIA Y LAS ACTIVIDADES RELATIVAS A LA DETECCIÓN, EVALUACIÓN, COMPRENSIÓN Y PREVENCIÓN DE LOS EFECTOS ADVERSOS DE LOS MEDICAMENTOS O CUALQUIER OTRO PROBLEMA RELACIONADO CON ELLOS.

UNA REACCIÓN ADVERSA A UN MEDICAMENTO (RAM) ES LA RESPUESTA NOCIVA, NO DESEADA Y NO INTENCIONADA QUE SE PRODUCE TRAS LA ADMINISTRACIÓN DE UN FÁRMACO, MISMA QUE PUEDEN SER LEVE, MODERADA, GRAVE Y EN ALGUNOS CASOS PUEDEN DERIVAR EN INGRESOS HOSPITALARIOS E INCLUSO MUERTE.

¿CONOCEMOS LAS REACCIONES ADVERSAS QUE PRODUCEN TODOS LOS MEDICAMENTOS?

SOLO CONOCEMOS LOS EFECTOS ADVERSOS MÁS FRECUENTES:

- MUY FRECUENTES - 1 DE CADA 10 PACIENTES
- FRECUNETES - 1 DE CADA 100 PACIENTES
- INFRECUENTES - 1 DE CADA 1000 PACIENTES

ALGUNOS DATOS

- ALGUNOS ESTUDIOS HAN CONCLUIDO QUE EL 41% DE LOS PACIENTES AMBULATORIOS TRATADOS CON FÁRMACOS Y HASTA EL 46% DE LOS INDIVIDUOS HOSPITALIZADOS SUFRIERON EN ALGÚN MOMENTO ALGUNA RAM.

- EN EEUU LAS RAM CONSTITUYEN LA 4ª – 6ª CAUSA DE MUERTE, REPRESENTAN MÁS DEL 10% DE LAS CAUSAS DE INGRESOS HOSPITALARIOS Y EL 15-20% DE LOS PRESUPUESTOS DE LOS HOSPITALES SE INVIERTE EN EL TRATAMIENTO DE RAM.

-UN INFORME ELABORADO POR EL SINDICATO ARGENTINO DE FARMACEÚTICOS Y BIOQUÍMICOS (SAFY) DA CUENTA QUE 8 DE CADA 10 ARGENTINOS SE AUTO MEDICAN Y EL 50 POR CIENTO DE LA POBLACIÓN TOMA LOS MEDICAMENTOS DE FORMA INCORRECTA, LO QUE GENERA MÁS DE 60 MUERTES POR DÍA.
WWW.SAFYB.ORG.AR

- JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION
(JAMA) PUBLICADO EN LA LITERATURA CIENTÍFICA, DOCUMENTA LA TRAGEDIA
DEL PARADIGMA MÉDICO TRADICIONAL

LA AUTORA ES LA DRA. BARBARA STARFIELD, DE LA ESCUELA DE HIGIENE Y SALUD PÚBLICA JOHN HOPKINS, Y DESCRIBE CÓMO EL SISTEMA DE SALUD DE LOS EEUU PUEDE CONTRIBUIR A LA MALA SALUD.

TODAS ESTAS SON MUERTES POR AÑO:
• 7.000 – ERRORES DE MEDICACIÓN EN HOSPITALES (9)
• 106.000 – EFECTOS NEGATIVOS DE MEDICAMENTOS (QUE NO SON ERRORES) (2)
¡ESTO SUMA 112.000 MUERTES POR AÑO POR CAUSAS IATROGÉNICAS!
WWW.JAMANETWORK.COM

- LA NOTA "STRENGTHENING PHARMACOVIGILANCE TO REDUCE ADVERSE EFFECTS OF MEDICINES", EN LA QUE EXPONE QUE EL 5% DE LOS INGRESOS A URGENCIAS SON DEBIDOS A RAM LO QUE SUPONE 145.000 MILLONES DE EUROS A LOS SISTEMAS DE SALUD Y ESTIMAN QUE SE PRODUCEN 197.000 MUERTES POR RAM AL AÑO EN LA UE.
(HTTPS://COFZARAGOZA.ORG/REACCIONES-ADVERSAS-A-MEDICAMENTOS-Y-LA-IMPORTANCIA-DE-NOTIFICARLAS/

EN ESTE SENTIDO SE HA VISTO CONVENIENTE EL UTILIZAR MECANISMOS QUE PERMITAN EL PROCESAMIENTO DE LA GRAN CANTIDAD DE INFORMACIÓN QUE SE GENERA PERMANENTEMENTE EN LA INDUSTRIA FARMACEUTICA Y ASI COADYUVAR EN LA FARMACOVIGILANCIA.

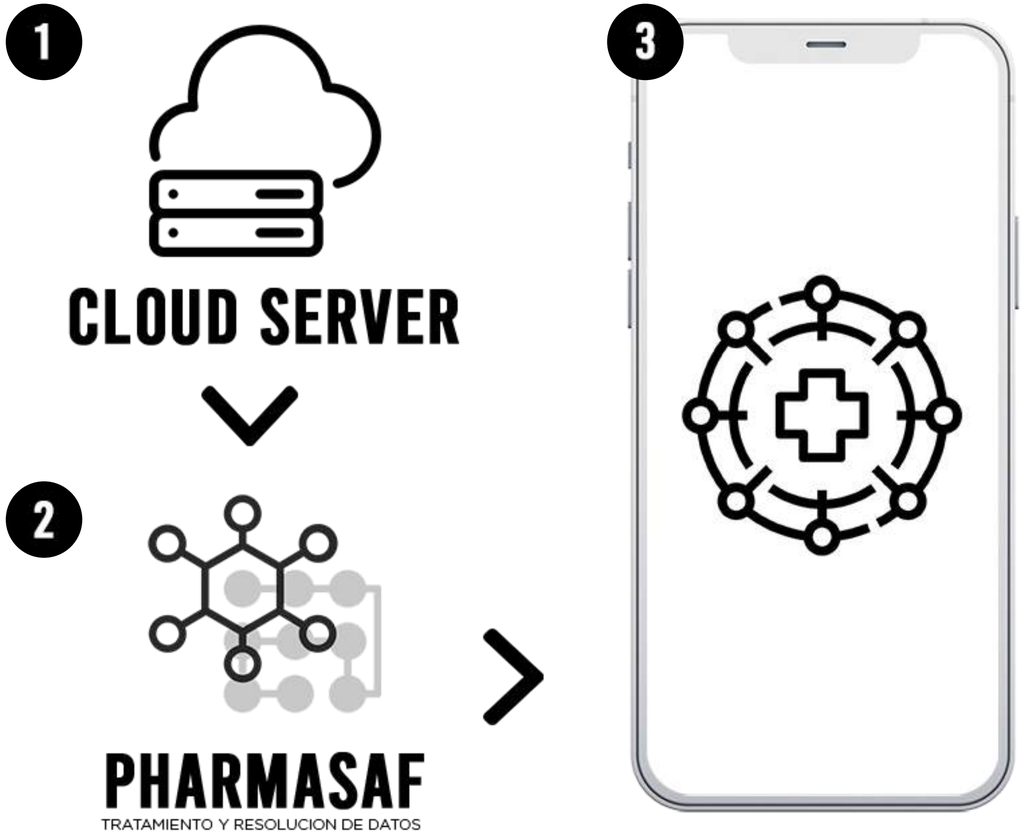
OBJETIVO

EN EL MARCO DE GENERAR MECANISMOS QUE CONTRIBUYAN A LA FARMACOVIGILANCIA, A TRAVÉS DE LA INTELIGENCIA ARTIFICIAL Y EL USO DE MODELOS DE MACHINE LEARNING, EL PRESENTE PROYECTO PERSIGUE DOS OBJETIVOS:

- EN BASE A LA LISTADO DE MEDICAMENTOS DE REFERENCIA 2020 EN MEXICO, REALIZAR LA AGRUPACIÓN DE MEDICAMENTOS EN BASE A SUS EFECTOS ADVERSOS Y CONTRAINDICACIONES.

- DETECCIÓN DE NOMBRES CONFUSOS DE MEDICAMENTOS, EN BASE A LA MEDICIÓN DE LAS DISTANCIAS ENTRE LOS MISMOS.

POR LO TANTO PLANTEAMOS UNA SOLUCION A GRAN ESCALA QUE SEA CAPAZ DE ALAMACENAR UNA BASE DE DATOS DISPONIBLE NO SOLO CON EL OBJETIVO DE CONSULTAR DICHA INFORMACION SINO TAMBIEN QUE LOS PROFESIONALES DE LA SALUD SEAN CAPACES DE COMPARAR Y ENTENDER LAS CONSECUENCIAS FARMACOLOGICAS DE CADA MEDICAMENTO Y SU USO ADECUADO SEGUN EL CASO Y QUE ADEMAS SEA ACCECIBLE Y DE SIMPLE USABILIDAD.



Saturdays.AI
La Paz

TECNICAS IMPLEMENTADAS

LIMPIANDO EL DATASET

DESPUES DE LA SELECCION DEL DATASET INDICADO:

- CUADRO BASICO DE MEDICAMENTOS
EL INSTITUTO MEXICANO DEL SEGURO SOCIAL
IMSS - HTTP://WWW.IMSS.GOB.MX

ORGANIZANDO LA INNFORMACION PARA CORRECTA LECTURA, PRIMERO SE LIMPIO EL DATASET PARA TENER UN HOMOGENEIDAD EN SU ESTRUCTURA, CORRIGIENDO:
- MINUSCULAS A MAYUCULAS
- REMOVER ES ESPACIOS EN BLANCO PROLONGADOS A MAS DE 1.
- SIGNOS DE PUNTUACION , . : ;
- ACENTOS
- REMOVER STEPWORDS COMO ARTÍCULOS, PRONOMBRES, PREPOSICIONES
- LEMATIZAR O DEVOLVER PALABRAS A SU RAIZ SEMANTICA (DIAS A DIA)

PARA LUEGO REALIZAR EL CLUSTERING DE MEDICAMENTOS.

K-MEANS

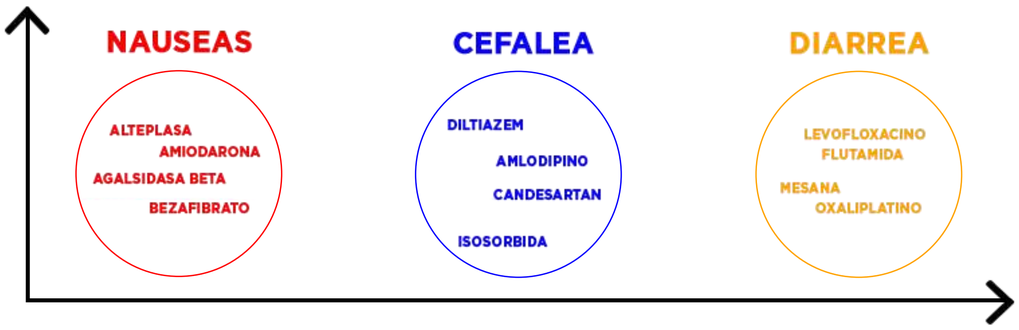
"K-MEDIAS ES UN MÉTODO DE AGRUPAMIENTO, QUE TIENE COMO OBJETIVO LA PARTICIÓN DE UN CONJUNTO DE N OBSERVACIONES EN K GRUPOS"

LOS DATOS RELEVANTES A EVALUAR DEL DATASET ERA: EFECTOS ADVERSOS Y CONTRAINDICACIONES.

	Grupo	Medicamento	Generalidades	Efectos adversos	Contraindicaciones
0	CARDIOLOGIA	ADENOSINA	NUCLEOTIDO END	DISNEA, ENROJECI	HIPERSENSIBILIDA
1	CARDIOLOGIA	ALPROSTADIL	ALPROSTADIL ES	APNEA, FIEBRE, RI	HIPERSENSIBILIDA
2	CARDIOLOGIA	ALTEPLASA	BLOQUEADOR DE	NAUSEA, VOMITO,	HIPERSENSIBILIDA
3	CARDIOLOGIA	AMIODARONA	BLOQUEADOR DE	NAUSEA, VOMITO,	HIPERSENSIBILIDA
4	CARDIOLOGIA	AMLODIPINO	BLOQUEADOR DE	CEFALEA, FATIGA,	HIPERSENSIBILIDA
5	CARDIOLOGIA	CANDESARTAN CIL	ANTAGONISTA DE	CEFALEA, DOLOR I	HIPERSENSIBILIDA
6	CARDIOLOGIA	CILOSTAZOL	DERIVADO DE LA	UINOLINONA, CON E	HIPERSENSIBILIDA
7	CARDIOLOGIA	CLOPIDOGREL	ANTAGONISTA DE	DIARREA, SANGRA	HIPERSENSIBILIDA
8	CARDIOLOGIA	DIAZOXIDO	VASODILATADOR	HIPERGLUCEMIA, I	HIPERSENSIBILIDA
9	CARDIOLOGIA	DIGOXINA	REFUERZAN LA CO	ANOREXIA, NAUSE	HIPERSENSIBILIDA
10	CARDIOLOGIA	DILTIAZEM	BLOQUEADOR DE	CEFALEA, CANSAN	INFARTO AGUDO D
11	CARDIOLOGIA	DIPIRIDAMOL	ANTIPLAQUETARI	DOLOR ABDOMINA	HIPERSENSIBILIDA
12	CARDIOLOGIA	DOBUTAMINA	INOTROPICO DE A	TAQUICARDIA, HIP	HIPERSENSIBILIDA
13	CARDIOLOGIA	DOPAMINA	EFECTO ADRENEF	NAUSEA, VOMITO,	HIPERSENSIBILIDA
14	CARDIOLOGIA	EFEDRINA	SIMPATICOMIMET	INSOMNIO, DELIRIO	HIPERSENSIBILIDA
15	CARDIOLOGIA	ENALAPRIL O LISIN	INHIBEN A LA ENZ	CEFALEA, MAREO,	HIPERSENSIBILIDA
16	CARDIOLOGIA	EPINEFRINA	ESTIMULA A LOS	C HIPERTENSION A	INSUFICIENCIA VA

18	CARDIOLOGIA	ESTREPTOQUINAS	C FORMA UN COM	HEMORRAGIA, ARF	HIPERSENSIBILIDA
19	CARDIOLOGIA	FELODIPINO	BLOQUEADOR DE L	DEBIDOS A SU EFE	CHOQUE CARDIOG
20	CARDIOLOGIA	IBUPROFENO	NITRATO QUE DISM	TAQUICARDIA, ARF	HIPERSENSIBILIDA
21	CARDIOLOGIA	ISOSORBIDA	NITRATO QUE DISM	TAQUICARDIA, MAI	HIPOTENSION ART
22	CARDIOLOGIA	ISOSORBIDA, DINIT	NITRATO QUE DISM	TAQUICARDIA, ARF	HIPERSENSIBILIDA
23	CARDIOLOGIA	ISOSORBIDA, MONO	NITRATO QUE INCI	CEFALEA, VERTIGI	HIPERSENSIBILIDA
24	CARDIOLOGIA	LEVOSIMENDAN	AUMENTA LA CONT	CEFALEA, HIPOTEI	HIPERSENSIBILIDA
25	CARDIOLOGIA	LIDOCAINA	BLOQUEADOR DE	HIPOTENSION, AGI	HIPERSENSIBILIDA
26	CARDIOLOGIA	LOSARTAN	PROFARMACO ANT	SEDACION, HIPOTE	HIPERSENSIBILIDA
27	CARDIOLOGIA	METILDOPA	PROFARMACO ANT	SEDACION, HIPOTE	HIPERSENSIBILIDA
28	CARDIOLOGIA	METOPROLOL	ANTAGONISTA CA	HIPOTENSION ART	HIPERSENSIBILIDA
29	CARDIOLOGIA	MILRINONA	INHIBIDOR SELECT	ARRITMIAS SUPRA	HIPERSENSIBILIDA
30	CARDIOLOGIA	NOREPINEFRINA	NEUROTRANSMISO	CEFALEA, TAQUIC	HIPERSENSIBILIDA
31	CARDIOLOGIA	PENTOXIFILINA	DERIVADO METILX	ANTINICO QUE REDI	HIPERSENSIBILIDA
32	CARDIOLOGIA	POTASIO, SALES D	ELECTROLITO ESE	ARRITMIAS CARDI	HIPERSENSIBILIDA
33	CARDIOLOGIA	PRAZOSINA	BLOQUEA LA COR	ANOREXIA, NAUSE	HIPERSENSIBILIDA
34	CARDIOLOGIA	PROPRANOLOL	ANTAGONISTA B A	BRADICARDIA, HIP	HIPERSENSIBILIDA

YA QUE AL AGRUPARLOS EN DISTINTOS CLUSTERS NOS PERMITIRIA ENTENDER LA RELACION ENTRE GRUPOS DE MEDICAMENTOS Y PREDECIR SUS EFECTOS ADVERSOS ADEMAS DE OBTENER MAYOR FECUENCIA DE UN EFECTO ADEVERSO SEGUN EL GRUPO AL QUE PERTENECE



Saturdays.AI
La Paz

EVALUCACION DEL METODO BASE

SHILOUETTE // CALINSKI-HARABASZ // DAVIES-BOULDIN

"EL COEFICIENTE DE SILUETA ES UNA MÉTRICA
PARA EVALUAR LA CALIDAD DEL AGRUPAMIENTO OBTENIDO CON
ALGORITMOS DE CLUSTERING. EL OBJETIVO DE SILUETA
ES IDENTIFICAR CUÁL ES EL NÚMERO ÓPTIMO DE
AGRUPAMIENTOS. "

PUDIMOS CONSTATAR QUE CLUSTER DEL K-MEANS PREVIO ERA EL MAS OPTIMO PARA PARA OBTENER GRUPO HOMOGENEOS EN BASE A SUS EFECTOS ADVERSOS Y CONTRAINDICACIONES, POR LO TANTO OBNETER MAYOR CALIDAD.

PARA ESTO DEBIMOS APLICAR DISTINTAS CONFIGURACIONES PARA PODER OBTENER LA MEJOR COMBIANCION SIN LA INTERFENCIA DE OTRO PARAMETROS O PALABRAS INNECESARIAS EN EL DATASET

- MINUSCULAS A MAYUCULAS
- REMOVER ES ESPACIOS EN BLANCO PROLONGADOS A MAS DE 1.
- SIGNOS DE PUNTUACION , . : ;
- ACENTOS
- REMOVER STEPWORDS COMO ARTÍCULOS, PRONOMBRES, PREPOSICIONES
- LEMATIZAR O DEVOLVER PALABRAS A SU RAIZ SEMANTICA (DIAS A DIA)
- MATENER LA FUENTES EN UNICODE PARA EVITAR ERRORES ENTRE LA ORTOGRAFIA EN ESPAÑOL E INGLES.

EN LOS ALGORITMOS DE APRENDIZAJE NO SUPERVISADO, LA CANTIDAD DE GRUPOS PUEDE SER UN PARÁMETRO DE ENTRADA DEL ALGORITMO O PUEDE SER DETERMINADO AUTOMÁTICAMENTE POR EL ALGORITMO.

EN EL PRIMER CASO, COMO OCURRE CON EL ALGORITMO DE K-MEAN, LA DETERMINACIÓN DEL NÚMERO ÓPTIMO DE CLUSTERS TIENE QUE SER REALIZADO MEDIANTE ALGUNA MEDIDA EXTERNA AL ALGORITMO. EL COEFICIENTE DE SILUETA ES INDICADOR DEL NÚMERO IDEAL DE CLUSTERS.

UN VALOR MÁS ALTO DE ESTE ÍNDICE INDICA UN CASO MÁS DESEABLE DEL NÚMERO DE CLUSTERS

COEFICIENTE DE SHILOUETTE
COMPARA LA DISTANCIA DE UNA MUESTRA RESPECTO AL CENTRO DE SU GRUPO CON LA DISTANCIA AL GRUPO MÁS CERCANO, OTORGÁNDOLE UN VALOR ENTRE -1 Y 1. UN VALOR NEGATIVO INDICA QUE LA MUESTRA DEBERÍA ENCONTRARSE EN OTRO GRUPO, MIENTRAS QUE UN VALOR DE 1 INDICARÍA QUE ESTÁ JUNTO A OTRAS MUESTRAS DEL GRUPO EL EN QUE SE ENCUENTRA.

CALISNKI-HARABASZ
QUE PERMITE COMPARAR DOS AGRUPAMIENTOS ENTRE SÍ PARA CONOCER CUAL DE ELLOS TIENE SUS GRUPOS MEJOR DEFINIDOS; ES DECIR, CUANDO MAYOR SEA ÉSTE PARÁMETRO, SUS MUESTRAS SE ENCUENTRAN MENOS DISPERSADAS Y, A SU VEZ, ALEJADAS DE OTROS GRUPOS. ESTE ALGORITMO NO ESTÁ ACOTADO SUPERIORMENTE POR LO QUE ES DIFÍCIL EVALUARLO DE FORMA AISLADA, SIEMPRE DEBE SER UTILIZADO PARA COMPARAR DISTINTAS CLASIFICACIONES.

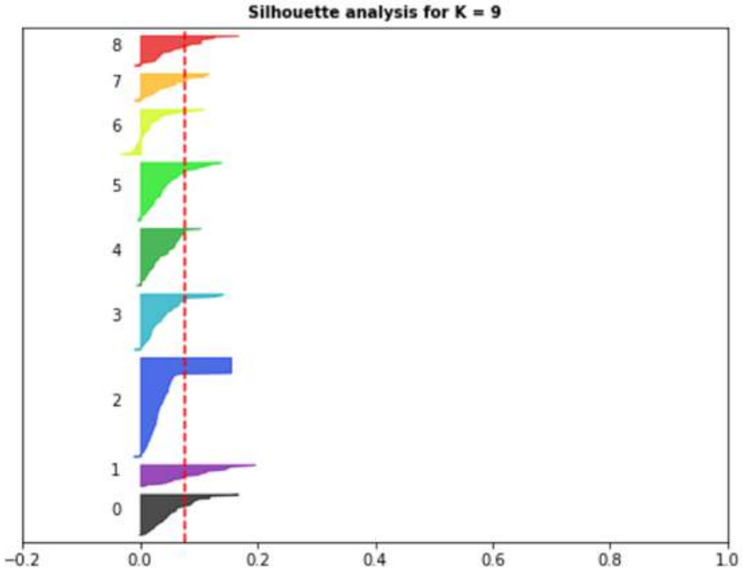
ÍNDICE DAVIES-BOULDIN
BUSCA MINIMIZAR UNA FUNCIÓN OBJETIVO: LA DISTANCIA DE LAS MUESTRAS AL CENTRO DE SU GRUPO. CUANTO MÁS CERCANO SEA EL VALOR A 0, MEJOR SE CONSIDERA LA AGRUPACIÓN, PUES IMPLICA QUE LAS MUESTRAS ESTÁN JUNTAS EN SUS GRUPOS Y A SU VEZ ESTÁN MUY SEPARADOS DEL RESTO. UNA AGRUPACIÓN QUE INDIQUE GRUPOS COMPACTOS Y SEPARADOS TENDRÁ VALORES INFERIORES A 0,7 EN ESTE ÍNDICE, MIENTRAS QUE UN VALOR SUPERIOR PUEDE INDICAR GRUPOS SOLAPANTES O DISPERSOS.

EL COEFICIENTE DE SILHOUETTE Y CALINSKI-HARABASZ, ESTE PARÁMETRO NOS DA UNA IDEA DE COMO DE BUENA ES LA AGRUPACIÓN Y SI PODEMOS UTILIZARLA PARA COMPARAR LAS VARIEDADES DE TOMATE. COMPARANDO LOS TRES PARÁMETROS, PODEMOS SELECCIONAR EL MODELO0 QUE MEJOR OPTIMICE ESTOS COEFICIENTES PARA ESTABLECER CUAL ES EL MEJOR MODELO.

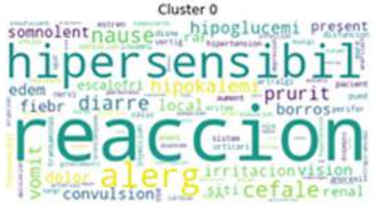
RESULTADOS
EFECTOS ADVERSOS

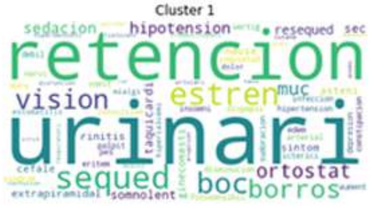
SILHOUETTE: 0.0764
CALINSKI HARABASZ: 10.2622
DAVIES BOULDIN: 3.9869

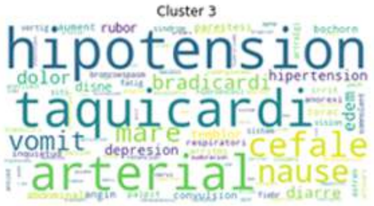
M_STOPWORDS = TRUE
M_STEM = TRUE
M_LEMMATIZE = TRUE
N_PALABRAS_MAS_FRECUENTES = 3
PALABRAS_FRECUENTES = FALSE
PALABRAS_MENOS_FRECUENTES = TRUE
N_GRAMAS_SIZE = (1,1)
K = 10

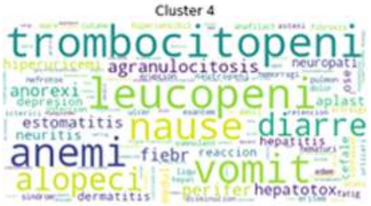


EFECTOS ADVERSOS









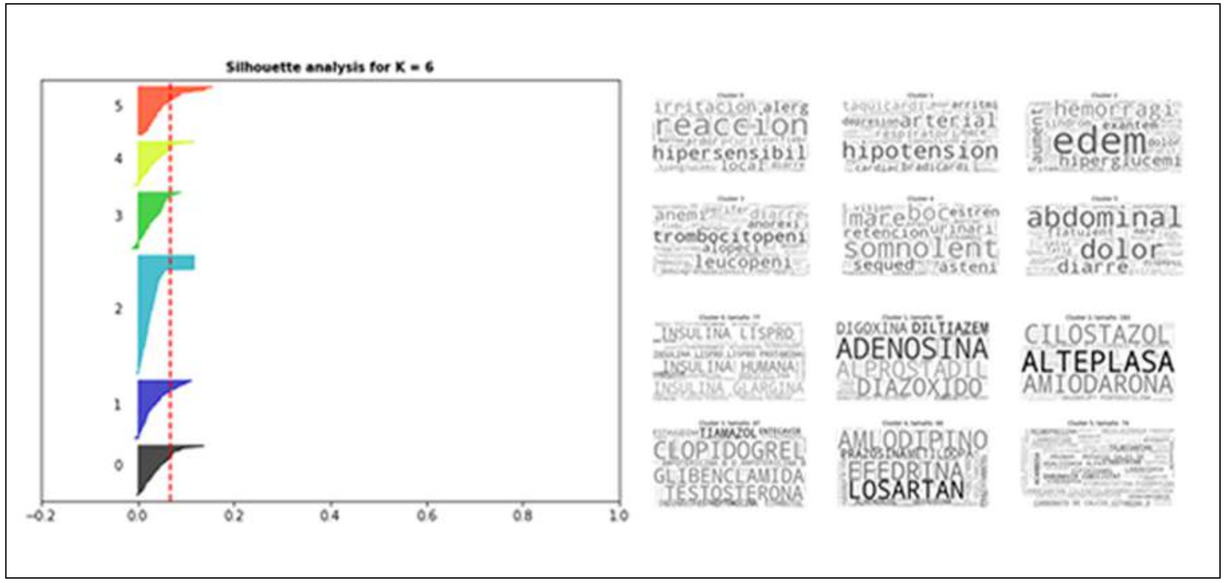
MEDICAMENTOS

Saturdays.AI
La Paz

PHARMASAF:
FARMACOVIGILANCIA INTELIGENETE

SILHOUETTE: 0.0666
CALINSKI HARABASZ: 11.7709
DAVIES BOULDIN: 4.5434

M_STOPWORDS = TRUE
M_STEM = TRUE
M_LEMMATIZE = TRUE
N_PALABRAS_MAS_FRECUENTES = 3
PALABRAS_FRECUENTES = TRUE
PALABRAS_MENOS_FRECUENTES = TRUE
N_GRAMAS_SIZE = (1,1)
K = 10



SILHOUETTE: 0.0718
CALINSKI HARABASZ: 10.6725
DAVIES BOULDIN: 4.1231

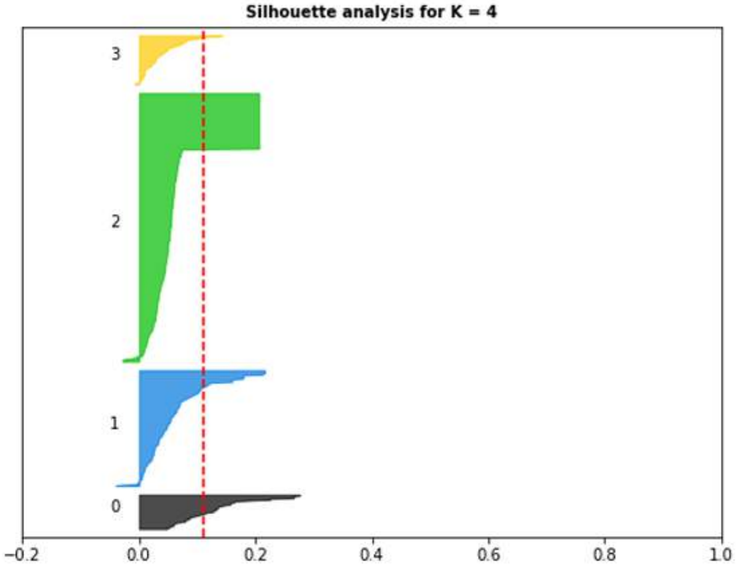
M_STOPWORDS = FALSE
M_STEM = FALSE
M_LEMMATIZE = TRUE
N_PALABRAS_MAS_FRECUENTES = 3
PALABRAS_FRECUENTES = TRUE
PALABRAS_MENOS_FRECUENTES = TRUE
N_GRAMAS_SIZE = (1,1)
K = 10

Saturdays.AI
La Paz

RESULTADOS SILLHOUETTE CONTRA INDICACIONES

SILHOUETTE: 0.1103
CALINSKI HARABASZ: 19.1528
DAVIES BOULDIN: 3.5482

M_STOPWORDS = TRUE
M_STEM = FALSE
M_LEMMATIZE = FALSE
N_PALABRAS_MAS_FRECUENTES = 3
PALABRAS_FRECUENTES = TRUE
PALABRAS_MENOS_FRECUENTES = TRUE
N_GRAMAS_SIZE = (1,1)
K = 10

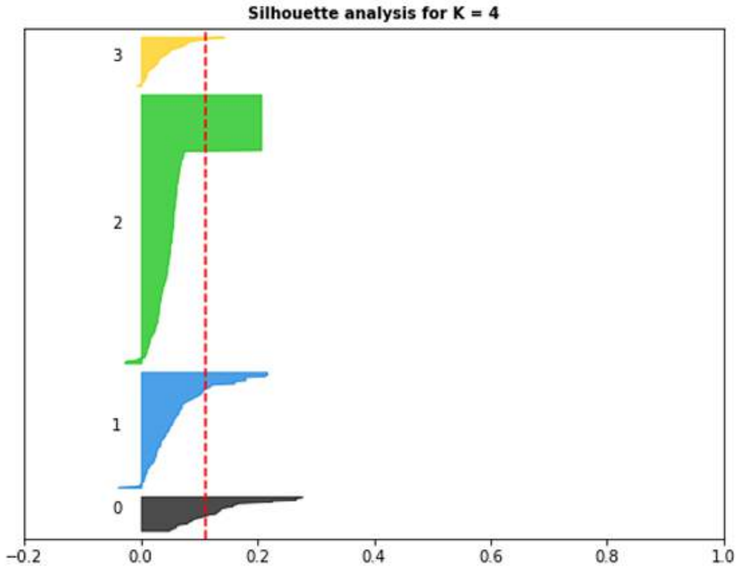


Saturdays.AI
La Paz

RESULTADOS SILLHOUETTE CONTRA INDICACIONES

SILHOUETTE: 0.1103
CALINSKI HARABASZ: 19.1528
DAVIES BOULDIN: 3.5482

M_STOPWORDS = TRUE
M_STEM = FALSE
M_LEMMATIZE = TRUE
N_PALABRAS_MAS_FRECUENTES = 3
PALABRAS_FRECUENTES = TRUE
PALABRAS_MENOS_FRECUENTES = TRUE
N_GRAMAS_SIZE = (1,1)
K = 10



Saturdays.AI
La Paz

TECNICAS IMPLEMENTADAS

LIMPIANDO EL DATASET

DESPUES DE LA SELECCION DEL DATASET INDICADO:

- LISTADO DE MEDICAMENTOS DE REFERENCIA
COMISIÓN FEDERAL PARA LA PROTECCIÓN CONTRA RIESGOS SANITARIOS
COFEPRIS - WWW.GOB.MX/COFEPRIS

ORGANIZANDO LA INNFORMACION PARA CORRECTA LECTURA, PRIMERO SE LIMPIO EL DATASET PARA TENER UN HOMOGENEIDAD EN SU ESTRUCTURA, CORRIGIENDO:

- MINUSCULAS A MAYUCULAS
- REMOVER ES ESPACIOS EN BLANCO PROLONGADOS A MAS DE 1.
- SIGNOS DE PUNTUACION , . : ;
- ACENTOS
- REMOVER STEPWORDS COMO ARTÍCULOS, PRONOMBRES, PREPOSICIONES
- LEMATIZAR O DEVOLVER PALABRAS A SU RAIZ SEMANTICA (DIAS A DIA)
- MATENER LA FUENTES EN UNICODE PARA EVITAR ERRORES ENTRE LA ORTOGRAFIA EN ESPAÑOL E INGLES.

DETECCION DE NOMBRES DE MEDICAMENTOS - SIMILITUDES

EXTRACCIÓN
DADO QUE LA INFORMACIÓN VIENE CONTENIDA EN FORMATO PDF, FUE NECESARIO UTILIZAR LA LIBRERÍA PDF2DOCX LA CUAL PERMITE EXTRAER DATOS DE ARCHIVOS PDF SIEMPRE Y CUANDO ESTOS ESTÉN ALMACENADOS EN FORMA TABULAR.

SELECCIÓN
LA LISTA DE COFEPRIS CONTIENE 1385 REGISTROS DE MEDICAMENTOS, DONDE ALGUNOS DE LOS MEDICAMENTOS SON LOS MISMO CON UNA VARIANTE EN SU DOSIS. ASÍ MISMO, INCLUYE MEDICAMENTOS CON MAS DE UNA PALABRA EN SU NOMBRE COMERCIAL. PARA ESTE PROYECTO, SE SELECCIONARON SOLO LOS NOMBRES DE MEDICAMENTOS ÚNICOS Y QUE CONTUVIERAN SOLO UNA PALABRA EN SU NOMBRE COMERCIAL, DEJANDO ASÍ UNA LISTA DE 1035 NOMBRES DE MEDICAMENTOS.

TRANSFORMACIÓN
A LA LISTA DE LOS 1035 NOMBRES DE MEDICAMENTOS, SE LES APLICO UN PROCESO DE LIMPIEZA PARA ASÍ LOGRAR TENER UNA BASE CONSISTENTE: SE TRANSFORMARON LOS DATOS A MAYÚSCULAS, SE ELIMINARON ACENTOS, DÍGITOS Y ESPACIOS EN BLANCO. ESTO EN RECOMENDADO CUANDO SE TRABAJA CON DATOS TEXTUALES.

CÁLCULO DE DISTANCIA
EL OBJETIVO ES ENCONTRAR UNA LISTA DE PARES DE NOMBRES DE MEDICAMENTOS CONFUSOS, Y PARA ELLO ES NECESARIO CALCULAR QUE TAN PARECIDOS SON DOS ELEMENTOS ENTRE SÍ, PARA ESTE CASO, ES NECESARIO CALCULAR QUE TAN PARECIDOS SON TODOS LOS DATOS.
AL TRABAJAR CON SISTEMAS DE PROCESAMIENTO DE LENGUAJE NATURAL (NLP) ES FRECUENTE ENCONTRAR LA NECESIDAD DE COMPARAR DIFERENTES PALABRAS O FRASES ENTRE SÍ O DE BUSCAR PATRONES DE CARACTERES DENTRO DE UN TEXTO, ES DE INTERÉS EL ENCONTRAR NO SOLAMENTE LAS COINCIDENCIAS EXACTAS ENTRE DOS CADENAS DE TEXTO, SINO TAMBIÉN EL TENER UNA MEDIDA DE APROXIMACIÓN O SIMILITUD ENTRE ESTAS CUANDO LA COINCIDENCIA NO ES PERFECTA. UNA MÉTRICA ES UNA FUNCIÓN MATEMÁTICA QUE DEFINE UNA DISTANCIA ENTRE CADA PAR DE ELEMENTOS DE UN CONJUNTO, DONDE EL VALOR 0 INDICA UNA DISTANCIA NULA (LOS ELEMENTO SON SIMILARES) Y 1 INDICA QUE LOS ELEMENTOS SON TOTALMENTE DIFERENTES.
EXISTE UN ALTO NÚMERO DE MÉTRICAS DE DISTANCIA, CADA UNA CON DIFERENTES PECULIARIDADES QUE LA HACEN MÁS ADEPTA PARA ALGUNA APLICACIÓN EN PARTICULAR.

EVALUCACION DEL METODO BASE

ABYDOS

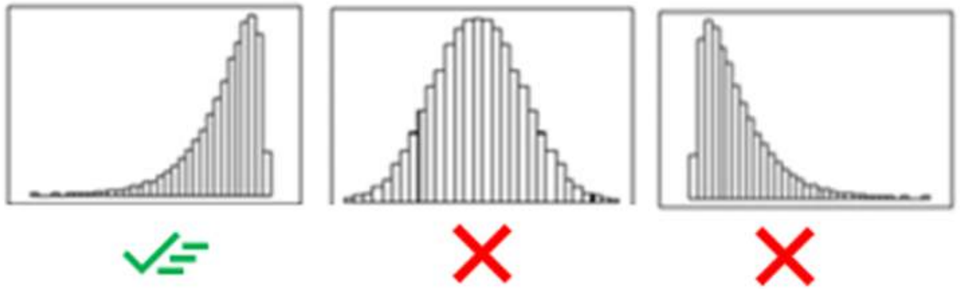
- A PARTIR DE LOS 1035 NOMBRES DE MEDICAMENTOS COMERCIALES ÚNICOS, REALIZAR LA COMPARACIÓN DE TODOS CONTRA TODOS, PARA DE ESTA FORMA ENCONTRAR LOS PARES CON MENOR VALOR DE DISTANCIA (MÁS SIMILARES).
- ESTO SE OBTIENE A TRAVÉS DEL PRODUCTO PUNTO DE LA LISTA DE LOS 1035 ELEMENTOS, LO CUAL NOS GENERA 1,071,225 PARES DE NOMBRES DE MEDICAMENTO (1035 * 1035).

- PARA CADA PAR DE NOMBRES DE MEDICAMENTOS, DE DEBE APLICAR UNA MÉTRICA DE DISTANCIA, PARA ASÍ OBTENER LOS VALORES CORRESPONDIENTES A CADA PAR Y ENCONTRAR LOS MÁS SIMILARES. EN ESTE PROYECTO, SE SELECCIONARON 14 MÉTRICAS DE DISTANCIA INCLUIDAS EN ABYDOS LAS CUALES SON DE COMPARACIÓN DE CADENAS DE TEXTO: LEVENSHTEIN, OSA, FLEXMETRIC, BISIM, PHONETIC EDIT DISTANCE, STRCMP95, PREFIX, SUFFIX; Y DE COMPARACIÓN FONÉTICA: MRA, DAVIDSON, DOLBY, PARMAR KUMBHARANA, PHONETIC SPANISH, SPANISH METAPHONE, SOUNDEX.

- A CADA UNO DE LOS 1,071,225 PARES DE NOMBRES SE LES APLICA CADA UNA DE LAS 14 MÉTRICAS DE DISTANCIA INDICADAS.

SELECCIÓN DE RANGO

PARA CADA UNA DE LAS MÉTRICAS DE SIMILITUD APLICADAS, SE SELECCIONÓ EL RANGO DONDE SE CONSIDERARÍA QUE LOS PARES DE NOMBRES DE MEDICAMENTOS SON CONFUSOS. ESTO SE REALIZÓ MEDIANTE EL ANÁLISIS DE DISTRIBUCIÓN. EN GENERAL, SI LA DISTRIBUCIÓN DE LA MÉTRICA DE DISTANCIA ES UNA DISTRIBUCIÓN SEGADA A LA IZQUIERDA, INDICA UNA BUENA MÉTRICA DE DISTANCIA; Y SI LA DISTRIBUCIÓN ES NORMAL (SIMÉTRICA CENTRAL) O SESGADA A LA DERECHA, LA MÉTRICA NO ES BUENA. DESPUÉS, EL RANGO DE SELECCIÓN SE OBTIENE DE LOS PASOS DE LA DISTRIBUCIÓN, DONDE 0 < N < MC, DONDE MC REFLEJA EL PASO MÁXIMO ANTES DE QUE LA DISTRIBUCIÓN AGRUPARA DEMASIADOS VALORES (SI LA MÉTRICA AGRUPA MUCHOS VALORES BAJO UN MISMO RANGO, SIGNIFICA QUE HA DEJADO DE ENCONTRAR PARES SIMILARES).



Saturdays.AI
La Paz

RESULTADOS ABYDOS

ABYDOS

A CONTINUACIÓN, SE MUESTRAN ALGUNAS MÉTRICAS APLICADAS, ASÍ COMO LOS PARES DE NOMBRES CONFUSOS OBTENIDAS POR ESTAS.

BISIM

Medicamento1	Medicamento2	Distancia
SYNALARC	SYNALARN	0.063
SYNALARC	SYNALARO	0.063
DAIVOBET	DAIVONEX	0.188
MICROLAX	MICROLUT	0.188
DURACEF	DURATER	0.214
CIPROLISINA	CIPROXINA	0.273
APROVEL	ATROVENT	0.313
ASPIRINA	SPIRIVA	0.313
ALDOMET	ALMETEC	0.357
ZOVIRAX	ZOFRAN	0.357

FlexMetric

Medicamento1	Medicamento2	Distancia
RECOVERON	RECOVERONNC	0.036
LUTORAL	LUTORALE	0.063
SYNALARC	SYNALARN	0.125
SYNALARC	SYNALARO	0.125
COMBODART	COMPETACT	0.156
SARIDON	VARITON	0.157
ANALGEN	ANTALGIN	0.163
ASTELIN	ESTECLIN	0.163
SINTROM	SYNTHROID	0.211
TEMPRA	KEPPRA	0.233

Levenshtein

Medicamento1	Medicamento2	Distancia
RECOVERONN	RECOVERONNC	0.091
SYNALARC	SYNALARN	0.125
LIBERAXIM	LIBERTRIM	0.222
ASPIRINA	SPIRIVA	0.250
ASTELIN	ESTECLIN	0.250
CASODEX	FASLODEX	0.250
CIPROLISINA	CIPROXINA	0.273
ALDOMET	ELOMET	0.286
VILAMIN	DIAMIN	0.286
ACROMICINA	BUCOMICINA	0.300

Dolby

Medicamento1	Medicamento2	Distancia
SUPRADOL	SUPRADOLF	0.143
TEMPOCAPS	DIMICAPS	0.143
TOBRADEX	TOBREX	0.143
UROMITEXAN	ARIMIDEX	0.143
ZYPREXA	CIPROXINA	0.143
AFUMIX	ALMAX	0.200
ALMETEC	ELOMET	0.200
ALDOMET	ELOMET	0.286
VILAMIN	DIAMIN	0.286
ACROMICINA	BUCOMICINA	0.300

Prefix

Medicamento1	Medicamento2	Distancia
SYNALARC	SYNALARN	0.125
LIBERAN	LIBERAXIM	0.143
PROPESHIA	PROPESS	0.143
FLAGYSTATINV	FLAGYL	0.167
GENTILAX	GENTILITO	0.250
TRITAZIDE	TRITACE	0.286
ALDACTONE	ALDARA	0.333
TENORMIN	TENORETIC	0.375
ADVANTAN	ADVIL	0.400
CARDINIT	CARDURA	0.429

Suffix

Medicamento1	Medicamento2	Distancia
DIABINESE	OBINESE	0.143
AVIRENA	MIRENA	0.167
BISOLVON	TOLVON	0.167
PRIMACOR	OMACOR	0.167
SUPRADOL	TRADOL	0.167
RELPAX	ULPAX	0.200
SPRIAFIL	TAFIL	0.200
ACTRON	SINESTRON	0.333
ZESTORETIC	TENORETIC	0.333
TENORETIC	MODURETIC	0.444

UNA VEZ QUE SE OBTUVIERON LAS 14 LISTAS DE POSIBLES PARES DE NOMBRES DE MEDICAMENTOS CONFUSOS, SE REALIZO UN PESADO PONDERADO PARA DETERMINAR CON BASE A LA CANTIDAD DE MÉTRICAS QUE ENCONTRARON EL MISMO PAR DE NOMBRES, CUALES SON LOS PARES DE NOMBRES QUE APARECIERON EN MÁS MÉTRICAS, MIENTRAS MÁS MÉTRICAS INDICAN QUE UN PAR DE NOMBRES ES CONFUSO, MÁS PROBABLE ES QUE SI LO SEA. A CONTINUACIÓN, SE MUESTRA UNA LISTA DE ALGUNOS DE ESTOS PARES DE NOMBRES Y SU PONDERACIÓN.

Medicamento1	Medicamento2	Ponderación
FLIXONASE	FLAGENASE	12
ANALGEN	ANALFIN	12
FLAGENASE	FLIXONASE	12
TOBRADEX	TOBREX	12
TOBREX	TOBRADEX	12
ANALFIN	ANALGEN	12
MICROLUT	MICROLAX	11
COMBIVENT	COMBIGAND	11
MINIRIN	MINOCIN	11
SYNALARN	SYNALARO	11
NORVAS	NORVIR	11
SUPRA	SUPRANE	10
PROVIRON	PROVERA	10
SUPRANE	SUPRA	10
DICETEL	MICETAL	10

Medicamento1	Medicamento2	Ponderación
MICETAL	DICETEL	10
SYNALARNEO	SYNALARC	10
TRITAZIDE	TRITACE	10
ALMETEC	DUOALMETEC	10
TRITACE	TRITAZIDE	10
CONCOR	BICONCOR	9
BICONCOR	CONCOR	9
EXELON	EXEL	9
MINITRAN	GYNOTRAN	9
CARDISPAN	CARDIOXANE	9
SINEQUAN	SINOGAN	9
APROVASC	APROVEL	8
COMBIVIR	COMBIVENT	8
APROVEL	APROVASC	8
PRIMALAN	PRIMPERAN	7

AL FINAL SE OBTUVO UNA LISTA DE 10078 PARES DE NOMBRES DE MEDICAMENTOS CONFUSOS, PONDERADAS DESDE 1 HASTA 12, DONDE 1 INDICA QUE EL PAR FUE ENCONTRADO POR UNA MÉTRICA Y 12 INDICA QUE FUE ENCONTRADO POR 12 MÉTRICAS. SE SELECCIONARON SOLO LOS PARES DE NOMBRES QUE APARECIERAN EN AL MENOS 4 MÉTRICAS DE DISTANCIA, DEJANDO ASÍ UN TOTAL DE 998 PARES DE NOMBRES DE MEDICAMENTOS CONFUSOS.

CONCLUSIONES

K-MEANS Y SILLHOUETTE

EL ESTUDIO NOS PERMITIRA EN EL FUTURO CERCANO Y LEJANO IMPLEMENTAR MEJORAS A DATASET INCLUSO CATEGORIZARLOS POR ESPECIALIDADES, ES NECESARIO LA COLABORACION A GRAN ESCALA DE DISTINTOS PAISES Y SUS INSTITUCIONES PARA LOGRAR UN "STANDAR" Y CONSEGUIR UNIFICAR LAS MEDICIONES.

SIMILITUDES Y ABYDOS

LA DETECCIÓN DE NOMBRES DE MEDICAMENTOS CONFUSOS HA SIDO AMPLIAMENTE ESTUDIADA EN IDIOMA INGLÉS, ESTUDIOS EN LOS CUALES SE PRESENTAN MÉTODOS NOVEDOSOS PARA DETECTAR CUANDO EL NOMBRE DE UN NUEVO MEDICAMENTO ES ALTAMENTE PROBABLE SE CONFUNDA CON OTRO. PERO PARA LOGRAR ESOS AVANCES, PRIMERO SE DEBE CONTAR CON UNA LISTA DE REFERENCIA QUE INDIQUE QUE PARES DE NOMBRES DE MEDICAMENTOS SON CONFUSOS. EN ESTE PROYECTO SE APOYA AL AVANCE DE ESTAS INVESTIGACIONES, GENERANDO UNA LISTA DE 998 PARES DE NOMBRES DE MEDICAMENTOS ALTAMENTE CONFUSOS PARA MEDICAMENTOS DEL IDIOMA ESPAÑOL.

COMPARTIDA

IMPLEMENTAR HERRAMIENTAS DE MAHCINE LEARNING Y DATA SCIENCE PARA PROCESAR Y OBTENER INFORMACION EFECTIVA RESPECTO A LOS EFECTOS ADVERSOS Y MEDICAMENTOS CONFUSOS QUE PERMITA ASISTIR EL TRABAJO DE LOS MEDICOS, FARMACEUTICOS Y HACER MAS EFICIENTE LA FARMACOVIGILANCIA.