



**CONACYT**

Consejo Nacional de Ciencia y Tecnología



**CentroGeo**

19°17'30"N 99°13'17"O 2489m

Centro de Investigación en  
Ciencias de Información Geoespacial, A.C.  
Centro Público de Investigación CONACYT

---

---

Detección de tweets en México relacionados con la  
depresión utilizando un esquema multilingual y destilación  
de conocimiento.

TESIS

Que para obtener el grado de

Maestro en Ciencias de Información Geoespacial

Presenta

Jorge Manuel Pool Cen

Director de la Tesis

Dr. Hugo Carlos Martínez

Ciudad de México 2023

Centro de Investigación en  
Ciencias de Información Geoespacial, A.C.  
Centro Público de Investigación CONACYT

Detección de tweets en México relacionados con la depresión utilizando un esquema  
multilingüal y destilación de conocimiento.

TESIS

Que para obtener el grado de  
Maestro en Ciencias de Información Geoespacial

Presenta

Jorge Manuel Pool Cen

Director de Tesis

Dr. Hugo Carlos Martínez

Codirector de Tesis:

Dr. Gandhi Samuel Hernández Chan

Sinodales

Dr. Alejandro Molina Villegas

Examinador Externo:

Dra. Sally Vanega Romero

Ciudad de México, Marzo, 2023

# Resumen

La salud mental es un estado de bienestar, la cual puede verse afectada por factores individuales y sociales. La pandemia provocada por Covid-19, causó diversos cambios en el estilo de vida, como el distanciamiento físico y el trabajo desde el hogar; éstos cambios y los efectos económicos de la pandemia tuvieron un impacto en la salud emocional de las personas. El estudio de la salud mental, ha despertado el interés de investigadores que se dedican al área de Aprendizaje Máquina y Procesamiento de Lenguaje Natural (NLP, por sus siglas en inglés), para desarrollar algoritmos que contribuyan en la detección de problemáticas de salud mental como ansiedad, depresión, estrés, ideación del suicidio, etcétera. El presente trabajo pretende aportar a este tipo de estudios construyendo un modelo de clasificación de tweets (únicamente el texto) relacionados a la depresión en idioma español mexicano, en el período 2018-2021.

Los resultados obtenidos por el modelo presentan un F1-Score de 0.95. Este modelo permitió clasificar un conjunto de tweets para realizar un análisis geoespacial en el período antes mencionado.

## ***Dedicatorias***

*A mi Familia por todo el apoyo.*

*A mis padres: María D. Cen. Ay y José I. Pool U. (q.e.p.d), por enseñarme sobre la disciplina para lograr los sueños.*

*A mi esposa Marisol por apoyarme en este sueño y en diversos momentos complicados siempre estuvo para motivarme. Lo logramos.*

*A mis hijas por escuchar y comprender, han sido un apoyo importante en este propósito.*

*A mis hermanos de quienes aprendí sobre la responsabilidad y afrontar los compromisos.*

*Una parte importante de esta experiencia han sido Don Manuel y Doña Reyni, gracias por la paciencia.*

*A mis amigos con quienes he tenido la oportunidad de compartir un café y contar con su apoyo.*

## ***Agradecimientos***

*A CENTRO GEO por la beca otorgada para realizar los estudios de maestría.*

*A mis tutores el Dr. Hugo Carlos y el Dr. Gandhi Hernández, por su tiempo y recomendaciones en todo momento que me guiaron para lograr este trabajo.*

*A los profesores de las diversas asignaturas de los cuales aprendí diversos temas durante la maestría.*

*A los compañeros del seminario de estudios dirigidos sobre de donde surge el presente trabajo.*



# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Planteamiento del problema . . . . .	1
1.2. Objetivos . . . . .	2
1.2.1. Objetivo General . . . . .	2
1.2.2. Objetivos específicos . . . . .	3
1.3. Preguntas de investigación . . . . .	3
1.4. Justificación . . . . .	3
1.5. Estudios relacionados . . . . .	4
<b>2. Marco Teórico</b>	<b>7</b>
2.1. Aprendizaje máquina . . . . .	7
2.1.1. Aprendizaje no supervisado . . . . .	8
2.1.2. Aprendizaje supervisado . . . . .	8
2.2. Algoritmos de clasificación . . . . .	9
2.2.1. Formulación matemática . . . . .	9
2.2.2. Regresión Logística . . . . .	10
2.2.3. Máquinas de Soporte Vectorial (Linear SVM) . . . . .	10
2.2.4. Clasificación mediante Procesos Gaussianos (Gaussssian Process) .	11
2.2.5. Análisis de Discriminante Cuadrático (QDA) . . . . .	12
2.3. Métricas de desempeño para modelos de clasificación . . . . .	13
2.4. Procesamiento de Lenguaje Natural . . . . .	14

2.4.1. Métodos de representación basados en conteo . . . . .	14
2.4.2. Modelado de Secuencias de Texto y Deep Learning . . . . .	15
2.4.3. Destilación del Conocimiento (knowledge distillation) . . . . .	18
2.5. Métodos de reducción de dimensionalidad . . . . .	19
2.5.1. PCA . . . . .	20
2.5.2. IVIS . . . . .	20
<b>3. Metodología</b>	<b>22</b>
3.1. Extracción de datos . . . . .	23
3.2. Preprocesamiento . . . . .	24
3.3. Extracción de características . . . . .	26
3.4. Reducción de dimensionalidad . . . . .	26
3.5. Entrenamiento de los modelos de clasificación . . . . .	29
3.6. Experimentación . . . . .	31
3.6.1. Experimentos con Traducciones . . . . .	32
3.6.2. Experimentos con KD . . . . .	35
<b>4. Resultados</b>	<b>39</b>
4.1. Resultados de los experimentos con traducciones . . . . .	39
4.1.1. Traducciones y PCA . . . . .	40
4.1.2. Traducciones e IVIS Supervisado . . . . .	41
4.1.3. Traducciones e IVIS No Supervisado . . . . .	43
4.1.4. Traducciones e IVIS Semisupervisado . . . . .	44
4.2. Resultados de los experimentos con KD . . . . .	46
4.2.1. KD y PCA . . . . .	46
4.2.2. KD e IVIS Supervisado . . . . .	48
4.2.3. KD e IVIS No Supervisado . . . . .	49
4.2.4. KD e IVIS Semisupervisado . . . . .	50
4.2.5. KD sin reducción de dimensionalidad . . . . .	52
4.3. Caso de estudio: Depresión en México . . . . .	53



4.3.1. Clasificación de tweets . . . . .	53
4.3.2. Análisis espacio-temporal . . . . .	55
4.3.3. Distribución Mensual por Entidad . . . . .	56
4.3.4. Correlación con datos de salud mental reportados por INEGI . . . . .	59
4.3.5. Distribución de palabras . . . . .	61
<b>5. Conclusiones y trabajo a futuro</b>	<b>63</b>
5.1. Conclusiones . . . . .	63
5.2. Trabajo a futuro . . . . .	65

# Capítulo 1

## Introducción

En este capítulo se presenta un panorama general del proyecto de investigación realizado. Se incluye el planteamiento del problema, el objetivo general, los objetivos específicos y la pregunta de investigación.

### 1.1. Planteamiento del problema

Cada año cerca de 703,000 personas se suicidan. Diversas investigaciones han documentado los factores de riesgo: depresión, ansiedad, estrés, entre otros (OMS, 2022).

El estudio realizado por la Organización Mundial de la Salud (OMS), muestra que la depresión afecta al 3,8% de la población mundial, de los cuales el 10,7% son población adulta. Una persona con depresión puede manifestar síntomas diversos como la pérdida de interés en actividades, baja concentración, ansiedad, infelicidad y en caso de no recibir tratamiento, puede llegar a tener pensamientos de automutilación o suicidio (OMS, 2021).

Se estima que en América Latina casi 16 millones de adolescentes padecen algún tipo de trastorno como la depresión, ansiedad, bipolaridad y autismo TDAH (Unicef, 2021).

En México, el Instituto Nacional de Estadística y Geografía (INEGI), presentó en el año 2021, los resultados de la encuesta Nacional de Bienestar (INEGI, 2021). Dicha encuesta fue aplicada a 84,4 millones de personas, de las cuales 39,8 millones son hombres y

44,6 millones mujeres; entre los hallazgos se encontró que el 15,4% de los encuestados presentan síntomas de depresión.

La detección de enfermedades mentales utilizando aprendizaje máquina ha despertado el interés de investigadores del área de ciencias computacionales, debido a que las tecnologías actuales permiten generar y obtener datos de forma masiva, lo que facilita la construcción de modelos de clasificación (Sang Eun Woo, 2020).

El *bigdata* ofrece la posibilidad de desarrollar investigación en el área de salud mental utilizando como fuente de datos los textos provenientes de redes sociales. Una de las redes sociales más utilizada para este propósito es Twitter, debido a sus políticas de acceso a datos (Govindasamy & Palanichamy, 2021). En este mismo sentido, las investigaciones realizadas en materia de salud mental han utilizado en su mayoría datos en idioma inglés; esto restringe la posibilidad de detectar problemas de salud mental en otros idiomas. Abordar este problema es importante porque los padecimientos de salud mental afectan a la población mundial. Debido a lo anterior, el problema abordado en este trabajo consiste en detectar tweets relacionados con la depresión en idioma español mexicano. Para ello se utilizaran herramientas de Procesamiento de Lenguaje Natural y Aprendizaje Máquina.

## **1.2. Objetivos**

### **1.2.1. Objetivo General**

Construir un modelo de clasificación para detectar tweets relacionados a la depresión en el idioma español mexicano empleando técnicas de Procesamiento de Lenguaje Natural.

### **1.2.2. Objetivos específicos**

Para el desarrollo de la presente investigación se proponen los siguientes objetivos específicos:

- Extraer tweets en español para el período 2018-2021.
- Construir el modelo de clasificación.
- Realizar un análisis espacio-temporal del contenido de los tweets.
- Realizar un análisis semántico del contenido de los tweets.

### **1.3. Preguntas de investigación**

En el presente estudio se plantean las siguientes preguntas de investigación:

1. ¿Es posible construir un modelo que detecte tweets relacionados con la depresión en español mexicano a partir de datos en inglés?
2. ¿Cuál es la mejor estrategia para construir dicho modelo?
3. ¿Es posible modelar el comportamiento de la depresión en México a partir de su manifestación en Twitter?
4. ¿Es posible la identificación de patrones temporales a partir de su manifestación en Twitter?

### **1.4. Justificación**

La Organización Mundial de la Salud (OMS), ha desarrollado el Plan de Acción Integral sobre la Salud Mental 2013-2030, en este plan se incluyen acciones para la detección temprana de enfermedades de salud mental, como es el caso de la depresión y la ideación suicida. El uso de herramientas computacionales podría ayudar en este objetivo.

Para México, el INEGI reporta un incremento en las tasas de suicidio del 4,2% en personas entre 15 y 29 años, lo cual lo convierte en uno de los problemas de salud prioritarios para el país. Debido a que la depresión es un factor de riesgo del suicidio, su detección oportuna es un componente importante en la solución del problema.

La pandemia ocasionada por el Covid-19 propició un cambio en el estilo de vida de las personas. Diversos estudios han mostrado los efectos de la pandemia en la salud mental (OMS, 2022). Lo que acentúa la importancia de este trabajo para estudiar patrones de comportamiento asociados con la depresión.

Por otra parte, existen pocos estudios que utilicen la información espacio-temporal extraída de redes sociales, sin embargo este componente permite analizar el comportamiento de la depresión en una dimensión geográfica. Por ejemplo, emplear análisis geoespacial para modelar la depresión en migrantes mexicanos puede ser de utilidad para diseñar medidas preventivas.

Debido a que la depresión es un problema de salud mental universal, contar con un modelo multilingual amplia la capacidad de analizar conjuntamente la información relacionada con la depresión. Retomando el problema de la depresión asociado a los migrantes mexicanos, contar con una herramienta multilingual permitiría abordar el problema debido a que los migrantes suelen comunicarse en ambos idiomas (inglés y español). En este sentido, la importancia de esta tesis radica en que el español es hablado por alrededor de 580 millones de personas en el mundo, incluyendo 126 millones en México, en donde la migración hacia Estados Unidos es común.

En la siguiente sección se presentan estudios relacionados con la presente investigación.

## **1.5. Estudios relacionados**

Los métodos computacionales con NLP, han sido utilizados para generar modelos de detección para problemas de salud mental con textos provenientes de redes sociales, transcripciones médicas y notas clínicas.

El estudio realizado por Zhang et al. (2022) presenta los resultados de un análisis sobre los métodos aplicadas al área de salud mental. La Figura 1.1 muestra los resultados considerando cuatro criterios: método, enfermedad mental, lenguaje y fuente de datos aplicados a la depresión y la ideación del suicidio. Los hallazgos muestran que la depresión es el principal padecimiento estudiado con métodos de aprendizaje maquina. Con respecto al criterio del idioma de la fuente de datos, el 81 % de los estudios analizados utilizan una fuente de datos en idioma inglés. La fuente de datos utilizada en los estudios son principalmente obtenidos de twitter y reddit.

Es notorio los trabajos orientados al idioma inglés, sin embargo el fenómeno de migración abre la necesidad de estudios relacionados en idioma español o multilingual.

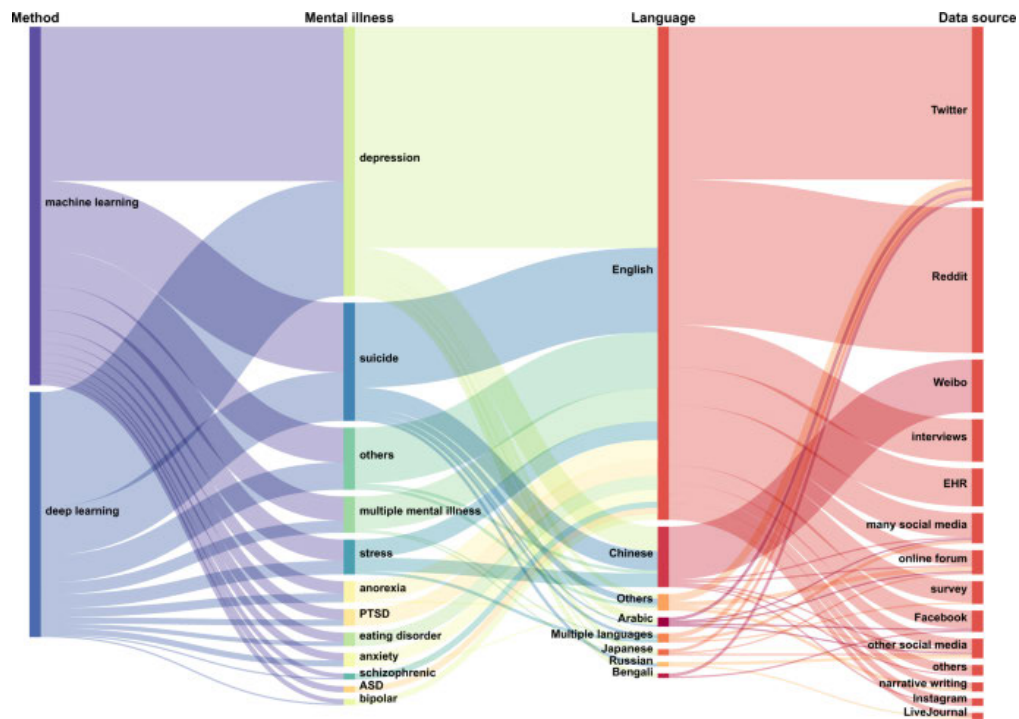


Figura 1.1: Métodos utilizados en el área de salud mental. Fuente: (Zhang et al., 2022)

El estudio presentado por Thieme et al. (2020) muestra el uso de algoritmos de clasificación aplicados a la depresión para realizar análisis de sentimientos, modelado de tópicos, agrupación con métodos no supervisados con fuentes de datos en formato de audio, notas clínicas, sensores, imágenes, vídeo y encuestas.

En el estudio realizado por Islam et al. (2018) aborda la detección de depresión a partir de comentarios de usuarios de la red social facebook. El análisis de los comentarios fue realizado con el software LIWC (Linguistic Inquiry and Word Count ) para la extracción de características. El análisis de los comentarios consistió la extracción de características sobre el proceso emocional, el estilo de escritura y la emoción como un proceso temporal.

La detección de procesos emocionales consistió en el uso de un vocabulario de emociones positivas y emociones negativas. Algunas de las palabras relacionadas con las emociones positivas son: *happy, love, nice, etcétera*. Algunas palabras relacionadas con las emociones negativa son: *worthless, loser, hurt, ugly, sad, etcétera*. La característica de temporalidad, consiste en el uso de palabras que denotan el pasado, presente o futuro de un texto, adicionalmente el uso de palabras como: *today, now, ago*. El estilo de escritura consistió en el análisis del uso de pronombres, pronombres personales, artículos, verbos, verbos auxiliares en un texto. Las características mencionadas anteriormente fueron utilizadas para generar modelos de clasificación mediante el uso de algoritmos como Decision Tree, KNN, SVM y otro basado en el uso de ensamble de clasificadores. Los resultados de las métricas de los modelos de clasificación se encuentran entre un 60 % y 80 %.

En el siguiente capítulo se presentan los algoritmos y los métodos de NLP utilizados en la presente investigación.

# Capítulo 2

## Marco Teórico

En este capítulo se describen los elementos teóricos sobre los cuales se desarrolla esta investigación.

### 2.1. Aprendizaje máquina

Los algoritmos de aprendizaje máquina son definidos como un conjunto de métodos que pueden detectar patrones en datos y utilizarlos para inferir información o tomar decisiones bajo incertidumbre (Murphy, 2022). Los datos son la parte fundamental de los algoritmos de aprendizaje máquina, en inglés Machine Learning (ML). En la era del *big-data* los datos pueden encontrarse y extraerse en formatos como audio, vídeo, imágenes y texto.

Cuando los datos se encuentran en formato texto, la comunidad de ciencia de datos denomina su procesamiento como: minería de textos, analítica de textos o aprendizaje máquina para texto. La fusión de las áreas de ciencia de datos y el Procesamiento de Lenguaje Natural (NLP) se ha aplicado en la resolución de tareas de clasificación, análisis de sentimientos, análisis de discurso, extracción y resumen de textos, por mencionar algunas. Lo anterior ha requerido del desarrollo de métodos para la representación de textos en vectores (vectorización) a partir de la extracción de sus características.



Una vez extraídas las características es posible entrenar algoritmos que, de manera automática, realicen tareas de clasificación (aprendizaje supervisado) o agrupación de textos (aprendizaje no supervisado). Esto se describe en la siguiente sección.

### 2.1.1. Aprendizaje no supervisado

El aprendizaje no supervisado consiste en algoritmos que basan su proceso de entrenamiento en datos sin etiquetas y tienen como propósito resolver tareas de agrupamiento o también llamado *clustering* tomando en consideración las características comunes de los datos. Según Murphy (2022), estos algoritmos se pueden clasificar en: basados en centroides, basados en densidad y basados en jerarquía.

Entre los algoritmos basados en centroides, se encuentra el algoritmo conocido como *k-means*. En los algoritmos basados en densidad, se encuentra el modelo de mezclas gaussianas (*model mixture gaussian*). En el caso de los algoritmos basados en jerarquía se encuentra el denominado DBScan (Murphy, 2022).

### 2.1.2. Aprendizaje supervisado

El aprendizaje supervisado tiene como propósito aprender a predecir un valor continuo o discreto a partir de un conjunto de variables independientes llamadas características o *features*. Según Murphy (2022) cuando se predice una variable continua se le denomina problema de regresión, mientras que en el caso de predecir un valor discreto se le denomina clasificación.

Según Murphy un modelo de clasificación o predicción puede representarse como:

$$f : \mathbb{R}^n \rightarrow \mathbb{R} \quad (2.1)$$

Donde  $f$  es una función que recibe un vector de características  $\mathbf{x}$  con dimensión  $n$  y genera como salida un escalar.

La construcción de la función  $f$  puede ser a través de diversos algoritmos como regresión logística, máquinas de soporte vectorial(SVM), análisis de discriminante(QDA), entre otros existentes en la literatura.

En la siguiente sección se presentan los algoritmos utilizados en la presente investigación para la construcción de modelos de clasificación.

## 2.2. Algoritmos de clasificación

Los algoritmos para la construcción de modelos de clasificación requieren de un conjunto de datos distribuidos en clases. Cuando los datos se encuentran distribuidos en más de dos clases, se dice que el modelo es multiclase, en caso contrario se dice que el modelo de clasificación es binario.

Según Aggarwal (2022) en un modelo de clasificación binario, el conjunto de posibles valores de salida puede ser denotada como  $\{+1, -1\}$ , la cual se puede representar de la forma:

$$f : \mathbb{R}^n \rightarrow \{+1, -1\} \quad (2.2)$$

Un modelo de clasificación es generado a partir de dos etapas: entrenamiento y prueba (Aggarwal, 2022). En la etapa de entrenamiento o aprendizaje el modelo de clasificación recibe un conjunto de datos etiquetados, mientras que en la etapa de prueba el modelo recibe un conjunto de datos no etiquetados que no fueron utilizados en la etapa de entrenamiento y para los cuales predice una etiqueta. El propósito de la etapa de prueba es conocer el desempeño del modelo a partir de sus métricas.

### 2.2.1. Formulación matemática

Sea  $\{\mathbf{x}_i, y_i\}_{i=1,2,\dots,m}$  el conjunto de observaciones, donde  $\mathbf{x} \in \mathbb{R}^n$  es el vector de características y  $y \in \mathbb{R}$  corresponde a la etiqueta, deseamos encontrar una función:

$$\hat{y} = f(\mathbf{x}) \quad (2.3)$$

donde  $\hat{y}$  corresponde a la predicción del modelo. Cada algoritmo de clasificación difiere en el tipo de función que asume el modelo (e.g. lineal) o en los supuestos del problema (e.g. si es un esquema probabilístico). Las siguientes secciones explican de forma general los algoritmos de clasificación considerando la formulación anterior.

## 2.2.2. Regresión Logística

La regresión logística pertenece a una clase de modelos probabilísticos denominados modelos discriminativos. El modelo está definido como:

$$p(y|\mathbf{x}; \mathbf{w}) = \sigma(\theta) = \sigma(\mathbf{w}^T \mathbf{x}), \quad (2.4)$$

donde  $\mathbf{w} \in \mathbb{R}^n$  es el vector de parámetros y  $\sigma(\cdot)$  corresponde a la función sigmoide. Para estimar el vector  $\mathbf{w}$  en una regresión logística binaria definimos la siguiente función de costo:

$$\min_{\mathbf{w}} J(\mathbf{w}) = -\frac{1}{m} \sum_{i=1}^m [y_i \log(\sigma(\theta)) + (1 - y_i) \log(1 - \sigma(\theta))]. \quad (2.5)$$

En práctica  $J(\mathbf{w})$  se minimiza empleando el descenso de gradiente estocástico (Murphy, 2022).

## 2.2.3. Máquinas de Soporte Vectorial (Linear SVM)

Máquinas de Soporte Vectorial (SVM) es otro algoritmo de clasificación para datos linealmente separables. Es un clasificador binario de la forma:

$$y = h(\mathbf{x}) = \text{sign}(f(\mathbf{x})),$$

donde el *límite de decisión* está dado por  $f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0$ . En SVM se asume que los datos se encuentran etiquetados con  $y \in \{+1, -1\}$  (Murphy, 2022).

Este algoritmo tiene como función de costo:

$$\begin{aligned}
& \underset{\mathbf{w}, w_0}{\text{minimize}} && g(\mathbf{w}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} \\
& \text{subject to} && y_i(\mathbf{w}^T \mathbf{x}_i + w_0) \geq 1, \quad i = 1, \dots, m
\end{aligned} \tag{2.6}$$

La restricción asegura que todos los puntos están del lado correcto del límite de decisión a una distancia de al menos una  $\epsilon$ .

En la Figura 2.1 se observa la representación de la distancia entre el punto más cercano de una clase y el límite de decisión, representado por la línea en color rojo.

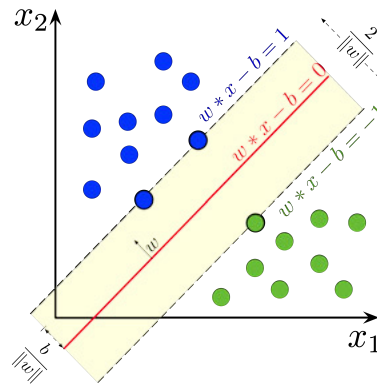


Figura 2.1: Linear SVM. Fuente: Wikipedia

## 2.2.4. Clasificación mediante Procesos Gaussianos (Gaussian Process)

Un proceso Gaussiano para una tarea de clasificación binario consiste en obtener una función continua a priori  $f(\mathbf{x})$  y estimar la probabilidad de  $f(\mathbf{x})$  con la función sigmoide, representada como  $\pi(\mathbf{x}) = p(y = +1|\mathbf{x}) = \sigma(f(\mathbf{x}))$ .

La función  $f(\mathbf{x})$  es una distribución multivariada definida de la forma  $f(\mathbf{x}) \sim \mathcal{N}(\mu(\mathbf{x}), \Sigma)$ .

Para estimar  $\mathbf{f}$  se utiliza la regla de Bayes mediante  $p(\mathbf{f}|\mathbf{x}, \mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{f})p(\mathbf{f}|\mathbf{x})}{p(\mathbf{y}|\mathbf{x})}$ . De acuerdo con (Rasmussen y Williams, 2005) para estimar  $\mathbf{f}$ , se maximiza la función:

$$\Psi(\mathbf{f}) = \log p(\mathbf{y}|\mathbf{f}) - \frac{1}{2} \mathbf{f}^T \Sigma^{-1} \mathbf{f} - \frac{1}{2} \det \Sigma - \frac{n}{2} \log(2\pi) \tag{2.7}$$

## 2.2.5. Análisis de Discriminante Cuadrático (QDA)

El análisis de discriminante Gaussiano es un algoritmo para construir modelos de clasificación generativos mediante la estimación de  $p(\mathbf{x}|y = c; \theta)$ . En este algoritmo tiene como un supuesto que  $p(\mathbf{x}|y)$  es una distribución gaussiana multivariada. Para un clasificador binario se tienen los siguientes supuestos:

$$p(y) = \phi^y(1 - \phi)^{1-y} \quad (2.8)$$

$$p(\mathbf{x}|y = 0) = \frac{1}{(2\pi)^n \det|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_0)\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}_0)\right) \quad (2.9)$$

$$p(\mathbf{x}|y = 1) = \frac{1}{(2\pi)^n \det|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_1)\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}_1)\right) \quad (2.10)$$

Los parámetros del modelo son  $\phi$ ,  $\Sigma$ ,  $\boldsymbol{\mu}_0$ ,  $\boldsymbol{\mu}_1$ .  $\Sigma$  es la matriz de covarianza. Cuando se asume que la matriz de covarianza  $\Sigma$  es diferente para cada clase, el algoritmo se conoce como análisis de discriminante cuadrático (por sus siglas en inglés QDA).

Para estimar los parámetros, definimos  $m_0$  y  $m_1$  como la cantidad de datos en la clase 0 y 1 respectivamente y los parámetros se obtienen por medio de:

$$\phi = \frac{m_0}{m} \quad (2.11)$$

$$\boldsymbol{\mu}_0 = \frac{1}{m_0} \sum_{i:y_i=0} \mathbf{x}_i \quad (2.12)$$

$$\boldsymbol{\mu}_1 = \frac{1}{m_1} \sum_{i:y_i=1} \mathbf{x}_i \quad (2.13)$$

$$\Sigma_0 = \frac{1}{m_1} \sum_{i:y_i=0} (\mathbf{x}_i - \boldsymbol{\mu}_0)(\mathbf{x}_i - \boldsymbol{\mu}_0)^T \quad (2.14)$$

$$\Sigma_1 = \frac{1}{m_1} \sum_{i:y_i=1} (\mathbf{x}_i - \boldsymbol{\mu}_1)(\mathbf{x}_i - \boldsymbol{\mu}_1)^T \quad (2.15)$$

Posterior a la generación del modelo de clasificación (etapa de entrenamiento), la segunda etapa es la prueba del modelo y obtener sus métricas de desempeño.

En la siguiente sección se explican las métricas utilizadas en la presente investigación.

## 2.3. Métricas de desempeño para modelos de clasificación

En la etapa de prueba de un modelo de clasificación, el modelo recibe un conjunto de datos que no se utilizaron en la etapa de entrenamiento. Sin embargo, las etiquetas de estos datos son conocidas. En esta etapa se pueden estimar diversas métricas para evaluar el desempeño de un clasificador binario. Según (Hackeling, 2014) las métricas utilizadas en un clasificador binario son: *accuracy*, *precision*, *recall*, *F1 measure*. Estas métricas dependen de los conceptos de verdaderos positivos (TP), verdaderos negativos (TN), falsos positivos (FP) y falsos negativos (FN). Cuando se habla de positivos y negativos hace referencia a la clase. La palabra verdadero y falso hace referencia a la comparación entre la clase resultante de la predicción y la clase del dato.

- Exactitud (*accuracy*). Mide la fracción de predicciones correctas del clasificador. Esta métrica no distingue de los errores de un falso positivo y un falso negativo.
- Precisión (*Precision*). La precisión es la fracción de predicciones positivas que son correctas. Se obtiene por medio de  $P = \frac{TP}{TP+FP}$ .
- Recall. Es la fracción de verdaderos positivos definido como:  $R = \frac{VP}{VP+FN}$ .
- F1 Score. Definida como  $F_1 = 2 \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$  es una medida de desempeño que combina *Precision* y *Recall* que suele utilizarse para seleccionar al mejor modelo.

En las secciones anteriores se mencionó que los modelos de clasificación reciben como entrada un vector de características  $\mathbf{x}$ ; para generar dicho vector en textos se requiere

de técnicas de NLP. En la siguiente sección se explican las técnicas de NLP utilizadas en esta investigación.

## **2.4. Procesamiento de Lenguaje Natural**

El Procesamiento de Lenguaje Natural (NLP), es una de las áreas de la IA que se refiere al conjunto de técnicas que involucran la aplicación de métodos estadísticos para modelar características de un lenguaje; y llevar a cabo tareas como: traducciones, resumen de textos, extracción de palabras claves, etcétera. La idea es generar representaciones vectoriales de los textos, que pueden ser discretas o continuas (Rao & McMahan, 2019).

Entre los métodos de representación de textos han surgido algoritmos basados en conteo, funciones de distribución y en los últimos años a través de redes neuronales.

### **2.4.1. Métodos de representación basados en conteo**

En NLP existen métodos para la representación de textos basados en conteo. Uno de estos métodos es Frecuencia de Palabras - Frecuencia Inversa de Documentos (TF-IDF). Este método está basado en una normalización de las frecuencias de las palabras en el texto (Jurafsky & Martin, 2009). Una de las ventajas de este algoritmo es que tienen un bajo costo computacional. Por otra parte, la desventaja de este tipo de métodos es que no capturan contexto de una manera eficiente.

En un problema de clasificación donde el contexto de una palabra tiene importancia TF-IDF puede no ser útil, por lo cual se sugiere utilizar otros métodos que capturen el contexto de la palabra. Una alternativa son los métodos de representación empleando distribuciones continuas que se describen en la siguiente sección introduciendo los conceptos básicos.

## 2.4.2. Modelado de Secuencias de Texto y Deep Learning

El orden de las palabras en una secuencia de textos contiene información semántica que no puede ser inferida a través de la representación de textos basados en conteo. Una forma para extraer la información semántica de una palabra en una secuencia de textos es mediante algoritmos que utilizan redes neuronales artificiales (Aggarwal, 2022).

El algoritmo *word2vec* (Mikolov et al., 2013) consiste de una red neuronal artificial para generar la representación de la palabra en un contexto dado. La representación obtenida con *word2vec* se le denomina *embedding*. Según Rao y McMahan (2019), el *embedding* es la representación vectorial de una palabra utilizando distribuciones continuas en baja dimensionalidad. Esto contribuye a modelar diversos aspectos de una palabra con la idea de que dos palabras que tienen posiciones cercanas en el espacio construido tienen significados similares, como se aprecia en la Figura 2.2. La importancia de los *embeddings* radica en su capacidad para predecir una palabra a partir de un contexto y viceversa.

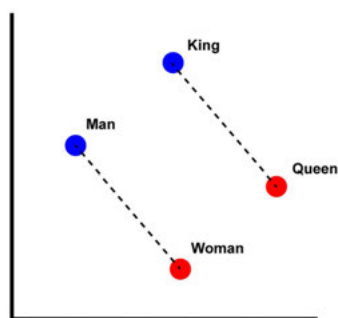


Figura 2.2: Embedding de palabras. Fuente: Wikipedia

*Word2vec* genera *embeddings* a nivel de palabra, sin embargo los avances en el área de NLP han permitido el desarrollo de algoritmos para generar *embeddings* a nivel de sentencias y documentos, estos *embeddings* pueden ser aplicados a la solución de problemas como la traducción automática de textos.



En años recientes, para la tarea de traducción automática de textos, se ha tenido uno de los avances más significativos con el uso de mecanismos de atención (transformers) propuesto por Vaswani et al. (2017). El mecanismo de atención aplicado a la traducción de textos requiere como entrada un texto, representado como *embeddings* y genera como salida la traducción correspondiente. En la Figura 2.3 se muestra la arquitectura del mecanismo de atención.

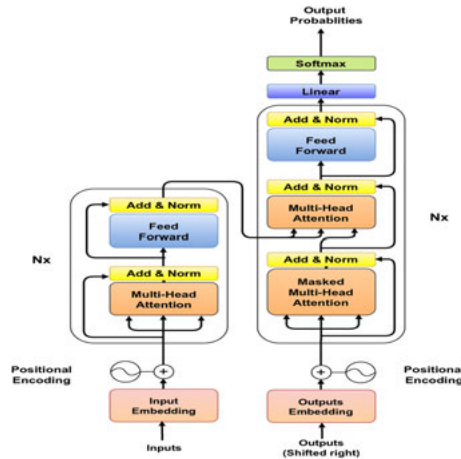


Figura 2.3: Arquitectura del modelo de atención. Fuente: Aurpa et al., 2022

La arquitectura de los *transformers* consiste de un codificador-decodificador (*encoder-decoder*). El codificador recibe un conjunto de palabras denotadas como  $x$  para generar una representación vectorial de la secuencia de texto denotada como  $z$ . El decodificador toma la entrada  $z$  y genera una secuencia de salida  $y$ .

Los *transformers* utilizan el mecanismo de atención. Este mecanismo aplicado a NLP tiene como propósito asignar contexto a la palabra en toda la secuencia de entrada y cuantificar la relación del uso de una palabra con otras palabras. En la Figura 2.4 se ejemplifica el proceso del mecanismo de atención.

La matriz de atención se obtiene por medio de  $Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$ , donde  $Q$ ,  $K$  y  $V$  representan al *Query*, *Keys* y *Values* respectivamente. La Figura 2.5 corresponde a la visualización de la matriz de atención. En este caso las palabras que tienen menor relación se encuentran en color oscuro mientras que las de mayor relación

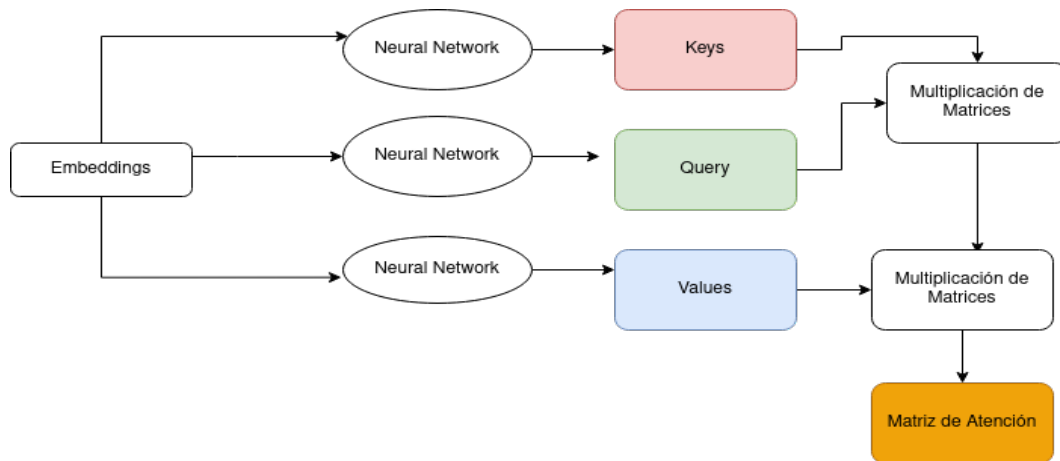


Figura 2.4: Mecanismo de atención

se encuentran en color blanco.

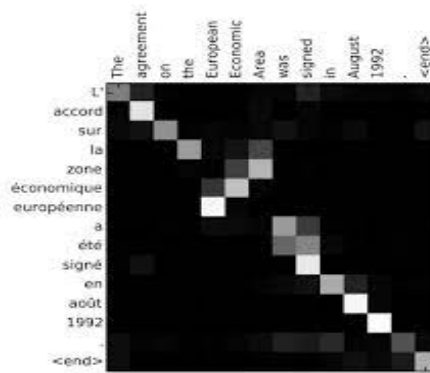


Figura 2.5: Visualización de la matriz de atención. Fuente: Bahdanau et al., 2015

BERT (Devlin et al., 2019) es un modelo para NLP para el idioma inglés, el cual implementa el mecanismo de atención. Por otra parte han surgido modelos para el idioma español. Un ejemplo es BETO (Cañete et al., 2020), el cual es la versión en español del modelo BERT. Este modelo genera representaciones vectoriales de dimensión 768. La arquitectura consiste de 110 millones de parámetros y un vocabulario de 31002 palabras.

La construcción de modelos para NLP como BERT requirió de grandes recursos computacionales. Técnicas recientes para la transferencia de conocimiento pretenden construir modelos ligeros con menor costo computacional.

En la siguiente sección se presenta la técnica destilación del conocimiento.

### 2.4.3. Destilación del Conocimiento (knowledge distillation)

La generación de modelos de redes neuronales (NN) pueden necesitar altos recursos computacionales, debido a esto se han desarrollado técnicas para construir modelos de NN ligeros a través de la transferencia de conocimiento a partir de grandes modelos de NN (Gou et al., 2021).

Una técnica para la transferencia de conocimiento es llamada *knowledge distillation* (KD), la cual se compone de un algoritmo y la arquitectura *teacher-student*. La Figura 2.6 muestra el esquema genérico del KD (Gou et al., 2021).

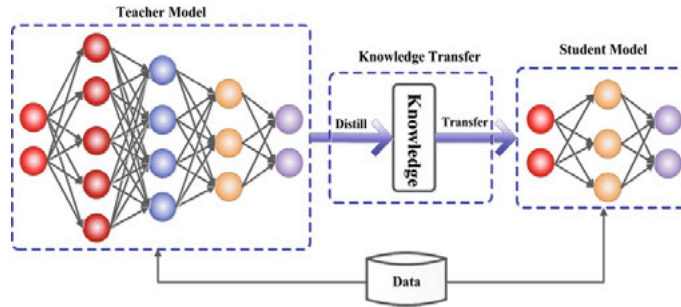


Figura 2.6: Framework genérico de KD. Fuente: Gou et al., 2021

Reimers y Gurevych (2020) utilizaron la técnica KD generar el modelo multilingual *distiluse-base-multilingual-cased-v1* (*student*) a partir del modelo SBERT (*teacher*). A partir del modelo *Teacher* representado por  $M$  para un lenguaje  $s$  y un conjunto paralelo de sentencias traducidas denotado como  $((s_1, t_1), \dots, (s_n, t_n))$  ( $s_i$  puede estar en diferentes lenguajes), el componente *Student* denotado como  $\hat{M}$  es entrenado como  $\hat{M}(s_i) \approx M(s_i)$  y  $\hat{M}(t_i) \approx M(s_i)$ .

La función costo es definida como:

$$\frac{1}{|\beta|} \sum_{j \in \beta} [(M(s_j) - \hat{M}(s_j))^2 + (M(s_j) - \hat{M}(t_j))^2] \quad (2.16)$$

En la primera parte de la ecuación  $(M(s_j) - \hat{M}(s_j))^2$  el modelo *Student* aprende sobre la proyección de la sentencia en el mismo espacio vectorial del lenguaje fuente.

La segunda parte de la función de costo denotado como  $(M(s_j) - \hat{M}(t_j))^2$ , reduce el

error entre la representación vectorial de la sentencia del modelo *Teacher* y la representación de la sentencia traducida con el modelo *Student*.

El modelo *distiluse-base-multilingual-cased-v1* (KDMV1) permite extraer características de textos longitud máxima de 128 palabras y generar una representación con dimensión 512. KDMV1 tiene soporte para 15 idiomas entre ellos el inglés y el español (Reimers y Gurevych, 2020).

A partir de los modelos multilingües es posible obtener la representación de un texto en diversos idiomas. Estas representaciones corresponden a vectores de alta dimensionalidad y en muchos casos son vectores esparsos. Ante esta situación existen métodos para obtener vectores de baja dimensionalidad que no sean esparsos.

En la siguiente sección se presentan dos métodos de reducción de dimensionalidad utilizados en la presente investigación.

## 2.5. Métodos de reducción de dimensionalidad

Como se mencionó anteriormente, las representaciones vectoriales de textos tienen dos características importantes: i) su alta dimensionalidad y ii) su esparcidad. Estas características incrementan el costo computacional y tienen un fuerte impacto en los algoritmos de clasificación. Para atenuar el problema suelen emplearse métodos de *reducción de dimensionalidad*. La idea detrás de estos métodos es sencilla, dada una representación vectorial  $\mathbf{x} \in \mathbb{R}^N$  se desea encontrar una función  $f(\mathbf{x}) : \mathbb{R}^N \rightarrow \mathbb{R}^D$  con  $D < N$ . Cada método de reducción de dimensionalidad varía en las características que asigna a  $f$  (e.g. si es lineal) y en los atributos que se desean conservar del espacio original (e.g. su variabilidad).

En general los algoritmos de reducción de dimensionalidad son no supervisados; sin embargo, existen formulaciones supervisadas que permiten incorporar información sobre el *tipo* de dato. En la práctica esto equivale a construir una función  $f(\mathbf{x}, y)$  siendo  $y$  la etiqueta asignada a la observación. La ventaja de estas formulaciones es que permiten extraer de forma más precisa las propiedades de los datos. Otra variante es la que se

conoce como estrategia semi-supervisada, una suerte de híbrido entre en los esquemas no supervisado y supervisado, este esquema asume que se conocen parcialmente las etiquetas de los datos. El objetivo de los modelos semi-supervisados es extraer propiedades generales de las observaciones a la vez que se intentan abstraer propiedades particulares dependiendo del tipo de dato.

En la literatura existen diversos algoritmos de reducción de dimensionalidad como PCA, LDA, t-SNE, IVIS e ISOMAP. Sin embargo, de acuerdo a los resultados presentados en Szubert y Drozdov (2019) IVIS tiene un mejor desempeño; debido a esto es el algoritmo empleado en este trabajo. Por otro lado, con fines únicamente comparativos se incluyeron los resultados obtenidos a través del algoritmo conocido como Análisis de Componentes Principales (PCA por sus siglas en inglés). En las siguientes secciones se describe de forma general cada uno de los algoritmos.

### 2.5.1. PCA

El análisis de componentes principales es un método de reducción de dimensionalidad *lineal* basado en la *variabilidad de los datos* (Aggarwal, 2022). La idea es proyectar los datos en un espacio  $\mathbb{R}^D$  que maximice la varianza de la proyección. Bajo supuestos de gaussianidad en la distribución de los datos, se puede demostrar que las direcciones de proyección corresponden a los eigenvectores asociados a los eigenvalores más grandes de la matriz de covarianza de los datos (Bishop, 2007).

### 2.5.2. IVIS

El algoritmo IVIS (Szubert & Drozdov, 2019) es un método de reducción de dimensionalidad no lineal basado en un modelo de red neuronal con tres métodos de entrenamiento: supervisado, no supervisado y semisupervisado.

La función de costo corresponde a:

$$J(\theta) = \sum_{i=1}^m \max(\|\mathbf{x}_i - \mathbf{x}_p\|^2 - \|\mathbf{x}_i - \mathbf{x}_n\|^2 + \alpha, 0) \quad (2.17)$$

Donde  $x_i, x_p, x_n$  corresponden al *anchor*, una muestra similar al *anchor*, una muestra negativa respectivamente y  $\alpha$  es el margen. La función de costo minimiza la distancia entre el punto de interés y la muestra positiva mientras maximiza la distancia entre el punto de interés y la muestra negativa. En cada punto de interés del conjunto de datos recibe una muestra positiva y negativa de acuerdo con el algoritmo denominado KNN (*k-nearest neighbors*).

En el siguiente capítulo se presenta la metodología construida a partir de los elementos teóricos descritos.

# Capítulo 3

## Metodología

Para resolver el problema de detección de tweets relacionados con la depresión utilizando ML y NLP se propusieron dos estrategias. La primera estrategia se basó en el uso de traducciones, es decir, se utilizó tweets traducidos de inglés a español. La segunda estrategia se basó en el método de destilación de conocimiento (KD) a partir de tweets en inglés.

En la Figura 3.1 se presentan las etapas para la construcción de los modelos de clasificación para las estrategias mencionadas en el párrafo anterior.



Figura 3.1: Metodología para generar los modelos de clasificación

En las siguientes secciones se explican cada una de las etapas presentadas en la Figura 3.1.

### 3.1. Extracción de datos

Los conjuntos de datos fueron extraídos de Kaggle<sup>1</sup> y de la plataforma AGEI<sup>2</sup> desarrollada por el Laboratorio Nacional de Geointeligencia (GeoInt)<sup>3</sup>.

De *Kaggle*, se obtuvo un conjunto de datos etiquetados al cual se denominó  $D_1$  que consiste de 4,648 textos en idioma inglés provenientes de twitter, de los cuales 2,385 pertenecen a la clase depresión  $\{1\}$  y 2,263 pertenecen a la clase no depresión  $\{0\}$ .

Del AGEI se obtuvieron 2,000 tweets en idioma español publicados durante el año 2019. Este conjunto de datos es denominado  $D_2$  y fue etiquetado manualmente por expertos en psicología.

Los conjuntos  $D_1$  y  $D_2$  son utilizados en los experimentos en los esquemas supervisado y no supervisado. A partir de los conjuntos  $D_1$  y  $D_2$ , se construyeron los conjuntos de datos  $D_3$  y  $D_4$  que fueron utilizados en los experimentos con el esquema semisupervisado.

El conjunto  $D_3$  contiene 5,248 tweets y está formado por todos los elementos de  $D_1$  (en inglés) y 600 tweets (en español) obtenidos de manera aleatoria a partir de  $D_2$ . La muestra aleatoria de 600 tweets fue etiquetada con el valor de  $-1$  indicando ausencia de etiqueta. Por lo tanto, los registros en  $D_3$  se encuentran etiquetados con  $\{1, 0, -1\}$ .

El conjunto de datos  $D_4$  es un subconjunto de  $D_2$  el cual contiene 1,400 tweets en español, distribuidos en 50 % para la clase depresión y el otro 50 % para la clase no depresión.

Para generar los modelos de clasificación empleando traducciones,  $D_1$  fue traducido de inglés a español con la API de Google, dando como resultado el conjunto  $T_1$ . Además,

<sup>1</sup><https://www.kaggle.com/general/234873>

<sup>2</sup><http://agei.geoint.mx/>

<sup>3</sup><http://geoint.mx/>



se construyó un conjunto  $T_3$  utilizando todos los elementos de  $T_1$  y los 600 elementos en español contenidos en  $D_3$ ;  $T_3$  se construyó para el esquema semisupervisado.  $T_4$  es equivalente a  $D_4$  y fue utilizado para el esquema semisupervisado.

En las Tablas 3.2 y 3.1 se presenta un resumen de los datos utilizados en los experimentos con traducciones y basados en KD respectivamente.

Tabla 3.1: Conjuntos de datos para experimentos basado en KD

Conjunto	Idioma	Registros	Clase: 1	Clase: 0	Clase: -1
$D_1$	Inglés	4,648	2,385	2,263	0
$D_2$	Español	2,000	1,000	1,000	0
$D_3$	Inglés + Español	5,093	2,685	2,263	600
$D_4$	Español	1,400	700	700	0

Tabla 3.2: Conjuntos de datos para experimentos con traducciones

Conjunto	Idioma	Registros	Clase: 1	Clase: 0	Clase: -1
$T_1$	Español Traducido	4,648	2,385	2,263	0
$T_2$	Español	2,000	1,000	1,000	0
$T_3$	Español Traducido + Español	5,248	2,685	2,563	600
$T_4$	Español	1,400	700	700	0

## 3.2. Preprocesamiento

Esta etapa consistió las siguientes fases:

- Normalizar los textos a minúsculas
- Eliminar espacios en blanco al principio y final de los textos.
- Eliminar registros nulos.

- Eliminar registros duplicados.
- Remover *emojis*.
- Remover hipervínculos.
- Remover menciones.
- Remover signos de puntuación y palabras que contiene el símbolo @ o el símbolo #.
- Eliminar los registros con frases de una sola palabra (longitud = 1).

Al finalizar el preprocesamiento se mantuvieron el 96 % de los datos.

Las Tablas 3.3 y 3.4 presentan los conjuntos resultantes del preprocesamiento.

Tabla 3.3: Datos preprocesados para experimentos con traducciones

Conjunto	Idioma	Registros	Clase: 1	Clase: 0	Clase: -1
$T_1$	Español Traducido	4,451	2,292	2,159	0
$T_2$	Español	1,985	1,000	985	0
$T_3$	Español Traducido + Español	5,046	2,292	2,159	595
$T_4$	Español	1,389	697	692	0

Tabla 3.4: Datos preprocesados para experimentos con KD

Conjunto	Idioma	Registros	Clase: 1	Clase: 0	Clase: -1
$D_1$	Inglés	4,448	2,291	2,157	0
$D_2$	Español	1,985	1,000	985	0
$D_3$	Inglés + Español	5,043	2,291	2,157	595
$D_4$	Español	1,389	697	692	0

### 3.3. Extracción de características

La etapa de extracción de características tiene como propósito generar las representaciones vectoriales de los textos como se describió en el capítulo anterior. En la estrategia basada en traducciones se utilizó el modelo BETO. Para la estrategia basada en KD se utilizó el modelo KDMV1 <sup>4</sup>.

### 3.4. Reducción de dimensionalidad

Dado que las representaciones vectoriales de los textos son de alta dimensionalidad y pueden ser raras o esparsas, existen métodos de reducción de dimensionalidad que contribuyen a compactar la información. En la Figura 3.2 se puede observar la distribución de frecuencias de las representaciones vectoriales de los textos y se puede observar una distribución normal con media 0, lo cual confirma que la representación de los textos en alta dimensionalidad pueden ser esparsos.

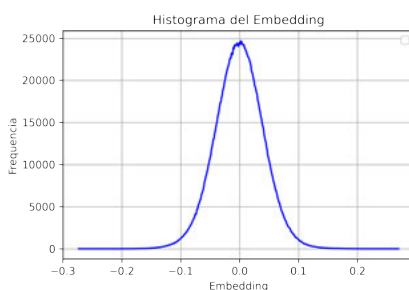


Figura 3.2: Distribución de los embedding

Para prevenir que los vectores sean esparsos, se optó por generar modelos para reducir la dimensionalidad con los algoritmos PCA e IVIS. En la Figura 3.3 se muestra el proceso de las secciones 3.4 y 3.5, este proceso consiste en la introducción de los datos de entrenamiento en los algoritmos de reducción de dimensionalidad para obtener los

---

<sup>4</sup>distiluse-base-multilingual-cased-v1

modelos de reducción, y posteriormente los datos reducidos introducirlos a los diversos algoritmos de ML.

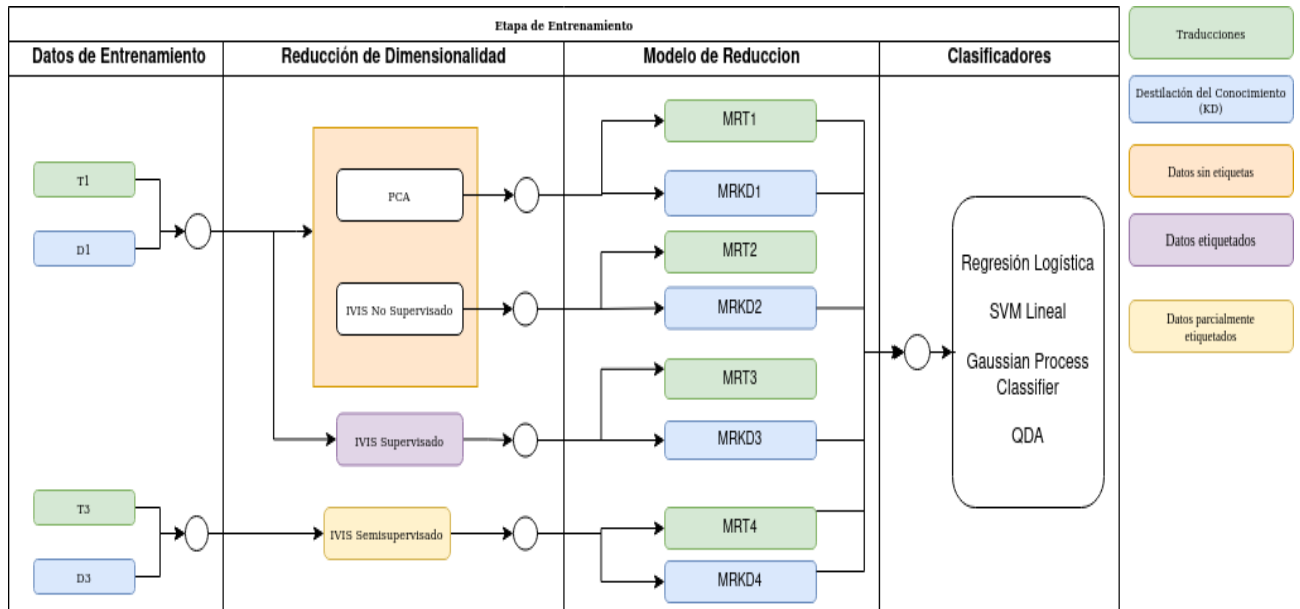


Figura 3.3: Modelos de reducción y entrenamiento de clasificadores

La Tabla 3.5 presenta los datos utilizados para generar los modelos mencionados previamente.

Tabla 3.5: Entrenamiento de los modelos de reducción de dimensionalidad

Estrategia	Conjunto	Idioma	Algoritmo	Entrenamiento	Modelo
Traducciones	$T_1$	Español	PCA	No supervisado	MRT1
			IVIS	No supervisado	MRT2
			IVIS	Supervisado	MRT3
	$T_3$	Español	IVIS	Semisupervisado	MRT4
KD	$D_1$	Inglés	PCA	No supervisado	MRKD1
			IVIS	No supervisado	MRKD2
			IVIS	Supervisado	MRKD3
	$D_3$	Inglés y Español	IVIS	Semisupervisado	MRKD4

Los modelos de reducción de dimensionalidad obtenidos a partir de la Tabla 3.5 se utilizaron para generar los datos de entrenamiento de los clasificadores. La Tabla 3.6 presenta el conjunto de datos y modelo de reducción utilizado para obtener los datos de prueba.

Tabla 3.6: Datos generados empleando los modelos de reducción

Estrategia	Conjunto	Idioma	Modelo de Reducción
			MRT1
Traducciones	$T_2$	Español	MRT2
			MRT3
	$T_4$	Español	MRT4
			MRKD1
KD	$D_2$	Inglés	MRKD2
			MRKD3
	$D_4$	Español	MRKD4

El resultado de la etapa de reducción de dimensionalidad consiste en obtener representaciones en dos dimensiones para el entrenamiento y prueba de los clasificadores.

### 3.5. Entrenamiento de los modelos de clasificación

A partir de las representaciones vectoriales reducidas se generaron modelos de clasificación basados en los algoritmos de Regresión Logística (LR), Máquinas de Soporte Vectorial (Linear SVM), Procesos Gaussianos (GP) y Análisis de Discriminante Cuadrático (QDA). Los hiperparámetros de los modelos se obtuvieron por medio del método *grid search cross-validation*. Además, en el presente estudio se utilizó el framework *scikit-learn*.

En la Figura 3.4 se pueden observar los datos utilizado en la etapa de prueba de los clasificadores.

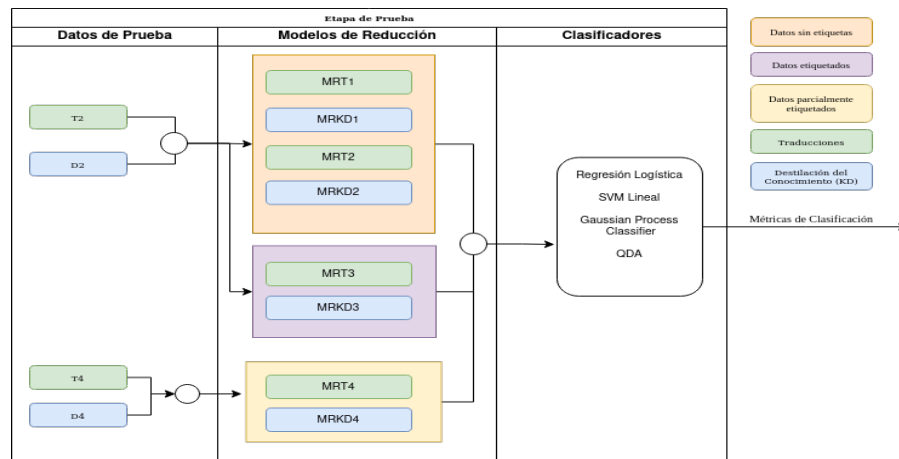


Figura 3.4: Etapa de prueba para los clasificadores

Los clasificadores obtenidos fueron evaluados en la etapa de prueba mediante la generación de las siguientes métricas:

- Exactitud (Accuracy)
- Precisión (Precision)

- Sensibilidad (Recall)
- F1-Score

La Tabla 3.7 presentan los datos utilizados para la etapa de entrenamiento y prueba de los modelos de clasificación.

Tabla 3.7: Entrenamiento y prueba para los clasificadores utilizando traducciones

Algoritmo	Modelo de reducción	Entrenamiento	Prueba
Linear SVM	MRT1	$T_1$	$T_2$
	MRT2	$T_1$	$T_2$
	MRT3	$T_1$	$T_2$
	MRT4	$T_3$	$T_4$
SVM Lineal	MRT1	$T_1$	$T_2$
	MRT2	$T_1$	$T_2$
	MRT3	$T_1$	$T_2$
	MRT4	$T_3$	$T_4$
GP	MRT1	$T_1$	$T_2$
	MRT2	$T_1$	$T_2$
	MRT3	$T_1$	$T_2$
	MRT4	$T_3$	$T_4$
QDA	MRT1	$T_1$	$T_2$
	MRT2	$T_1$	$T_2$
	MRT3	$T_1$	$T_2$
	MRT4	$T_3$	$T_4$

La Tabla 3.8 presentan los datos utilizados para el entrenamiento de clasificadores utilizando KD.

En la siguiente sección se explican los experimentos.

Tabla 3.8: Entrenamiento y prueba de clasificadores con KD

Algoritmo	Modelo de reducción	Entrenamiento	Prueba
Linear SVM	MRKD1	$D_1$	$D_2$
	MRKD2	$D_1$	$D_2$
	MRKD3	$D_1$	$D_2$
	MRKD4	$D_3$	$D_4$
SVM Lineal	MRKD1	$D_1$	$D_2$
	MRKD2	$D_1$	$D_2$
	MRKD3	$D_1$	$D_2$
	MRKD4	$D_3$	$D_4$
GP	MRKD1	$D_1$	$D_2$
	MRKD2	$D_1$	$D_2$
	MRKD3	$D_1$	$D_2$
	MRKD4	$D_3$	$D_4$
QDA	MRKD1	$D_1$	$D_2$
	MRKD2	$D_1$	$D_2$
	MRKD3	$D_1$	$D_2$
	MRKD4	$D_3$	$D_4$

### 3.6. Experimentación

La etapa de experimentación tiene como propósito determinar la mejor estrategia para generar un clasificador de tweets en idioma español considerando las métricas de: *accuracy*, *precision*, *recall* y *F1 Score*.

Tomando como punto de inicio que los textos de los tweets han sido preprocesados, los experimentos consisten de los siguientes pasos:

1. Obtener la representación vectorial de los datos de entrenamiento y prueba.
2. Entrenar los modelos de reducción de dimensionalidad.



3. Entrenar los modelo de clasificación con los algoritmos presentados en la Tabla 3.7.
4. Reducir las dimensiones de los datos de prueba con el modelo de reducción obtenido en el paso 3.
5. Con los datos de prueba obtener las métricas de los clasificadores obtenidos en el paso 4.

En las siguientes secciones se explican los experimentos para el esquema basado en traducciones y el esquema basado en destilación del conocimiento.

### **3.6.1. Experimentos con Traducciones**

En los experimentos con traducciones se utilizaron los conjuntos  $\{T_1, T_2, T_3, T_4\}$  y las representaciones vectoriales de los textos se obtuvieron con el algoritmo BETO. Se realizaron cuatro experimentos variando el uso de los algoritmos de reducción de dimensionalidad PCA e IVIS. En las siguientes secciones se explican cada uno de los experimentos realizados.

#### **Experimento con Traducciones y PCA**

Para este experimento se utilizaron  $T_1$  y  $T_2$  como datos de entrenamiento y prueba respectivamente. Se obtuvieron con BETO las representaciones vectoriales de los tweets contenidos en  $T_1$  y  $T_2$ . Con las representaciones de  $T_1$  se entrenó un modelo de reducción de dimensionalidad utilizando PCA, al cual se le denominó MRT1. Con las representaciones reducidas de  $T_1$  se entrenaron diversos clasificadores utilizando los algoritmos especificados en la Tabla 3.7. Para obtener las métricas de los clasificadores, las representaciones de  $T_2$  fueron reducidas utilizando MRT1 y posteriormente se aplicó la etapa de prueba para obtener las métricas.

### **Experimento con Traducciones e IVIS No supervisado**

Para este experimento se utilizaron  $T_1$  y  $T_2$  como datos de entrenamiento y prueba respectivamente. Con las representaciones obtenidas con BETO de  $T_1$  se entrenó un modelo de reducción de dimensionalidad utilizando IVIS no supervisado, con este esquema de entrenamiento el algoritmo no utiliza las etiquetas de los datos. El modelo de reducción obtenido se denominó MRT2. Con las representaciones reducidas de  $T_1$  se entrenaron diversos clasificadores utilizando los algoritmos especificados en la Tabla 3.7. En la etapa de prueba se utilizaron las representaciones de  $T_2$ . Éstas representaciones fueron reducidas utilizando MRT2 y posteriormente en la etapa de prueba se obtuvieron las métricas de los clasificadores entrenados.

### **Experimento con Traducciones e IVIS Supervisado**

Para este experimento se utilizaron  $T_1$  y  $T_2$  como datos de entrenamiento y prueba respectivamente. Con las representaciones obtenidas con BETO de  $T_1$  se entrenó un modelo de reducción de dimensionalidad utilizando IVIS supervisado, con este esquema de entrenamiento el algoritmo utiliza las etiquetas de los datos. El modelo de reducción obtenido se denominó MRT3. Con las representaciones reducidas de  $T_1$  se entrenaron diversos clasificadores utilizando los algoritmos especificados en la Tabla 3.7. En la etapa de prueba se utilizaron las representaciones de  $T_2$ . Éstas representaciones fueron reducidas a dos dimensiones utilizando MRT3 y posteriormente se obtuvieron las métricas de los clasificadores entrenados.

### **Experimento con Traducciones e IVIS Semisupervisado**

Para este experimento se utilizaron  $T_3$  y  $T_4$  como datos de entrenamiento y prueba respectivamente. Con las representaciones obtenidas de  $T_3$  con el modelo BETO, se entrenó un modelo de reducción de dimensionalidad utilizando IVIS semisupervisado.  $T_3$  tiene dos columnas para representar a los datos etiquetados, una denominada *label* y otra *label semisupervisado*. La columna *label* contiene valores  $\{0, 1\}$ , representa la etiqueta original

de los datos. La columna *label semisupervisado* contiene valores de  $\{0, 1, -1\}$  y se utilizó para entrenar el modelo de reducción de dimensionalidad con IVIS semisupervisado. El modelo de reducción obtenido se denominó MRT4. Con las representaciones reducidas de  $T_3$  se entrenaron diversos clasificadores utilizando los algoritmos especificados en la Tabla 3.7. La columna *label* fue utilizada durante el entrenamiento de los clasificadores. En la etapa de prueba se utilizaron las representaciones de  $T_4$ . Éstas representaciones fueron reducidas utilizando MRT4 y posteriormente se obtuvieron las métricas de los clasificadores entrenados. Los datos de  $T_4$  se encuentran etiquetados con  $\{0, 1\}$ .

La Figura 3.5 presenta los datos de las clases con cada uno de los modelos de reducción de dimensionalidad obtenidos en los experimentos con traducciones. Con MRT4 se observa una mejor separación entre los datos de cada clase, mientras que con MRT1 se observa una mayor mezcla entre los datos de las clases.

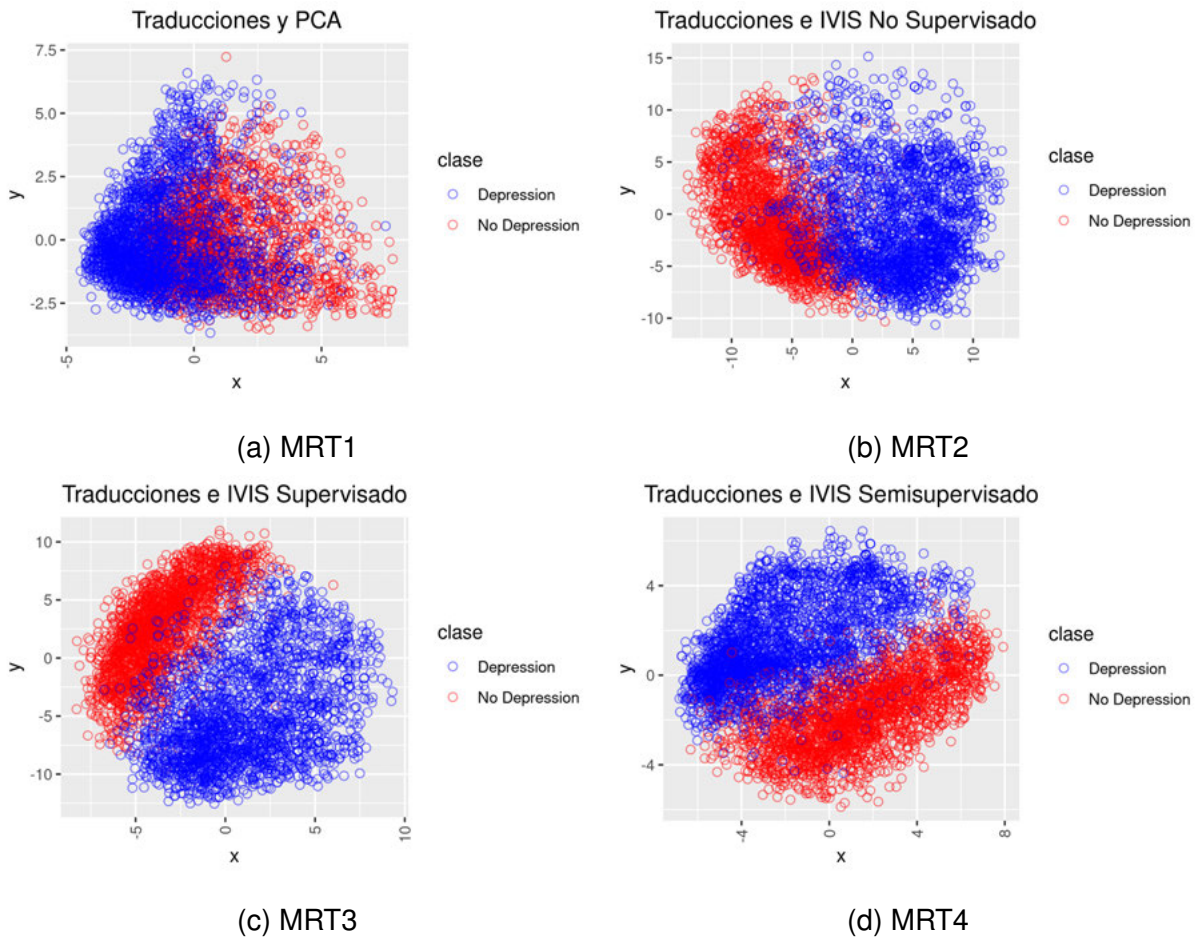


Figura 3.5: Entrenamiento de modelos de reducción de dimensionalidad con traducciones

### 3.6.2. Experimentos con KD

En los experimentos con KD se utilizaron los conjuntos  $\{D_1, D_2, D_3, D_4\}$  y las representaciones vectoriales de los textos se obtuvieron con KDMV1. Se realizaron cuatro experimentos variando el uso de los algoritmos de reducción de dimensionalidad PCA e IVIS. En las siguientes secciones se explican cada uno de los experimentos realizados.

#### Experimento con KD y PCA

Para este experimento se utilizaron  $D_1$  y  $D_2$  como datos de entrenamiento y prueba respectivamente. Se obtuvieron con KDMV1 las representaciones vectoriales de los

tweets contenidos en  $D_1$  y  $D_2$ . Con las representaciones de  $D_1$  se entrenó un modelo de reducción de dimensionalidad utilizando PCA, al cual se le denominó MRKD1. Con las representaciones reducidas de  $D_1$  se entrenaron diversos clasificadores utilizando los algoritmos especificados en la Tabla 3.8. Para obtener las métricas de los clasificadores, las representaciones de  $D_2$  fueron reducidas utilizando MRKD1 y posteriormente se aplicó la etapa de prueba para obtener las métricas.

### **Experimento con KD e IVIS No supervisado**

Para este experimento se utilizaron  $D_1$  y  $D_2$  como datos de entrenamiento y prueba respectivamente. Con las representaciones obtenidas con KDMV1 de  $D_1$  se entrenó un modelo de reducción de dimensionalidad utilizando IVIS no supervisado, con este esquema de entrenamiento el algoritmo no utiliza las etiquetas de los datos. El modelo de reducción obtenido se denominó MRKD2. Con las representaciones reducidas de  $D_1$  se entrenaron diversos clasificadores utilizando los algoritmos especificados en la Tabla 3.8. En la etapa de prueba se utilizó  $D_2$  y empleando reducción de dimensionalidad con utilizando MRKD2. Posteriormente en la etapa de prueba se obtuvieron las métricas de los clasificadores entrenados.

### **Experimento con KD e IVIS Supervisado**

Para este experimento se utilizaron  $D_1$  y  $D_2$  como datos de entrenamiento y prueba respectivamente. Con las representaciones obtenidas con KDMV1 de  $D_1$  se entrenó un modelo de reducción de dimensionalidad utilizando IVIS supervisado, con este esquema de entrenamiento el algoritmo utiliza las etiquetas de los datos. El modelo de reducción obtenido se denominó MRKD3. Con las representaciones reducidas de  $D_1$  se entrenaron diversos clasificadores utilizando los algoritmos especificados en la Tabla 3.8. El conjunto  $D_2$  se utilizó en la etapa de prueba, las representaciones de éstos datos fueron reducidas con MRKD3. Posteriormente se obtuvieron las representaciones reducidas de  $D_2$  para obtener las métricas de los clasificadores entrenados.

## Experimento con KD e IVIS Semisupervisado

Para este experimento se utilizaron  $D_3$  y  $D_4$  como datos de entrenamiento y prueba respectivamente. Con las representaciones obtenidas con KDMV1 de  $D_3$  se entrenó un modelo de reducción de dimensionalidad utilizando IVIS semisupervisado. En el entrenamiento semisupervisado se requieren datos etiquetados con valores en  $\{0, 1, -1\}$ , por lo cual  $D_2$  contiene una columna denominada *label semisupervisado*. La columna *label semisupervisado* contiene valores de  $\{0, 1, -1\}$  y se utilizó para entrenar el modelo de reducción de dimensionalidad con IVIS semisupervisado. El modelo de reducción obtenido se denominó MRKD4. Por otra parte  $D_3$  tiene una columna *label* con valores  $\{0, 1\}$  la cual representa la etiqueta original de los datos. Con las representaciones reducidas del conjunto  $D_3$  se entrenaron diversos clasificadores utilizando los algoritmos especificados en la Tabla 3.8. La columna *label* fue utilizada durante el entrenamiento de los clasificadores. En la etapa de prueba se utilizaron las representaciones de  $D_4$ . Éstas representaciones fueron reducidas utilizando MRKD4 y posteriormente se obtuvieron las métricas de los clasificadores entrenados. Los datos de  $D_4$  se encuentran etiquetados con  $\{0, 1\}$ .

La Figura 3.6 presenta los datos de las clases con cada uno de los modelos de reducción de dimensionalidad obtenidos en los experimentos con KD. Se puede observar en los cuatro experimentos una mayor separación entre los datos.

En el siguiente capítulo se presentan los resultados de la etapa de prueba de los clasificadores generados.

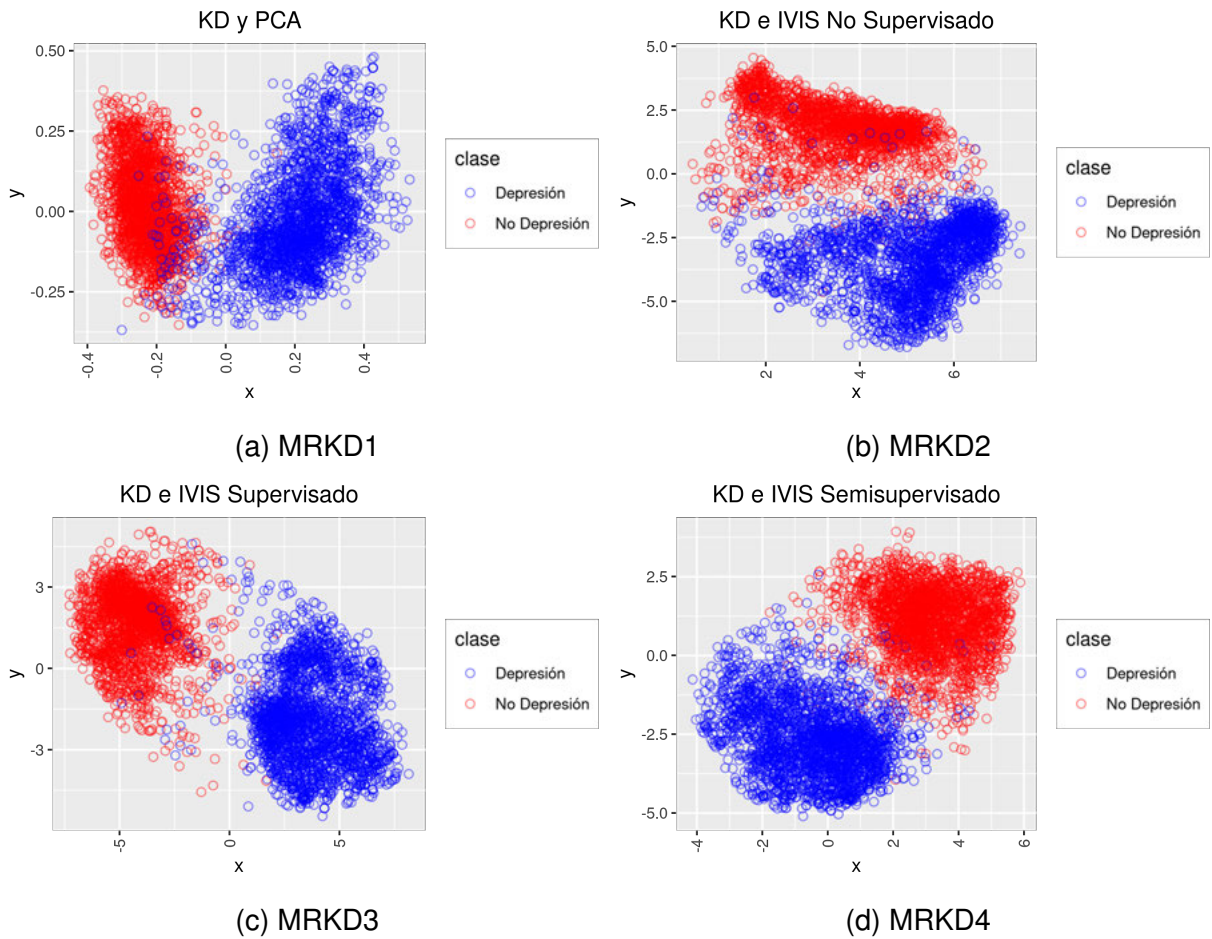


Figura 3.6: Entrenamiento de modelos de reducción de dimensionalidad con KD

# Capítulo 4

## Resultados

El presente capítulo se compone de tres secciones. Las primeras dos secciones muestran los resultados de los experimentos descritos en el capítulo anterior. La tercera sección consiste en un caso de estudio sobre la depresión en México, en la cual se muestran los resultados del análisis espacio-temporal en tweets recolectados con el AGEI en el período 2018 a 2021.

### 4.1. Resultados de los experimentos con traducciones

A continuación se presentan los resultados de los experimentos utilizando traducciones. En cada cada resultado se muestra la tabla con las métricas de cada clasificador en los experimentos. De igual forma se presentan las gráficas con las regiones de decisión, estas regiones fueron obtenidas con los datos de entrenamiento. Las regiones de decisión representan la probabilidad de que un dato pertenezca a la clase asignada. En el problema de clasificación binaria abordado en el presente estudio, se muestran dos regiones de decisión representadas en color azul y rojo para representar a cada una de las clases. En las gráficas muestran los datos de entrenamiento representados con un círculo ( $\circ$ ) y los datos de prueba con un triángulo ( $\triangle$ ).



### 4.1.1. Traducciones y PCA

En la Tabla 4.1 se presentan las métricas obtenidas por los clasificadores entrenados con traducciones y reducción de dimensionalidad con PCA. Considerando el Valor-F de 0,74, el clasificador obtenido a través de QDA tuvo el mejor desempeño. La Figura 4.1 presenta la distribución de los datos en las regiones de decisión en cada clasificador entrenado. Estas regiones representan la probabilidad de que un dato pertenezca a una clase, a mayor intensidad de color, más alta es la probabilidad de que el dato corresponda a dicha clase.

Tabla 4.1: Traducciones y PCA

Clasificador	Exactitud	Precisión	Exhaustividd	Valor-F
Regresión Logística	0.72	0.73	0.70	0.71
SVM Lineal	0.72	0.74	0.68	0.71
Procesos Gaussianos	0.71	0.82	0.54	0.65
QDA	0.72	0.71	0.76	<b>0.74</b>

En la Figura 4.1 se pueden observar los datos representados en color rojo y azul para las clases no depresión y depresión respectivamente. Los datos se encuentran mezclados. Incluso se pueden observar datos en color rojo debajo de los datos de color azul, lo cual explica los resultados de las métricas del clasificador. En las regiones de decisión se puede observar un área en color blanco, lo cual representa el límite decisión, es deseable que en esta área existan pocos datos. En las Figuras 4.1a y 4.1b se pueden observar que en la región de color blanco tiene una alta concentración de datos de prueba y entrenamiento. La concentración de datos en la franja blanca reflejan las métricas obtenidas con los clasificadores entrenados con regresión logística y SVM lineal. Las Figuras 4.1c y 4.1d muestran la distribución de los datos en las regiones de decisión de los clasificadores

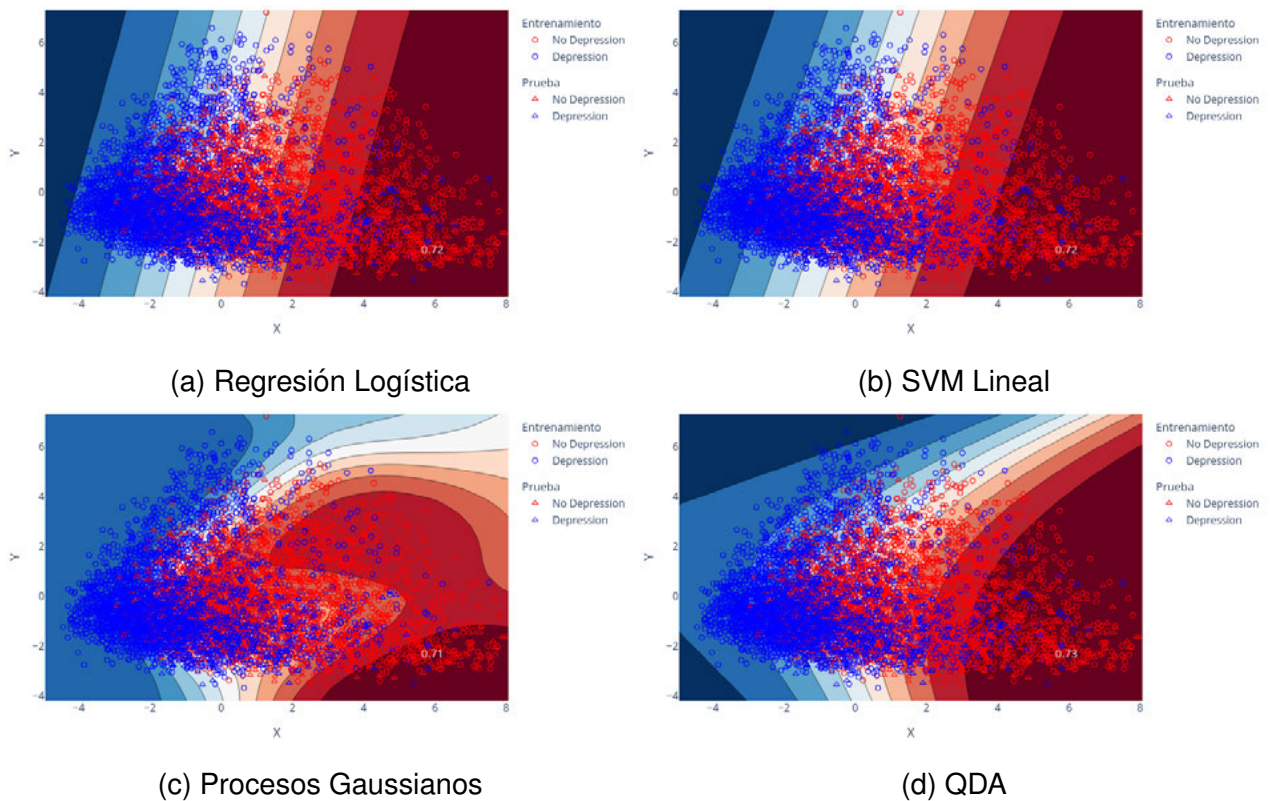


Figura 4.1: Resultados experimentos traducciones con PCA

obtenidos con procesos gaussianos y con QDA respectivamente, se puede observar que los datos de las clases se encuentran mezclados.

#### 4.1.2. Traducciones e IVIS Supervisado

En este experimento se utilizó IVIS Supervisado para la reducción de dimensionalidad. En la Tabla 4.2 se puede observar que el clasificador QDA tuvo el mejor desempeño con un Valor-F de 0,83. Comparando estos resultados con los presentados en la Tabla 4.1, el clasificador QDA en este experimento mejoró en un 9%, debido a la reducción de dimensionalidad obtenida con IVIS Supervisado.

La Figura 4.2 presenta la distribución de los datos en las regiones de decisión para los

Tabla 4.2: Traducciones e IVIS Supervisado

Clasificador	Exactitud	Precisión	Exhaustividad	Valor-F
Regresión Logística	0.82	0.84	0.81	0.82
SVM Lineal	0.82	0.84	0.81	0.82
Procesos Gaussianos	0.82	0.81	0.81	0.82
QDA	0.84	0.84	0.83	<b>0.83</b>

clasificadores. En las Figuras 4.2a y 4.2b se observa la distribución de las clases de depresión y no depresión de los utilizados en la etapa de entrenamiento y prueba. En las Figuras 4.2c y 4.2d se observa la separación de los datos y aunque en la región de color blanco se observa un concentración de datos, los datos de cada clase se ven más separados.

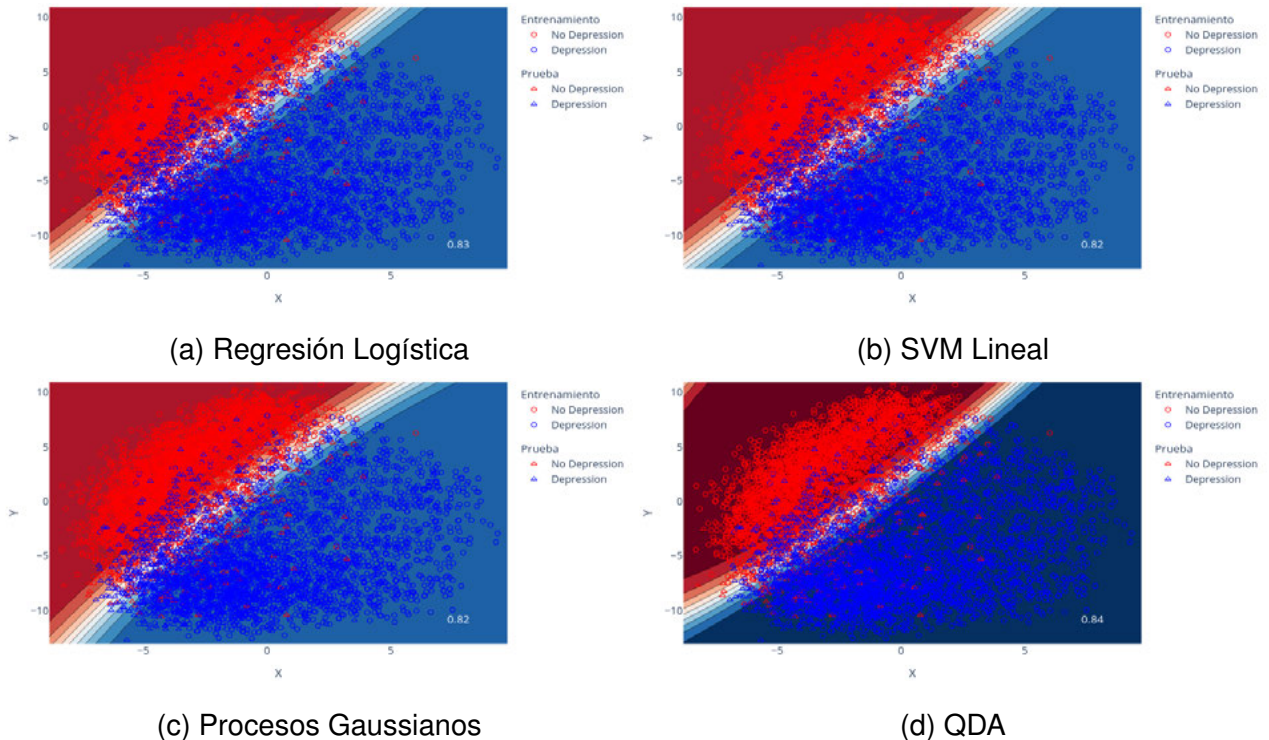


Figura 4.2: Resultados experimentos traducciones con IVIS Supervisado

### 4.1.3. Traducciones e IVIS No Supervisado

En este experimento se utilizó la reducción de dimensionalidad IVIS Supervisado. La Tabla 4.3 presenta los resultados de las métricas obtenidas por los clasificadores. En este experimento el clasificador obtenido con SVM Lineal tuvo el mejor desempeño con un valor-F de 0,85, sin embargo comparando los clasificadores QDA con los resultados de los experimentos anteriores, en este experimento QDA obtuvo un valor-F de 0,84. Comparando el resultado de QDA representa una ganancia del 10% con respecto al experimento con PCA. Esta mejora se debe a utilizar en la reducción de dimensionalidad IVIS No Supervisado.

La Figura 4.3 muestra la distribución de los datos de entrenamiento y de prueba en las regiones de decisión. En las Figuras 4.3a y 4.3b corresponde a la distribución de los datos en las regiones de decisión para los clasificadores obtenidos con regresión logística y SVM Lineal, respectivamente y se puede observar una similitud en las regiones de decisión. Por otro lado las Figuras 4.3c y 4.3d presentan la distribución de los datos en las regiones de decisión para los clasificadores obtenidos con procesos gaussianos y QDA, respectivamente. En cada una de las Figuras se puede observar una mejor separación de los datos.

Tabla 4.3: Traducciones e IVIS No Supervisado

Clasificador	Exactitud	Precision	Exhaustividad	Valor-F
Regresión Logística	0.84	0.84	0.85	0.85
SVM Lineal	0.84	0.84	0.85	<b>0.85</b>
Procesos Gaussianos	0.84	0.84	0.85	0.84
QDA	0.84	0.83	0.84	0.84

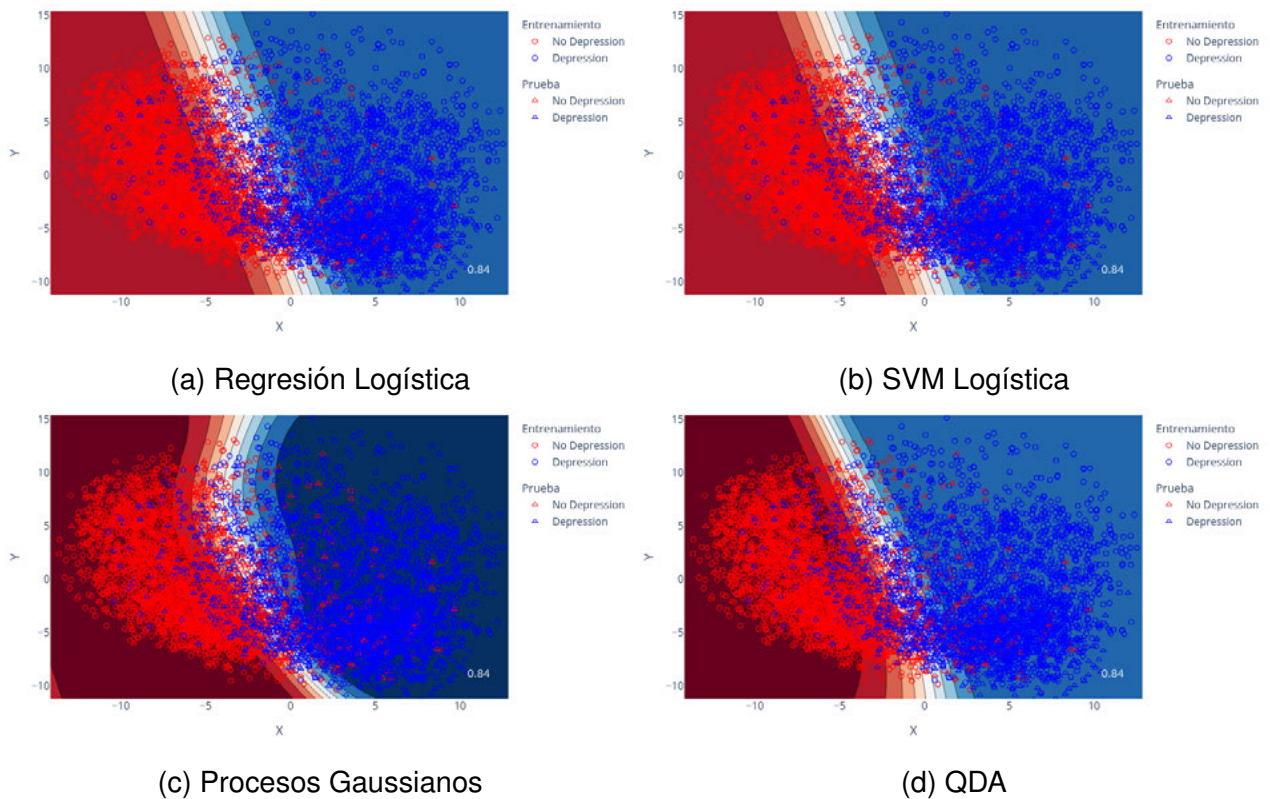


Figura 4.3: Resultados de los experimentos con traducciones e IVIS No Supervisado

#### 4.1.4. Traducciones e IVIS Semisupervisado

A continuación se presentan los resultados del experimento utilizando traducciones y reducción de dimensionalidad IVIS Semisupervisado. En la Tabla 4.4 se puede observar que el clasificador obtenido con regresión logística tuvo el mejor desempeño con un valor-F de 0,84.

Las Figuras 4.4a y 4.4b muestran la distribución de los datos en las regiones de decisión para los clasificadores con regresión logística y SVM Lineal. Se observan los grupos de datos de cada clase con una mejor separación comparados con los experimentos anteriores. Por otro lado en las Figuras 4.4c y 4.4d, muestran la distribución de los datos en las regiones de decisión para los clasificadores obtenidos con procesos gaussianos y QDA, respectivamente.

Tabla 4.4: Traducciones e IVIS Semisupervisado

Clasificador	Exactitud	Precision	Exhaustividad	Valor-F
Regresión Logística	0.82	0.82	0.84	<b>0.84</b>
SVM Lineal	0.82	0.83	0.83	0.83
Procesos Gaussianos	0.82	0.82	0.83	0.83
QDA	0.83	0.82	0.86	0.83

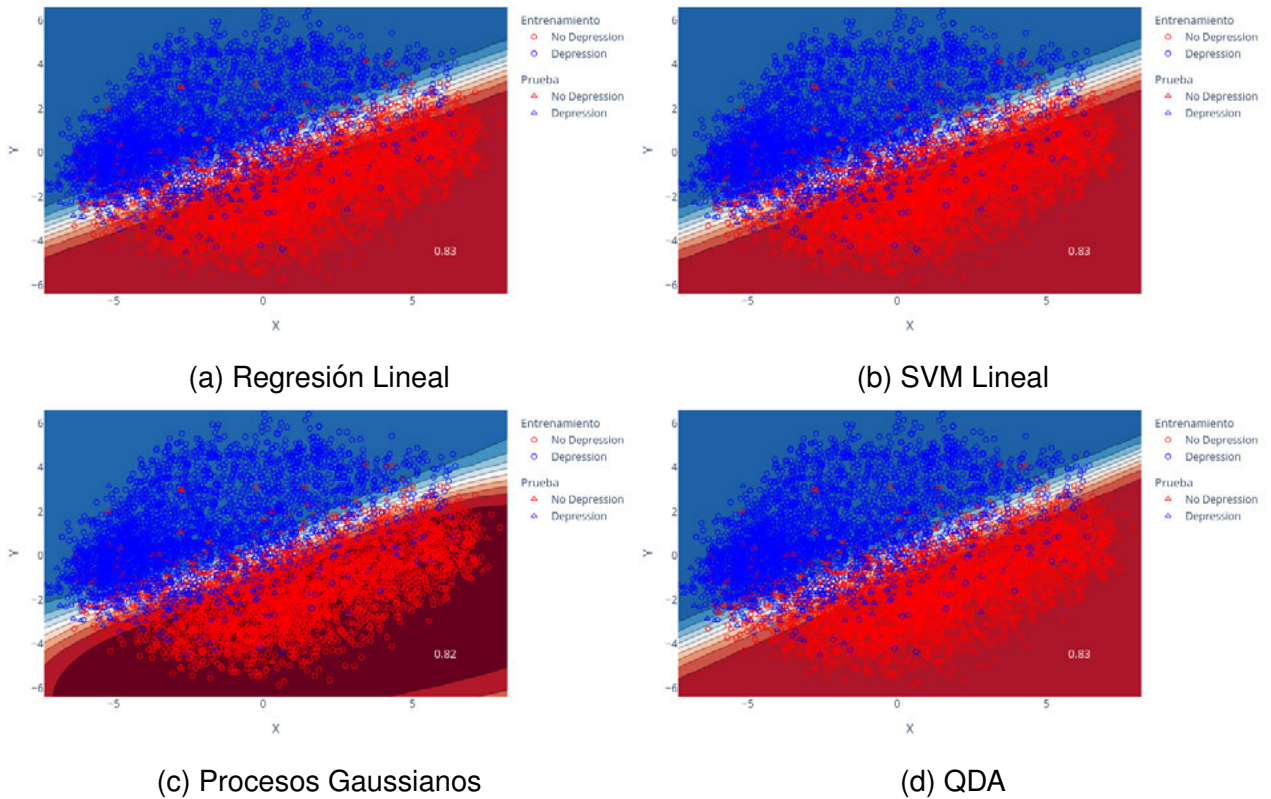


Figura 4.4: Resultados experimentos traducciones con IVIS Semisupervisado

A partir de los resultados obtenidos en los experimentos utilizando textos traducidos, extracción de características con BETO, reducción de dimensionalidad IVIS (supervisado, no supervisado y semisupervisado), genera mejores resultados a diferencia de utilizar PCA.

En cuanto a los algoritmos utilizados en el entrenamiento de clasificadores se concluye

que las mejores métricas se obtuvieron en el experimento con IVIS No Supervisado con los algoritmos de regresión logística y SVM Lineal con un Valor-F de 0,85.

Utilizando reducción de dimensionalidad IVIS en los diversos esquemas de entrenamiento, supervisado, no supervisado y semisupervisado, las métricas de los clasificadores obtenidos con los diferentes algoritmos no presentan diferencias significativas.

## 4.2. Resultados de los experimentos con KD

En esta sección se presentan los resultados de los experimentos descritos en el capítulo anterior para la estrategia basada en destilación de conocimiento (KD). En cada experimento se presenta la tabla de métricas obtenidas por los clasificadores y las gráficas de la distribución de los datos de entrenamiento y prueba en las regiones de decisión. Las regiones de decisión se obtuvieron mediante los datos de entrenamiento.

### 4.2.1. KD y PCA

La Tabla 4.5 presenta los resultados de las métricas obtenidas por los clasificadores en la etapa de prueba. En este experimento el clasificador obtenido con procesos gaussianos tuvo el mejor desempeño con un valor-F de 0,77.

Tabla 4.5: KD y PCA

Clasificador	Exactitud	Precisión	Exhaustividad	Valor-F
Regresión Logística	0.73	1	0.473	0.64
SVM Lineal	0.68	1	0.36	0.53
Procesos Gaussianos	0.81	0.99	0.63	<b>0.77</b>
QDA	0.82	0.99	0.60	0.75

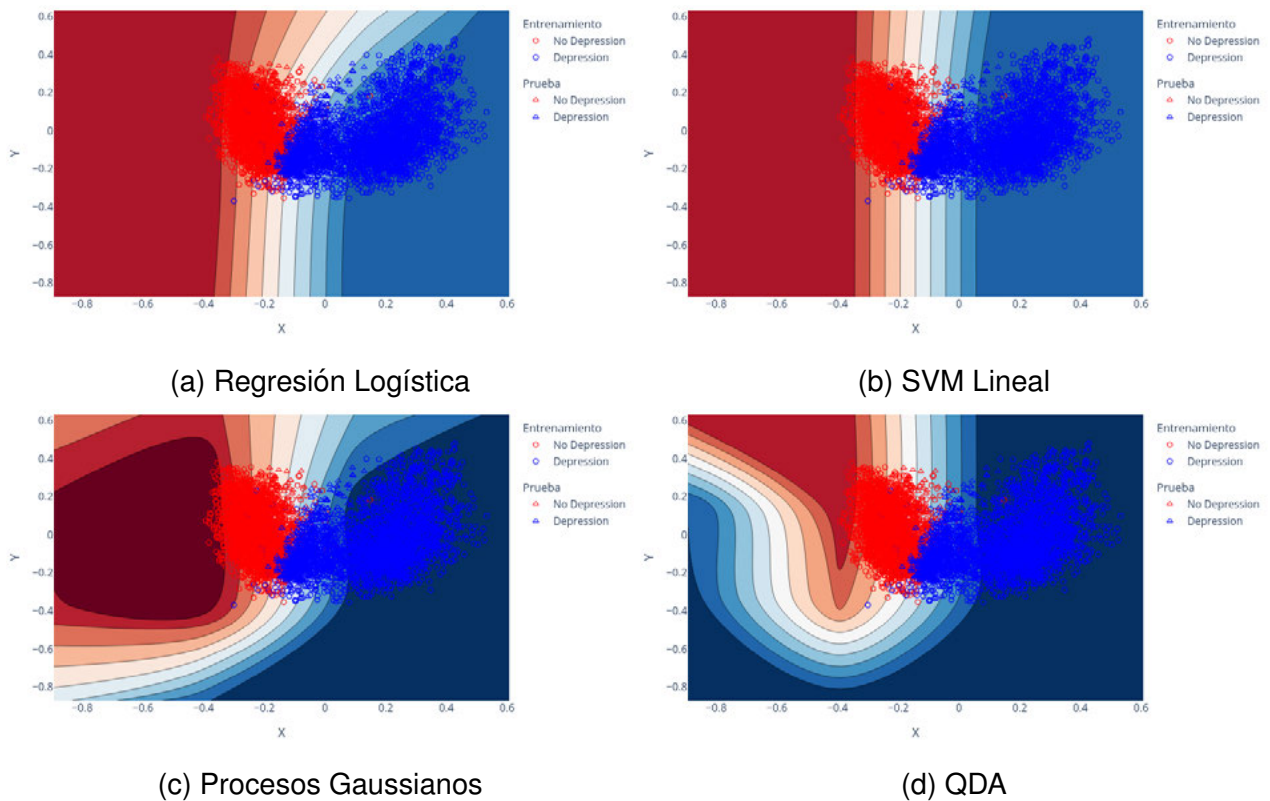


Figura 4.5: Resultados experimentos KD con PCA

Las Figuras 4.5a y 4.5b presentan las regiones de decisión de los clasificadores obtenidos con regresión logística y SVM Lineal, respectivamente, en las figuras se puede observar la distribución de los datos de cada clase, éstos se encuentran en las zonas donde la probabilidad de pertenencia a cada clase es baja. De igual forma en la zona de color blanco, la cual representa el límite de decisión, se pueden observar que los datos de las clases se encuentran mezclados, lo cual confirma las métricas obtenidas por los clasificadores. Las Figuras 4.5c y 4.5d presentan regiones de decisión para los clasificadores procesos gaussianos y QDA, respectivamente; de forma similar, los datos se encuentran concentrados en la zona de color blanco, lo que muestra que los datos no están separados.



## 4.2.2. KD e IVIS Supervisado

Los resultados de este experimento son presentados en la Tabla 4.6, la cual muestra las métricas obtenidas por los clasificadores. El clasificador que tuvo el mejor desempeño es QDA con un valor-F de 0,93. Comparando este valor-F con el mejor resultado obtenido con PCA y KD, se logró una mejora en un 16%. Las Figuras 4.6a y 4.6b corresponden a las regiones de decisión y la distribución de los datos para los clasificadores obtenidos con regresión logística y SVM Lineal respectivamente. En estas figuras se puede observar una mejor separación entre los datos, los cuales confirman las métricas presentadas en la Tabla 4.6. Las Figuras 4.6c y 4.6d representan las regiones de decisión y la distribución de los datos en las regiones. La Figura 4.6d muestran la distribución de los datos de las clases en las regiones de decisión para el clasificador entrenado con QDA, se observa una mejor separación de los datos de acuerdo a las zonas de decisión.

Tabla 4.6: KD e IVIS Supervisado

Clasificador	Exactitud	Precisión	Exhaustividad	Valor-F
Regresión Logística	0.91	0.98	0.84	0.90
SVM Lineal	0.90	0.99	0.80	0.89
Procesos Gaussianos	0.90	0.98	0.82	0.89
QDA	0.93	0.97	0.89	<b>0.93</b>

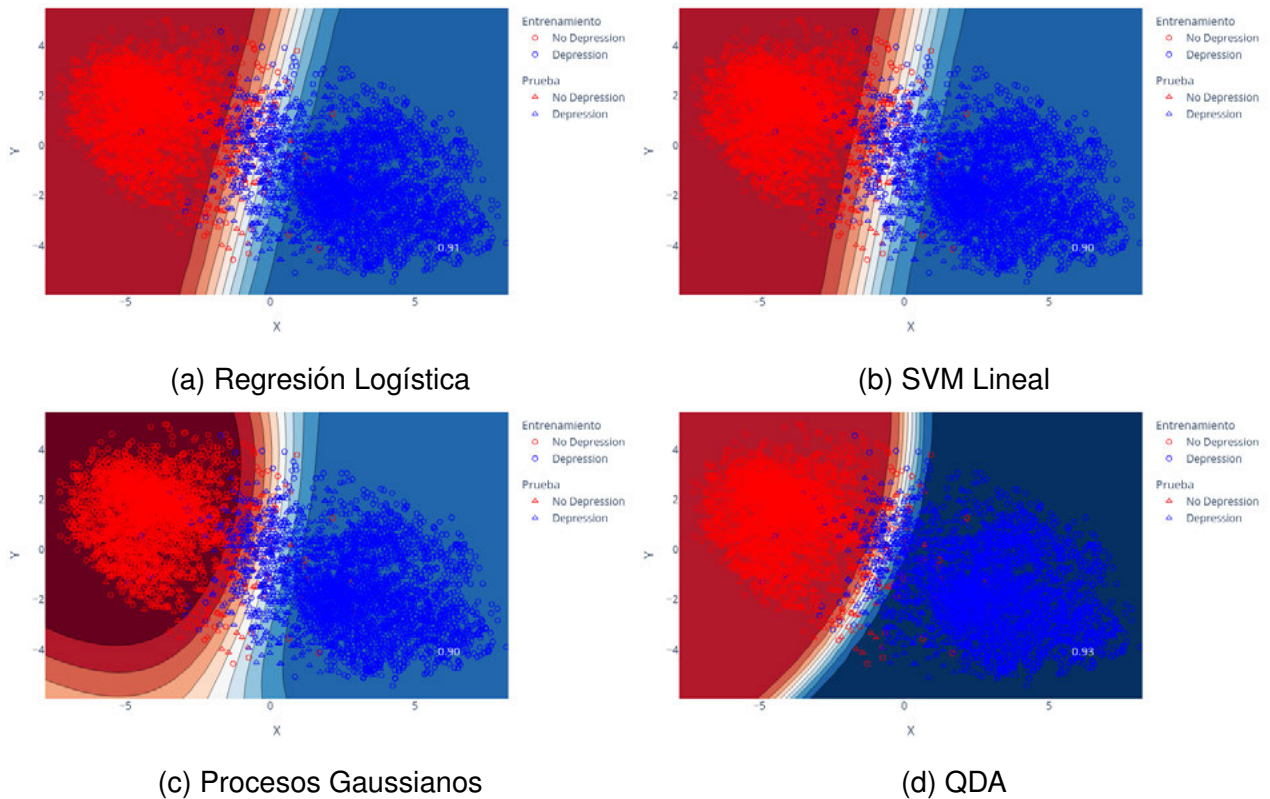


Figura 4.6: Resultados experimentos KD con IVIS Supervisado

### 4.2.3. KD e IVIS No Supervisado

Los resultados de este experimento son presentados en la Tabla 4.7. En este caso el clasificador obtenido con QDA tuvo el mejor desempeño con un valor-F de 0,93. Las métricas obtenidas en los diversos clasificadores no presentan diferencias significativas comparadas con el mejor resultado obtenido. Las Figuras 4.7a y 4.7b muestran la distribución de los datos en las regiones de decisión de los clasificadores obtenidos con regresión logística y SVM lineal, respectivamente. Las Figuras 4.7c y 4.7d corresponden a las regiones de decisión para los clasificadores obtenidos con procesos gaussianos y con QDA, respectivamente. En La Figura 4.7d se puede observar una mejor separación entre los datos de cada clase, los cuales confirman las métricas presentas en la Tabla 4.7.

Tabla 4.7: KD e IVIS No Supervisado

Clasificador	Exactitud	Precisión	Exhaustividad	Valor-F
Regresión Logística	0.92	0.93	0.90	0.91
SVM Lineal	0.92	0.93	0.90	0.92
Procesos Gaussianos	0.91	0.94	0.88	0.91
QDA	0.93	0.92	0.94	<b>0.93</b>

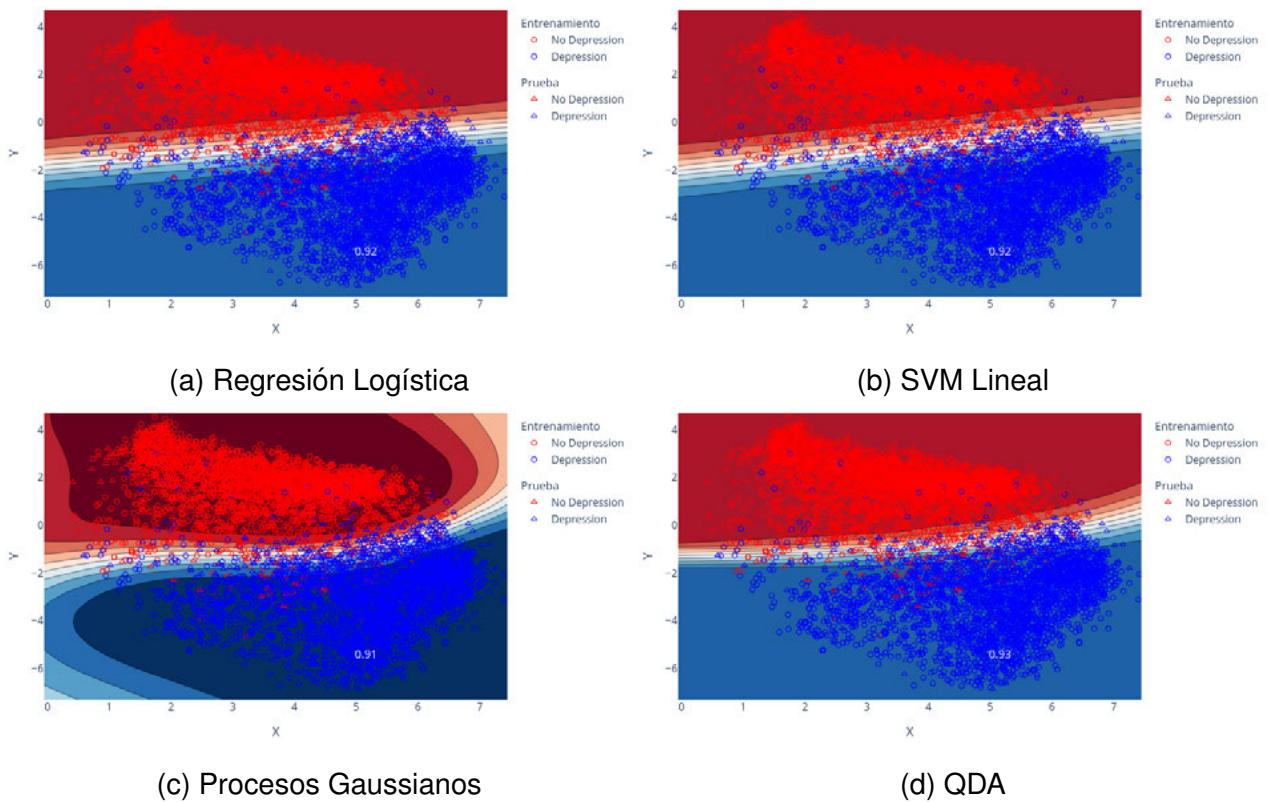


Figura 4.7: Resultados experimentos KD con IVIS No Supervisado

#### 4.2.4. KD e IVIS Semisupervisado

Los resultados de las métricas obtenidas por los clasificadores en este experimento son presentados en la Tabla 4.8. En este caso el clasificador generado con regresión logística tuvo el mejor desempeño con un valor-F de 0,95. La Figura 4.8 muestra la distribución de los datos de cada clase en las regiones de decisión para los clasificadores

obtenidos. En cada una de las Figuras se observa una mejor separación de los datos, lo cual confirman las métricas obtenidas. La diferencia entre los clasificadores y el mejor resultado con regresión logística es de 1 %, lo cual no es significativa. Este experimento tiene relevancia dado que el entrenamiento de los clasificadores fue con una mezcla de datos en inglés y una mínima parte de datos en español, lo cual puede explicar la mejora en las métricas.

Tabla 4.8: KD e IVIS Semi Supervisado

Clasificador	Exactitud	Precisión	Exhaustividad	Valor-F
Regresión Logística	0.95	0.93	0.94	<b>0.95</b>
SVM Lineal	0.95	0.96	0.94	0.95
Procesos Gaussianos	0.95	0.96	0.93	0.94
QDA	0.95	0.96	0.93	0.94

A partir de los resultados presentados con el esquema basado en destilación del conocimiento, reducción de dimensionalidad con IVIS (supervisado, no supervisado y semi-supervisado), genera mejores resultados que con PCA.

El mejor valor-F (0,95) se obtuvo con regresión logística en el experimento destilación de conocimiento y reducción de dimensionalidad con IVIS semisupervisado. Cabe señalar que los otros clasificadores generados con los algoritmos como SVM lineal, procesos gaussianos y QDA tienen una mínima diferencia con respecto a regresión logística.

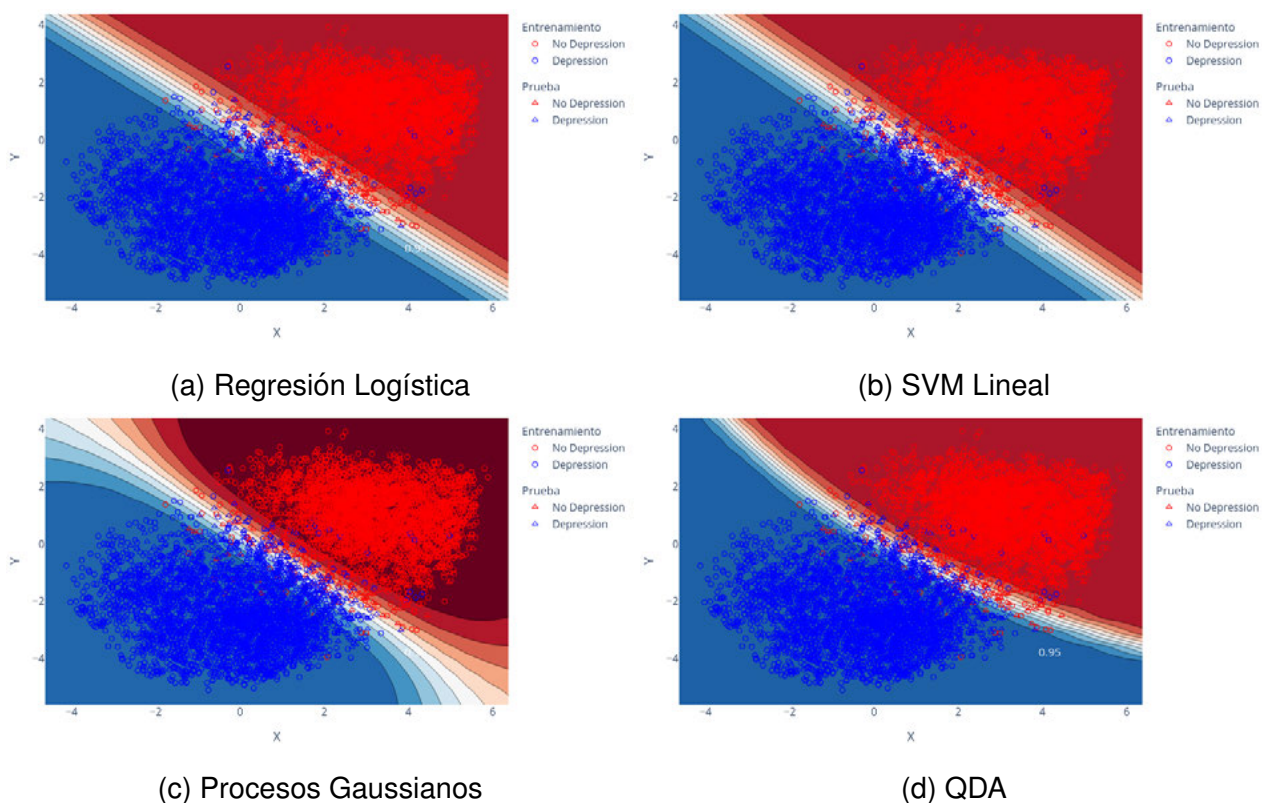


Figura 4.8: Resultados experimentos KD con IVIS Semisupervisado

#### 4.2.5. KD sin reducción de dimensionalidad

Un experimento adicional consistió en utilizar todas las dimensiones de los vectores para entrenar los clasificadores. En este experimento se entrenó con  $D_1$  y la etapa de prueba fue con  $D_2$ . En la Tabla 4.9 se puede observar que las métricas de los clasificadores son bajas.

Clasificador	Exactitud	Precisión	Exhaustividad	Valor-F
Linear Regression	0.72	1	0.456	0.62
SVM Linear	0.71	1	0.428	0.559
Procesos Gaussianos	0.79	0.98	0.60	0.74
QDA	0.80	0.89	0.68	0.74

Tabla 4.9: Resultados con todas las dimensiones esquema multilingual

### 4.3. Caso de estudio: Depresión en México

Con base en los resultados obtenidos, mismos que se presentaron en la sección anterior, se determinó que los mejores modelos de clasificación se obtienen con KD e IVIS Semisupervisado. Dado que los clasificadores obtenidos con el esquema semisupervisado no tiene una diferencia significativa, se utilizó QDA para clasificar tweets en el período 2018 a 2021 en toda la República Mexicana, con el propósito de determinar el impacto en la salud mental provocada por la pandemia Covid-19. Posteriormente con los tweets clasificados se realizaron diversos análisis.

#### 4.3.1. Clasificación de tweets

La recolección de los tweets fue por medio del AGEI, plataforma desarrollada por el Laboratorio Nacional de GeoInteligencia (GeoInt). La Tabla 4.10 presenta la distribución de los datos de acuerdo al año de publicación. Cada conjunto de datos está conformado por el texto, la fecha, la cuenta del usuario de twitter y los datos de geolocalización a través de dos columnas que almacena la latitud y longitud.

Tabla 4.10: Conjuntos de tweets de los años 2018-2021

Año	Tamaño	Tweets
2018	5.6 GB	16,912,135
2019	5.9 GB	17,198,272
2020	4.7 GB	13,635,630
2021	4.1 GB	12,106,644

Para realizar la clasificación los datos, éstos fueron procesados de manera similar a los datos utilizados en el entrenamiento del clasificador. La clasificación consistió en los siguientes pasos:

- Preprocesamiento.

- Extracción de características con KDMV1.
- Reducción de dimensionalidad con el modelo obtenido con IVIS Semisupervisado (MRKD4).
- Predicción de la etiqueta con el clasificador QDA.

La Tabla 4.11 muestra el resultado del proceso de clasificación, así como la distribución de los tweets por año y clase.

Tabla 4.11: Clasificación de tweets del período 2018-2021

Año	Tamaño	Tweets	Clase depresión	Clase No Depresión
2018	5.6 GB	16, 912, 135	2, 628, 694	14, 283, 441
2019	5.9 GB	17, 198, 272	2, 853, 876	14, 344, 396
2020	4.7 GB	13, 635, 630	2, 518, 144	11, 117, 486
2021	4.1 GB	12, 106, 644	2, 038, 312	10, 068, 332

### Georreferenciación a nivel estatal

Posterior a la clasificación se aplicó el proceso de georreferenciar cada tweet en el Estado de la República Mexicana. Para esto se utilizó el *shapefile* (SHPMEX) obtenido en el repositorio de [gits.igg](https://www.gits.igg.unam.mx/)<sup>1</sup>. SHPMEX contiene los Estados de la República Mexicana y el dato espacial del polígono de cada entidad, además de la clave de la entidad. Este proceso se realizó mediante el lenguaje python y las librerías *shapely*, *geopandas*. El proceso de georreferenciación se realizó utilizando la proyección espacial WGS 84 (World Geodetic System 1984).

Durante este proceso se detectaron tweets que se encuentran fuera de la República Mexicana, a los cuales se les asignó el valor de *FDM*, estos tweets se encontraban en la frontera Norte y Sur de la República Mexicana.

<sup>1</sup><https://www.gits.igg.unam.mx/idea/descarga>

La Tabla 4.12 presenta la distribución de los datos por clase y que pertenecen a la República Mexicana. Se puede observar en el 2020 un incremento del 2 % para los tweets etiquetados en la clase depresión.

Tabla 4.12: Datos de los Tweets clasificados en México

Año	Tamaño	Tweets	Clase depresión	%	Clase No depresión	%
2018	5.6 GB	15,535,919	2,414,887	15,5 %	13,121,032	84,5 %
2019	5.9 GB	15,775,594	2,618,534	16,5 %	13,157,060	83,5 %
2020	4.7 GB	12,301,589	2,279,891	18,5 %	10,021,698	81,5 %
2021	4.1 GB	10,947,488	1,849,643	16,8 %	9,097,845	83,2 %

A partir de los datos clasificados y georreferenciados se realizaron diversos análisis estadísticos los cuales son presentados en las siguientes secciones.

### 4.3.2. Análisis espacio-temporal

Este análisis se compone de diversos procesos para visualizar el comportamiento de los tweets de la clase depresión y posible correlaciones con datos obtenidos de INEGI, como son la tasa de suicidio y estado de ánimo.

#### Distribución acumulada

Para determinar si la pandemia tuvo algún impacto en el comportamiento de los tweets de la clase depresión, para cada año se estimó la distribución de probabilidad acumulada por mes y año de los tweets. La probabilidad de ocurrencia de un tweet en un mes y año se obtuvo de la siguiente forma:

$$\text{Probabilidad Tweet en la clase depresión}_{\text{año}} = \frac{\text{Número de Tweets por mes}}{\text{Total Tweets por año}} \quad (4.1)$$

La Figura 4.9 presenta la distribución de probabilidad acumulada de los tweets, se puede observar que la probabilidad de ocurrencia de un tweet en la clase depresión para



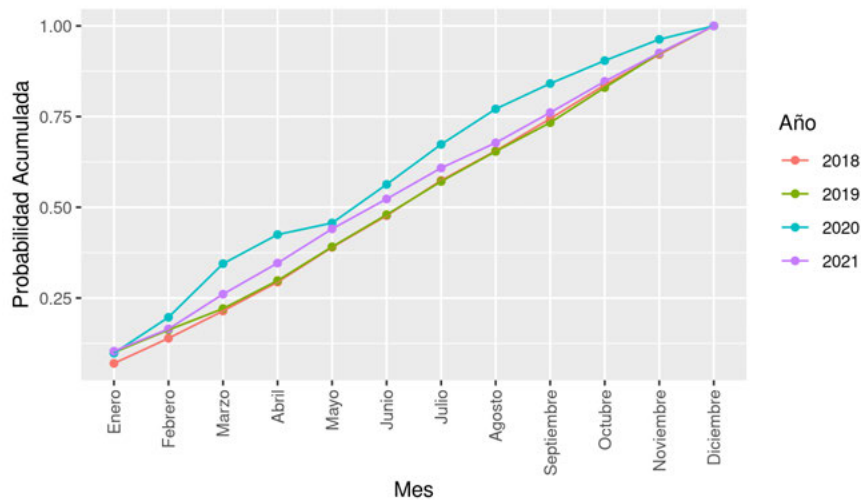


Figura 4.9: Distribución acumulada de tweets en la clase depresión

los años 2018 y 2019 tienen una tendencia similar, sin embargo en el 2020 presenta dos curvas de crecimiento en las probabilidades, lo cual se interpreta como un crecimiento de la probabilidad de ocurrencia de un tweet en la clase depresión a partir del mes de Marzo del 2020. Este crecimiento puede ser debido a la declaración del inicio de la pandemia Covid-19 en México. De igual forma, se puede observar que en el mes de mayo de 2020 existe un descenso de tweets, esto es debido a que no se tuvieron datos de todos los días del mes. Para el año 2021 se observa una curva de crecimiento de tweets, sin embargo no supera la del año 2020. Para el mes de agosto de 2021 la tendencia de los tweets de depresión empiezan a ser similares a los años 2018 y 2019.

### 4.3.3. Distribución Mensual por Entidad

El comportamiento de los tweets de depresión pueden ser diferentes en tiempo y lugar. Debido a lo anterior para determinar patrones de comportamiento considerando el dato espacial, se estimó la tasa mensual de tweets por Entidad. Esta tasa se obtuvo por medio de la siguiente expresión:

$$\text{Tasa Mensual}_{\text{Estado}} = \frac{\text{Número de Tweets por mes y Estado}}{\text{Total Tweets por Estado y año}} \quad (4.2)$$

La Figura 4.10 presenta la tasa mensual por Entidad de los tweets de la clase de depresión. Se puede observar en el mes de marzo de 2020, una alta tasa de tweets de depresión en todos los estados, esto puede ser un efecto de la declaración del inicio de la pandemia Covid-19<sup>2</sup>. Mientras que la franja de color azul en el mes de Mayo de 2020 se debe que no se contó con tweets de todos los días del mes.

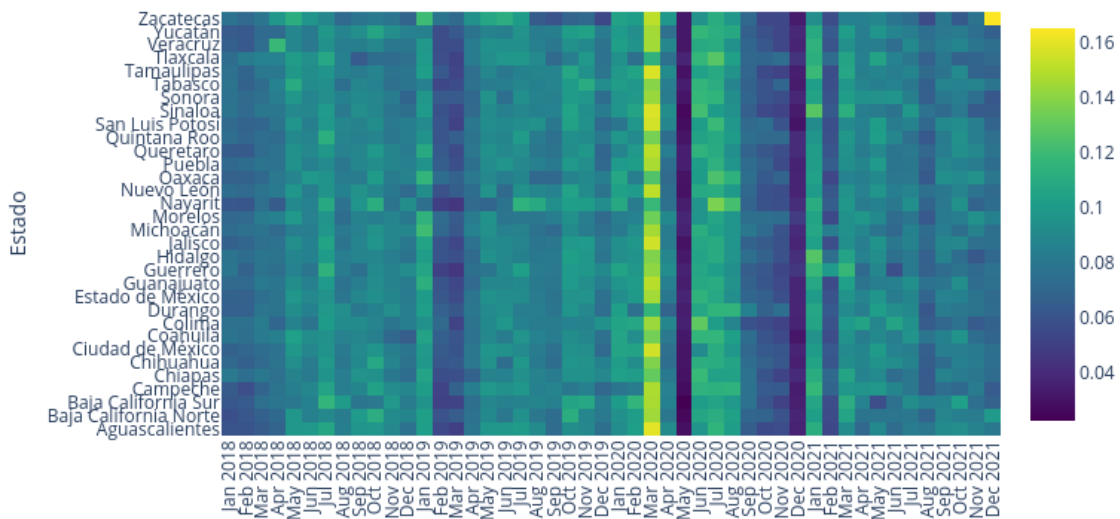


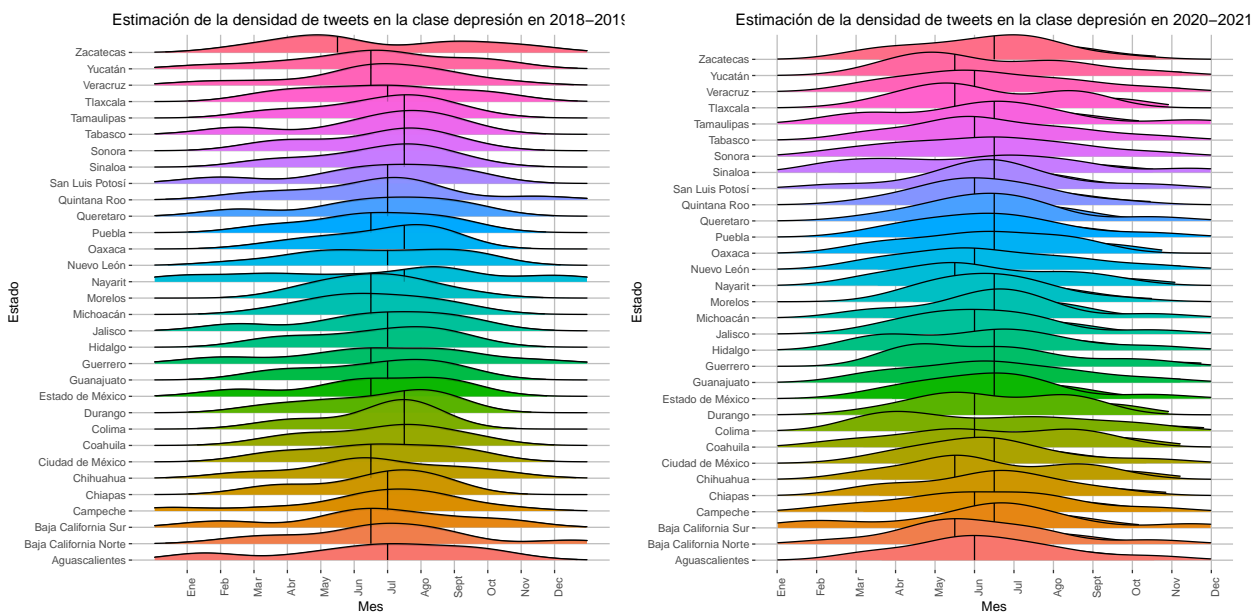
Figura 4.10: Tasa mensual de tweets de depresión por Entidad

De manera adicional para comparar el efecto de la pandemia, se estimaron las tasas mensuales de los tweets definiendo dos períodos, el primero corresponde a la pre-pandemia (2018-2019) y el segundo corresponde a la pandemia (2020-2021), a los cuales denominamos pre-pandemia y pandemia respectivamente. Esta tasa se obtuvo por medio de la siguiente expresión:

$$\text{Tasa Mensual}_{\text{Estado y Período}} = \frac{\text{Número de Tweets por mes y Estado en el período}}{\text{Total Tweets por Estado y período}} \quad (4.3)$$

<sup>2</sup><https://coronavirus.gob.mx/>

En las Figuras 4.11a y 4.11b, se muestra la densidad de los tweets por mes y estado en los períodos de prepandemia y pandemia. En la Figura 4.11a se observa que las entidades alcanzan una media entre los meses de Junio y Agosto, mientras que en la Figura 4.11b se observa que durante el período de la pandemia los Estados alcanzan una media entre los meses de Mayo a Julio. También se puede observar que estados como Tamaulipas, Tabasco, Sonora y Sinaloa, que en el período prepandemia tenían un comportamiento similar durante el período de pandemia tuvieron cambios en la distribución. Estos cambios en la distribución pueden ser derivados a cambios en el estilo de vida originados por el Covid-19 como las medidas de aislamiento, efectos económicos, pérdida de familiares, entre otros.



(a) Prepandemia 2018-2019

(b) Pandemia 2020-2021

Figura 4.11: Distribución de los tweets en períodos prepandemia y pandemia

#### 4.3.4. Correlación con datos de salud mental reportados por INEGI

A partir de los datos de la clase depresión se estimaron la tasa de tweets y tasa de usuarios por cada 100 mil habitantes. Los cálculos se obtuvieron de la siguiente forma:

$$\text{Tasa Tweets por Estado}_{\text{año}} = \frac{\text{Total de Tweets por Estado}}{\text{Población 2020 de la Entidad}} \times 100000 \quad (4.4)$$

$$\text{Tasa Usuarios Twitter por Estado}_{\text{año}} = \frac{\text{Total de Usuarios de Twitter por Estado}}{\text{Población 2020 de la Entidad}} \times 100000 \quad (4.5)$$

La población por entidad se tomó del último censo poblacional realizado por INEGI<sup>3</sup> en el año 2020. Adicionalmente por cada estado de la República Mexicana se obtuvo la tasa de suicidio<sup>4</sup> y para el caso del año 2021 se agregaron los datos de la tasa de depresión por estado, reportado en los resultados de la Encuesta Nacional de Bienestar Autorreportado( ENBIARE)<sup>5</sup>. Esta encuesta es una iniciativa para una vida mejor que mide diversos aspectos sobre la calidad de vida, entre ellos la salud mental considerando indicadores sobre la depresión y ansiedad. Para realizar el análisis de correlaciones se removieron los datos de los estados de Ciudad de México, Aguascalientes, Chihuahua y Yucatán, debido a que fueron considerados datos atípicos (outliers) por tener una alta tasa de suicidio. Los resultados de las correlaciones para el 2021 son presentadas en la Figura 4.12, se puede observar una correlación positiva entre la tasa de usuarios en la clase depresión y la tasa de suicidio, esto significa que a mayor tasa de usuarios y tasa tweets, se tendría una mayor tasa de suicidio. Por otro lado se observa una correlación negativa entre la tasa de tweets, la tasa de cuentas de usuarios y la tasa depresión, esto se puede deber a que la depresión tiene diversos niveles como leve, moderada y alta, y nuestros datos de tweets están relacionados con la depresión.

Con los datos de los estados considerados outliers (Ciudad de México, Aguascalientes, Chihuahua y Yucatán), se estimó la fracción mensual de los tweets de la clase no

<sup>3</sup><https://www.inegi.org.mx/app/tabulados/interactivos/>

<sup>4</sup><https://www.inegi.org.mx/app/tabulados/interactivos/>

<sup>5</sup><https://www.inegi.org.mx/rnm/index.php/catalog/730>

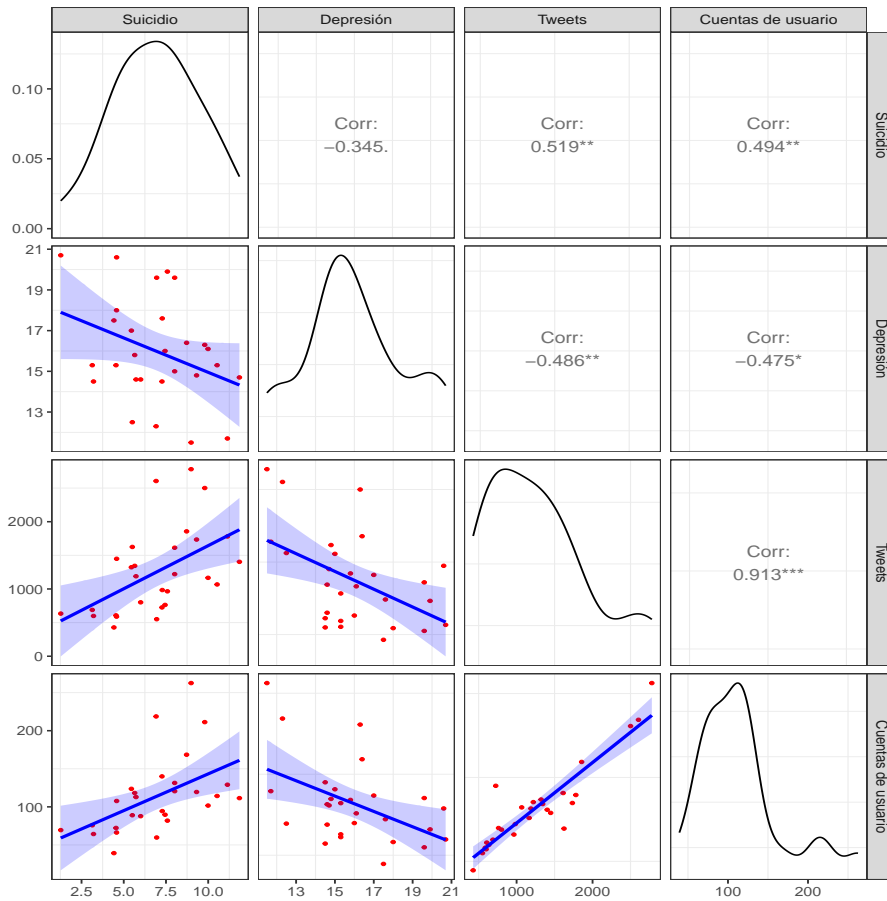


Figura 4.12: Resultados Correlaciones 2021

depresión entre los tweets de la clase depresión. Este resultado fue comparado con el indicador de estado de ánimo basado en tweets y analizados en INEGI<sup>6</sup>. INEGI estima el estado de ánimo de los tuiteros como una división entre los tweets positivos y los tweets negativos. En nuestro caso los tweets positivos corresponden a la clase no depresión y los tweets negativos a la clase depresión. Los resultados son presentados en la Figura 4.13.

En las Figuras 4.13a, 4.13b, 4.13d se puede observar una tendencia similar entre el estado de ánimo y los tweets clasificados con la metodología propuesta. Esto significa que a mejor estado de ánimo hay menos tweets relacionados con la depresión. En la Figura 4.13c correspondiente al estado de Chihuahua no se observa una tendencia similar.

<sup>6</sup><https://inegi.org.mx/app/animotuitero/#/app/multiline>

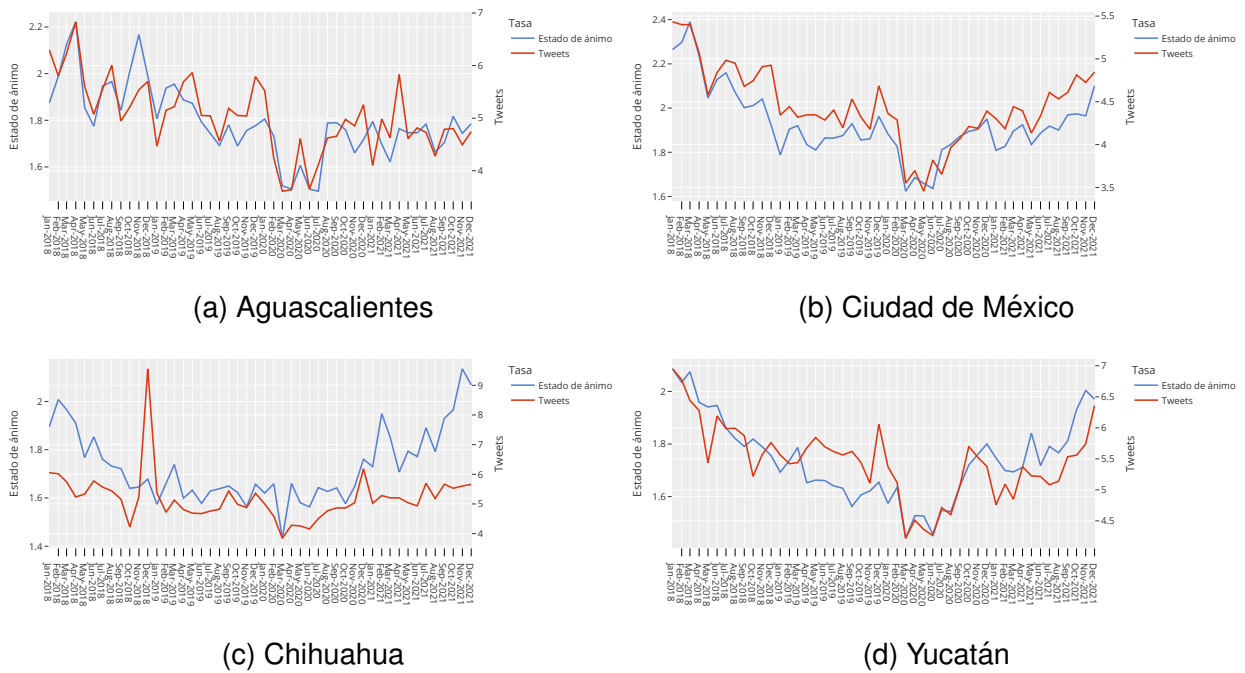


Figura 4.13: Comparación estado de ánimo y tweets de depresión

### 4.3.5. Distribución de palabras

Con el corpus de tweets de la clase depresión y el algoritmo *Term frequency – Inverse document frequency* (TF-IDF) se determinó la importancia de una palabra. A través de TF-IDF se determinó la distribución de las palabras con mayor puntuación en los tweets del período 2018 a 2021.

Para aplicar TF-IDF, los tweets pasaron por un proceso de lematización con el propósito de reducir las variantes morfológicas de una palabra a raíces comunes o lexemas.

La Figura 4.14, muestra para los años 2018 al 2021 los conjuntos de palabras con mayor puntuación. Para el 2018 se observa un sólo conjunto de palabras entre las que destacan *miedo, familia, realidad, terminar*. En 2019 y 2020 aparecen dos posibles grupos de palabras, entre los que aparece *morir, gracias, violencia, apoyo, recordar* y para el año 2021 aparecen palabras relacionadas con la pandemia como *covid, llorar, triste, esperar*.

En la Figura 4.14 se observa el comportamiento de los grupos de palabras con mayor puntuación en los años 2018 a 2021. Cada grupo de palabras se puede asociar a un tema

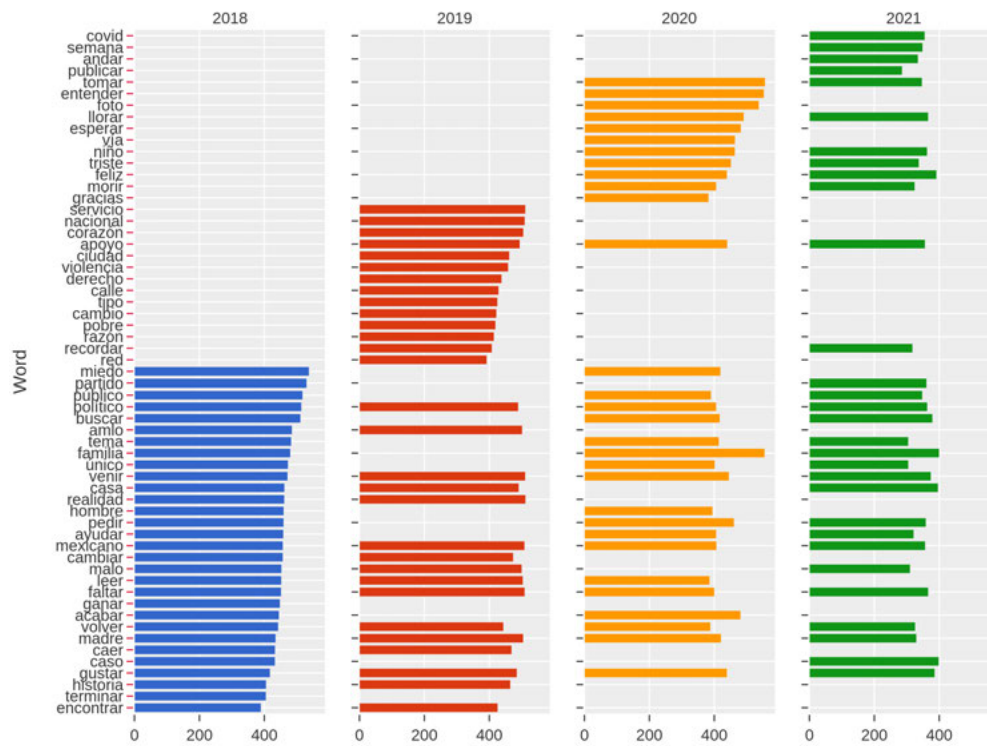


Figura 4.14: Distribución de las palabras con TF-IDF

relacionada con la depresión.

# Capítulo 5

## Conclusiones y trabajo a futuro

### 5.1. Conclusiones

En este trabajo se compararon dos métodos para entrenar clasificadores con el propósito de detectar tweets en idioma español relacionados con la depresión. El primer método consistió en utilizar traducciones de inglés a español y la representación de sentencias a vectores con BETO. El segundo método consistió en utilizar textos en inglés y la vectorización de textos con el modelo KDMV1 preentrenado con la metodología destilación de conocimiento.

A continuación se presentan las conclusiones:

- El clasificador con mejor desempeño se obtuvo a partir del método destilación de conocimiento, con una diferencia del 10 % por encima del mejor clasificador basado en traducciones, debido a la capacidad del modelo KDMV1 para representar sentencias en inglés y español, ya que el uso de traducciones puede generar pérdida de información de contexto de los textos.
- En los experimentos se observó que el algoritmo de reducción de dimensionalidad IVIS contribuye de mejor manera en conservar las propiedades de los datos originales en los vectores reducidos.



- Comparando IVIS con PCA, en IVIS se obtienen mejores resultados debido a que PCA utiliza la varianza como métrica para obtener los vectores reducidos, sin embargo en el caso de textos la varianza no es la mejor opción ya que se las representaciones vectoriales pueden ser esparsos.
- Para el 2020, año en el cual se registraron los primeros casos de covid-19 se encontraron 45597 tweets relacionados a la depresión, esto es aproximadamente un 2% comparado con el año 2019, lo cual significa que se publicaron más tweets relacionados a la depresión. Sin embargo es importante notar que existe una diferencia de aproximadamente 2 millones de tweets entre el total del año 2019 y 2020 debido a que no se tuvieron publicaciones de todos los días del mes de mayo de 2020.
- En Marzo de 2020, todos los estados tuvieron un incremento en la tasa de tweets en la clase depresión, esto puede ser causa de las medidas de aislamiento y cambio en el estilo de vida de la población debido a la pandemia.
- Se encontró una correlación positiva de 0,519 entre la tasa de tweets relacionados a la depresión y la tasa de suicidio reportada por INEGI, se esperaría que a mayor tasa de tweets de depresión mayor sería la tasa de suicidio.
- Se encontró correlación positiva de 0,494 entre la tasa de cuentas de twitter que han publicado tweets clasificados en la categoría depresión y la tasa de suicidio reportada por INEGI. Esta correlación tiene la interpretación de que a mayor tasa de cuentas de twitter en la clase depresión mayor podría ser la tasa de suicidio.
- En los estados de Aguascalientes, Ciudad de México y Yucatán, se encontro para el año 2020 una tendencia similar entre el estado de ánimo reportado por INEGI y la proporción de tweets obtenida entre la clase no depresión y depresión. En este caso los tweets clasificados por INEGE como positivos se interpretan como los tweets clasificados en la clase no depresión y los tweets negativos como los clasificados en la clase depresión.

## 5.2. Trabajo a futuro

El presente trabajo tendría las siguientes opciones a desarrollar en un futuro:

- Construir herramientas como apoyo para el área de salud mental en países a los cuales han emigrado hispanohablantes.
- Aplicar esta metodología para detectar textos relacionados con temas de violencia en redes sociales como el bullying y misoginia.
- Aplicar esta metodología para la detección de textos a temas de salud mental como la ansiedad y la ideación del suicidio.
- A partir de los tweets clasificados en la clase depresión, utilizar grafos para identificar posibles comunidades de usuarios de twitter que estén relacionados con el tema de la depresión.
- Identificar perfiles en twitter que publican tweets clasificados en la clase depresión para identificar rasgos depresivos.
- Analizar el conjunto de textos de la clase depresión para detectar patrones o vocabulario en español para construir un diccionario de palabras relacionados con la depresión en México.
- Analizar tweets en español georreferenciados en las frontera de México con Estados Unidos para abordar la depresión como un fenómeno derivado de la migración.

# Índice de figuras

1.1. Métodos utilizados en el área de salud mental. Fuente: (Zhang et al., 2022)	5
2.1. Linear SVM. Fuente: Wikipedia . . . . .	11
2.2. Embedding de palabras. Fuente: Wikipedia . . . . .	15
2.3. Arquitectura del modelo de atención. Fuente: Aurpa et al., 2022 . . . . .	16
2.4. Mecanismo de atención . . . . .	17
2.5. Visualización de la matriz de atención. Fuente: Bahdanau et al., 2015 . . .	17
2.6. Framework genérico de KD. Fuente: Gou et al., 2021 . . . . .	18
3.1. Metodología para generar los modelos de clasificación . . . . .	22
3.2. Distribución de los embedding . . . . .	26
3.3. Modelos de reducción y entrenamiento de clasificadores . . . . .	27
3.4. Etapa de prueba para los clasificadores . . . . .	29
3.5. Entrenamiento de modelos de reducción de dimensionalidad con traducciones . . . . .	35
3.6. Entrenamiento de modelos de reducción de dimensionalidad con KD . . .	38
4.1. Resultados experimentos traducciones con PCA . . . . .	41
4.2. Resultados experimentos traducciones con IVIS Supervisado . . . . .	42
4.3. Resultados de los experimentos con traducciones e IVIS No Supervisado .	44
4.4. Resultados experimentos traducciones con IVIS Semisupervisado . . . . .	45
4.5. Resultados experimentos KD con PCA . . . . .	47
4.6. Resultados experimentos KD con IVIS Supervisado . . . . .	49

4.7. Resultados experimentos KD con IVIS No Supervisado . . . . .	50
4.8. Resultados experimentos KD con IVIS Semisupervisado . . . . .	52
4.9. Distribución acumulada de tweets en la clase depresión . . . . .	56
4.10. Tasa mensual de tweets de depresión por Entidad . . . . .	57
4.11. Distribución de los tweets en períodos prepandemia y pandemia . . . . .	58
4.12. Resultados Correlaciones 2021 . . . . .	60
4.13. Comparación estado de ánimo y tweets de depresión . . . . .	61
4.14. Distribución de las palabras con TF-IDF . . . . .	62

# Índice de tablas

3.1. Conjuntos de datos para experimentos basado en KD . . . . .	24
3.2. Conjuntos de datos para experimentos con traducciones . . . . .	24
3.3. Datos preprocesados para experimentos con traducciones . . . . .	25
3.4. Datos preprocesados para experimentos con KD . . . . .	25
3.5. Entrenamiento de los modelos de reducción de dimensionalidad . . . . .	27
3.6. Datos generados empleando los modelos de reducción . . . . .	28
3.7. Entrenamiento y prueba para los clasificadores utilizando traducciones . .	30
3.8. Entrenamiento y prueba de clasificadores con KD . . . . .	31
4.1. Traducciones y PCA . . . . .	40
4.2. Traducciones e IVIS Supervisado . . . . .	42
4.3. Traducciones e IVIS No Supervisado . . . . .	43
4.4. Traducciones e IVIS Semisupervisado . . . . .	45
4.5. KD y PCA . . . . .	46
4.6. KD e IVIS Supervisado . . . . .	48
4.7. KD e IVIS No Supervisado . . . . .	50
4.8. KD e IVIS Semi Supervisado . . . . .	51
4.9. Resultados con todas las dimensiones esquema multilingual . . . . .	52
4.10. Conjuntos de tweets de los años 2018-2021 . . . . .	53
4.11. Clasificación de tweets del período 2018-2021 . . . . .	54
4.12. Datos de los Tweets clasificados en México . . . . .	55

# Bibliografía

- Aggarwal, C. C. (2022). *Machine Learning for Text, Second Edition*. Springer. <https://doi.org/10.1007/978-3-030-96623-2>
- Aurpa, T. T., Sadik, R., & Ahmed, M. S. (2022). Abusive Bangla comments detection on Facebook using transformer-based deep learning models. *Social Network Analysis and Mining*, 12. <https://doi.org/10.1007/s13278-021-00852-x>
- Bahdanau, D., Cho, K., & Bengio, Y. Neural machine translation by jointly learning to align and translate [3rd International Conference on Learning Representations, ICLR 2015 ; Conference date: 07-05-2015 Through 09-05-2015]. English (US). En: 3rd International Conference on Learning Representations, ICLR 2015 ; Conference date: 07-05-2015 Through 09-05-2015. 2015, enero.
- Bishop, C. M. (2007). *Pattern Recognition and Machine Learning (Information Science and Statistics)* (1.<sup>a</sup> ed.). Springer. <http://www.amazon.com/Pattern-Recognition-Learning-Information-Statistics/dp/0387310738%3FSubscriptionId%3D13CT5CVB80YFWJE26tag%3Dws%26linkCode%3Dxm2%26camp%3D2025%26creative%3D165953%26creativeASIN%3D0387310738>
- Cañete, J., Chaperon, G., Fuentes, R., Ho, J.-H., Kang, H., & Pérez, J. (2020). Spanish Pre-Trained BERT Model and Evaluation Data. *PML4DC at ICLR 2020*.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational*

- Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171-4186. <https://doi.org/10.18653/v1/N19-1423>
- Gou, J., Yu, B., Maybank, S. J., & Tao, D. (2021). Knowledge Distillation: A Survey. *Int. J. Comput. Vision*, 129(6), 1789-1819. <https://doi.org/10.1007/s11263-021-01453-z>
- Govindasamy, K. A., & Palanichamy, N. (2021). Depression Detection Using Machine Learning Techniques on Twitter Data. *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 960-966.
- Hackeling, G. (2014). *Mastering Machine Learning With Scikit-Learn*. Packt Publishing.
- INEGI. (2021). *Encuesta Nacional de Bienestar Autorreportado*. INEGI.
- Islam, M., Kabir, M., Ahmed, A., Kamal, A., Wang, H., & Ul-Haq, A. (2018). Depression detection from social network data using machine learning techniques [Includes bibliographical references.]. *Health Information Science and Systems*, 6(1), 1-12. <https://doi.org/10.1007/s13755-018-0046-0>
- Jurafsky, D., & Martin, J. H. (2009). *Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition*. Pearson Prentice Hall.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality (C. Burges, L. Bottou, M. Welling, Z. Ghahramani & K. Weinberger, Eds.). 26. <https://proceedings.neurips.cc/paper/2013/file/9aa42b31882ec039965f3c4923ce901b-Paper.pdf>
- Murphy, K. P. (2022). *Probabilistic Machine Learning: An introduction*. MIT Press. [probml.ai](https://probml.ai)
- OMS. (2021). *Mental Health*. World Health Organization.
- OMS. (2022). *Salud Mental y Covid-19: datos iniciales sobre las repercusiones de la pandemia*. Organización Mundial de la Salud.
- Rao, D., & McMahan, B. (2019). *Natural Language Processing with PyTorch: Build Intelligent Language Applications Using Deep Learning*. O'Reilly Media. <https://books.google.com.mx/books?id=NsuEDwAAQBAJ>

- Rasmussen, C. E., & Williams, C. K. I. (2005). *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press.
- Reimers, N., & Gurevych, I. (2020). Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 4512-4525. <https://doi.org/10.18653/v1/2020.emnlp-main.365>
- Sang Eun Woo, R. W. P., Louis Tay. (2020). *Big Data in Psychological Research* (A. P. Association, Ed.). <https://doi.org/http://dx.doi.org/10.1037/0000193-000>
- Szubert, B., & Drozdov, I. (2019). ivis: dimensionality reduction in very large datasets using Siamese Networks. *Journal of Open Source Software*, 4, 1596. <https://doi.org/10.21105/joss.01596>
- Thieme, A., Belgrave, D., & Doherty, G. (2020). Machine Learning in Mental Health: A Systematic Review of the HCI Literature to Support the Development of Effective and Implementable ML Systems. *ACM Trans. Comput.-Hum. Interact.*, 27(5). <https://doi.org/10.1145/3398069>
- Unicef. (2021). *Resumen Regional: América Latina y el Caribe. En mi mente, promover, proteger y cuidar la salud mental de la infancia*. UNICEF.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All you Need. En I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan & R. Garnett (Eds.), *Advances in Neural Information Processing Systems*. Curran Associates, Inc. <https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>
- Zhang, T., Schoene, A., Ji, S., & Ananiadou, S. (2022). Natural Language Processing Applied to Mental Illness Detection: A Narrative Review. *n p j Digital Medicine*.