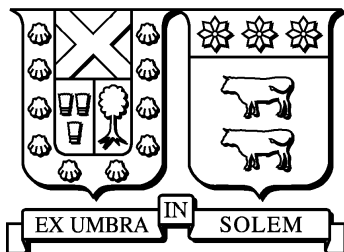


UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA

DEPARTAMENTO DE INFORMÁTICA

SANTIAGO – CHILE



“MODELOS DEL LENGUAJE EN REDES
SOCIALES PARA LA DETECCIÓN DE LA
DEPRESIÓN EN ESPAÑOL”

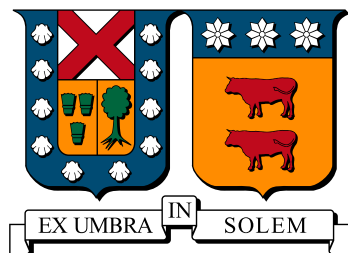
JEAN CARLOS MANUEL GONZÁLEZ ACEVEDO

MEMORIA DE TITULACIÓN PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INFORMÁTICO

PROFESOR GUÍA: MARCELO MENDOZA

DICIEMBRE 2020

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INFORMÁTICA
SANTIAGO – CHILE



**“MODELOS DEL LENGUAJE EN REDES
SOCIALES PARA LA DETECCIÓN DE LA
DEPRESIÓN EN ESPAÑOL”**

JEAN CARLOS MANUEL GONZÁLEZ ACEVEDO

**MEMORIA DE TITULACIÓN PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INFORMÁTICO**

PROFESOR GUÍA: MARCELO MENDOZA

PROFESOR CORREFERENTE: ANDRÉS MOREIRA

DICIEMBRE 2020

MATERIAL DE REFERENCIA, SU USO NO INVOLUCRA RESPONSABILIDAD DEL AUTOR O DE LA INSTITUCIÓN

Agradecimientos

Primero que nada debo hacer un agradecimiento enorme a mis padres, quienes a pesar de ellos mismo haber pasado por momentos muy difíciles, nunca se rindieron o minimizaron la importancia de una buena educación en mi vida. Aún recuerdo los días en que el refrigerador podía estar casi vacío, pero ustedes encontraban la forma de pagarme un colegio y suplir los materiales necesarios para este.

Mi madre que conforma una gran inspiración a este trabajo, pues ella misma me sirve de ejemplo para lo horrible que es sufrir una depresión severa cuando no es diagnosticada a tiempo, enfermedad que comparto con ella, pero que afortunadamente me diagnosticaron relativamente temprano. Mi padre que a pesar de él mismo no haber recibido una educación académica apropiada se esforzó al máximo para que a mí y a mis hermanos nunca nos faltara nada.

Mencionar al profesor Marcelo Mendoza, quien ha tenido una paciencia enorme con las demoras de este trabajo, entendiendo las situaciones personales de cada uno y quien me ha guiado y apoyado dándome material para continuar con este trabajo.

Por supuesto no menos importantes, a todos mis amigos que me han acompañado durante este largo viaje, haciéndome reír y mostrando su apoyo cuando más lo necesité pasando aquellas noches en vela por estudiar alguna materia que ninguno quiso estudiar antes, lo cual nos hacía compañeros en una travesía auto-infligida. Sin embargo, debo hacer una mención especial a Ignacio Tampe, quien me ayudó tanto que sin él este trabajo no hubiera sido posible, pues tuvo la paciencia de escuchar y resolver mis dudas respecto a mí propio trabajo, guiarme cuando estaba perdido e incluso resolver fallas cuando él no tenía por qué hacerlo.

Por último quiero hacer mención a una parte importante de mi vida, mis mascotas, quienes me han dado apoyo emocional y mental a pesar de que ellos mismos no pueden hablar. Wily ,mi gato y compañero de vida quien falleció a sus 16 años de edad, me hubiera gustado que vivieras para este momento. Almendra quien nos dio todo el amor que podríamos querer en sus 13 años de vida, pero que lamentablemente falleció a mitad de mi vida universitaria. Morita, mi shar-pei que llegó a darnos amor en uno de los momentos más difíciles de mi vida y que aún continua acompañándonos. Blair, una gatita quien lamentablemente falleció en un accidente a una muy corta edad, a pesar de que fue poco el tiempo la marca que dejaste en nosotros nunca se olvidará. Finalmente a las dos gatas que nunca paran de jugar o pelear, Emi y Perlita, quienes espero sigan dándonos la oportunidad de amarlas por muchos años más.

Gracias a todos por ser parte de mi vida.

Resumen

Hoy en día la depresión es una de las enfermedades más frecuentes, según la OMS, se estima que afecta a 300 millones de personas en todo el mundo y puede llevar al suicidio siendo esto la segunda causa de muerte de personas entre 15 a 29 años. Más del 70 % de los pacientes no consultan doctores en etapas tempranas de la enfermedad. Además, las personas cada vez comparten más en redes sociales sus emociones y rutinas diarias. Con esto se propone una forma de detectar la depresión de manera inteligente y automática mediante la recolección de posts en redes sociales. En este trabajo se estudia el rendimiento de un modelo propuesto basado en una red neuronal de tipo LSTM que tiene como entrada un tensor con los tweets de un usuario transformados en vectores utilizando word embeddings para palabras en español, mientras que la salida corresponde a la predicción de si el usuario en cuestión tiene depresión o no. Los resultados son comparados con diferentes tipos de modelos que sirven como benchmark para la LSTM propuesta.

Palabras Clave - Salud Mental, Depresión, Modelos de lenguaje.

Abstract

Depression is one of the most frequent diseases, according to the WHO it is estimated that it affects 300 million people worldwide and in the worst cases it can lead to suicide, this being the second cause of death for people between 15 to 29 years old. More than 70 % of the patients will not consult doctors in the early stages of the disease. In addition, people are increasingly sharing their emotions and daily routines on social networks. With this, a way to detect depression in an intelligent and automatic way is proposed by collecting posts on social networks.

Keywords - Mental Health, Depression, Language Models.

Índice de Contenidos

Agradecimientos	III
Resumen	V
Abstract	VI
Índice de Contenidos	VII
Lista de Tablas	X
Lista de Figuras	XIII
Glosario	XIV
Introducción	1
1. Definición del Problema	3
1.1. Procesamiento de Lenguaje Natural (NLP)	3
1.1.1. NLP en redes sociales	4
1.1.2. NLP para idioma español	5
1.2. Objetivos	6
1.2.1. Objetivos Generales	6

1.2.2. Objetivos Específicos	6
2. Estado del Arte	8
3. Metodología	12
3.1. Datos	12
3.1.1. Extracción de los datos	12
3.1.2. Preprocesado de los datos	15
3.2. Arquitectura	17
3.3. Entrenamiento del Modelo	18
3.3.1. Método Experimental	19
3.4. Métrica de validación	19
4. Resultados	21
4.1. Comparación de los modelos entrenados	21
4.2. Resultados de las pruebas	25
4.2.1. Tweets Largos vs Tweets Cortos	25
4.2.2. Sentiment Score, más negativos vs más positivos	26
4.2.3. Cantidad de tweets por usuario	27
4.3. Ejemplos ilustrativos	29
4.3.1. Mejores 5 resultados	29
4.3.2. Peores 5 resultados	33
4.3.3. Resultados adicionales	36
5. Conclusiones	39
Conclusiones	39

6. Trabajo A Futuro	40
Anexo	41
Bibliografía	63

Índice de cuadros

4.1. Resultados obtenidos del entrenamiento para la precisión, val loss y f1-score para los distintos modelos	22
4.2. Comparación de resultados para la precisión, la exhaustividad y el f1-score entre el modelo entrenado con Pyss3 y la LSTM propuesta.	24
4.3. Resultados obtenidos del entrenamiento para la precisión, val loss y f1-score para los distintos modelos para el dataset con los tweets de entrenamiento que poseen la menor cantidad de tokens o palabras	25
4.4. Resultados obtenidos del entrenamiento para la precisión, val loss y f1-score para los distintos modelos para el dataset con los tweets de entrenamiento que poseen la mayor cantidad de tokens o palabras	26
4.5. Resultados obtenidos del entrenamiento para la precisión, val loss y f1-score para los distintos modelos para el dataset con los tweets de entrenamiento que poseen el sentiment score más bajo	27
4.6. Resultados obtenidos del entrenamiento para la precisión, val loss y f1-score para los distintos modelos para el dataset con los tweets de entrenamiento que poseen sentiment score más alto	28
4.7. Resultados para la LSTM con los dataset de pruebas que divide a los usuarios según cantidad de tweets publicados.	28

Índice de figuras

2.1. Fuente: Devlin 2018[6]. Representación de la entrada para BERT.	10
3.1. Ejemplo de un gráfico de entrenamiento para el val loss en una red con un buen ajuste.	20
4.1. Matriz de confusión obtenida a partir del subset de validación para la LSTM	23
4.2. Gráfico del entrenamiento para el perceptron multicapa dónde se muestra el accuracy y el val loss	24
4.3. Gráfico del entrenamiento para la LSTM dónde se muestra el accuracy y el val loss	24
6.1. Matriz de confusión obtenida a partir del subset de validación para el perceptron multicapa	41
6.2. Matriz de confusión obtenida a partir del subset de validación para el modelo de regresión logística	42
6.3. Matriz de confusión obtenida a partir del subset de validación para el modelo de SVM	43
6.4. Matriz de confusión obtenida a partir del subset de test con tweets con menor cantidad de tokens para el modelo de perceptron multicapa	44

6.5. Matriz de confusión obtenida a partir del subset de test con tweets con mayor cantidad de tokens para el modelo de perceptron multicapa	45
6.6. Matriz de confusión obtenida a partir del subset de test con tweets con menor cantidad de tokens para el modelo de regresión logística	46
6.7. Matriz de confusión obtenida a partir del subset de test con tweets con mayor cantidad de tokens para el modelo de regresión logística	47
6.8. Matriz de confusión obtenida a partir del subset de test con tweets con menor cantidad de tokens para el modelo SVM	48
6.9. Matriz de confusión obtenida a partir del subset de test con tweets con mayor cantidad de tokens para el modelo SVM	49
6.10. Matriz de confusión obtenida a partir del subset de test de usuarios con tweets con menor cantidad de tokens promedio para el modelo LSTM . . .	50
6.11. Matriz de confusión obtenida a partir del subset de test de usuarios con tweets con mayor cantidad de tokens promedio para el modelo LSTM . . .	51
6.12. Matriz de confusión obtenida a partir del subset de test con tweets con menor sentiment score para el modelo de perceptron multicapa	52
6.13. Matriz de confusión obtenida a partir del subset de test con tweets con mayor sentiment score para el modelo de perceptron multicapa	53
6.14. Matriz de confusión obtenida a partir del subset de test con tweets con menor sentiment score para el modelo de regresión logística	54
6.15. Matriz de confusión obtenida a partir del subset de test con tweets con mayor sentiment score para el modelo de regresión logística	55
6.16. Matriz de confusión obtenida a partir del subset de test con tweets con menor sentiment score para el modelo SVM	56

6.17. Matriz de confusión obtenida a partir del subset de test con tweets con mayor sentiment score para el modelo SVM	57
6.18. Matriz de confusión obtenida a partir del subset de test de usuarios con tweets con menor sentiment score promedio para el modelo LSTM	58
6.19. Matriz de confusión obtenida a partir del subset de test de usuarios con tweets con mayor sentiment score promedio para el modelo LSTM	59
6.20. Matriz de confusión obtenida a partir del subset de test de usuarios con menor cantidad de tweets para el modelo LSTM	60
6.21. Matriz de confusión obtenida a partir del subset de test de usuarios con mayor cantidad de tweets para el modelo LSTM	61
6.22. Histograma que muestra distribución en la cantidad de usuarios que publican cierta cantidad de tweets	62

Glosario

- **NLP:** Natural Language processing
- **LSTM:** Long Short-Term Memory, una red neuronal artificial utilizada en los campos de la inteligencia artificial y el aprendizaje profundo.
- **MD5:** Algoritmo de hashing, utilizado con el propósito de anonimizar la data.
- **LR:** Logistic Regression
- **SVM:** Support-Vector Machine
- **SS3:** Modelo de aprendizaje automático supervisado para la clasificación de texto

Introducción

Hoy en día las redes sociales forman una parte fundamental en las vidas de las personas, manteniendo conectadas a miles de millones de personas a través del globo. En 2020 la pandemia forzó a las personas a mantenerse en cuarentena y encerradas, mostrando un incremento en el uso de redes sociales como medio de comunicación y entretenimiento. Por minuto, los usuarios internet publican 575000 tweets y realizan 5700000 de búsquedas en google. Para el año 2021, la población de usuarios en internet alcanza los 5,17 mil millones de usuarios, correspondientes al 65 % de la población mundial ¹. Esta continuamente creciente población virtual genera zettabytes de data anual, y analizar la información en forma de texto de este total es un desafío casi imposible realizar para una persona. Sin embargo se crea una oportunidad para el desarrollo y estudio de mecanismos automáticos para extraer, ordenar y analizar esta data con el fin de obtener información relevante.

La construcción de modelos de clasificación en texto han sido tópico de investigación a lo largo de los últimos años. Utilizar este tipo de mecanismos para la salud mental tampoco es nuevo. La conferencia de detección de riesgos en internet, eRisk, explora la metodología de evaluación, las métricas de efectividad y las aplicaciones prácticas (particularmente las relacionadas con la salud y la seguridad) de la detección temprana de riesgos en Internet. El objetivo principal de esta conferencia es ser pioneros en una nueva área de investigación interdisciplinaria que sería potencialmente aplicable a una amplia variedad de situaciones y a muchos perfiles personales diferentes. Los ejemplos incluyen pedófilos potenciales, acosadores, individuos que podrían caer en manos de organizaciones criminales, personas con inclinaciones suicidas o personas susceptibles a la depresión².

¹<https://www.domo.com/learn/infographic/data-never-sleeps-9>

²<https://erisk.irlab.org/>

Estudios para la detección de enfermedades mentales como la depresión utilizando técnicas de NLP han sido desarrolladas en los últimos años principalmente para el idioma inglés. Desde el análisis del lenguaje utilizado por las personas diagnosticadas con PTSD[8] hasta el uso de técnicas como árboles de decisión, Naive Bayes y Perceptrones[2]. Sin embargo, esfuerzos para hacer lo mismo en otros idiomas como el español no han sido tan explorados, un ejemplo de estos son los realizados por Angela Leis et al. en [10] y [9].

Este trabajo se enfoca principalmente en estudiar y comparar los resultados obtenidos al construir un modelo de clasificación para la detección de la depresión utilizando una red neuronal de tipo Long Short-Term Memory (LSTM). Una unidad LSTM común se compone de una celda, una puerta de entrada, una puerta de salida y una puerta de olvido. La celda recuerda valores en intervalos de tiempo arbitrarios y las tres puertas regulan el flujo de información dentro y fuera de la celda. Las redes LSTM son adecuadas para clasificar, procesar y hacer predicciones basadas en datos de series temporales, ya que puede haber retrasos de duración desconocida entre eventos importantes en una serie temporal.

Los aportes específicos de este trabajo corresponden a:

- El entrenamiento de un modelo tipo LSTM, que permite la clasificación de usuarios en la red social twitter a partir de las publicaciones y patrones del lenguaje propios. Esto se logra con la ayuda de word-embeddings que potencian el aprendizaje del modelo en el lenguaje español
- Construcción de un dataset con usuarios que presentan depresión y usuarios que son clasificados como no depresivos, con el fin de expandir el estudio de estas metodologías en español.

El documento está organizado de la siguiente manera. En el capítulo 2 se muestran ejemplos de trabajos relacionados a la detección de enfermedades mentales a partir de técnicas NLP. En el capítulo 3 se muestra cómo es construido el dataset propuesto, los distintos modelos entrenados y el método experimental. Los resultados son discutidos en el Capítulo 4. En el capítulo 5 se presentan las conclusiones del trabajo y finalmente en el Capítulo 6 se discute el trabajo futuro pendiente y posibles mejoras al método experimental.

Capítulo 1

Definición del Problema

La depresión es el resultado de interacciones complejas entre factores sociales, psicológicos y biológicos. Al mismo tiempo, la enfermedad puede generar más estrés y disfunción, lo cual empeora la situación de la persona afectada y, por consiguiente, la propia depresión. En el peor de los casos puede llevar al suicidio, de hecho, cada año se suicidan cerca de 800 mil personas¹. Es por esto que su diagnóstico en etapas tempranas de la enfermedad es crucial para el tratamiento y prevención de fatalidades.

1.1. Procesamiento de Lenguaje Natural (NLP)

Todo lo que expresamos, ya sea de manera verbal o escrita contiene una gran cantidad de información, tal como los tópicos a tratar, estado anímico, selección de las palabras, entre otros proveen información valiosa que puede ser interpretada para entender e incluso predecir el comportamiento de la persona.

El procesamiento del lenguaje natural es un campo de la inteligencia artificial que se enfoca en la interacción entre la ciencia de datos y el lenguaje humano. En términos simples, NLP consiste en metodologías que permiten manejar la forma natural del lenguaje humano, ya sea oral o textual, con el objetivo de obtener información valiosa de manera automática e

¹<https://www.who.int/es/news-room/fact-sheets/detail/depression>

inteligente. Algunos ejemplos de aplicaciones de NLP son:

- Reconocimiento y predicción de enfermedades basadas en registros de salud y forma de hablar del paciente. Esta capacidad para el diagnóstico de condiciones de la salud van desde enfermedades cardiovasculares hasta trastornos mentales como la depresión, PTSD e incluso esquizofrenia.
- Las organizaciones pueden determinar qué es lo que opinan sus consumidores acerca de un producto o servicio que ofrecen mediante la identificación y extracción de información al respecto en las redes sociales.
- Google assistant, Alexa de Amazon y Siri de Apple son ejemplos de interfaces inteligentes basadas en voz que utilizan NLP para responder a estímulos vocales y realizar las acciones solicitadas por la persona.

1.1.1. NLP en redes sociales

Las redes sociales proveen una oportunidad de expandir los datos disponibles para personas que se dedican a la investigación sobre la salud mental. Tal y como la palabra lo indica, estas redes pueden mostrar patrones sociales, lo cual es crítico a la hora de entender enfermedades mentales.

La creciente popularidad de las redes sociales ha incrementado de manera dramática la cantidad de publicaciones por segundo que se realizan, incluyendo detalles personales y actividades diarias. Todos los estados, tweets, comentarios, posts y reviews pueden ser catalogados como “contenido generado por usuario”[7] y típicamente se refiere a imágenes, texto y vídeos que provienen de personas regulares y no están necesariamente en un formato estandarizado. Por lo tanto la calidad de los datos generados varían desde muy buenos hasta de muy pobre calidad, debido a la naturaleza difusa y desorganizada de las redes sociales. Sin embargo, toda esta data puede incorporar la opinión, comportamientos y pensamientos de los usuarios, lo cual hace que sea muy valioso el extraer información a partir de esto.

Varios estudios indican que el lenguaje utiliza expresiones sociales e interacciones que sirven como indicadores sobre la salud mental de una persona. El Linguistic Inquiry Word Count (LIWC), desarrollado por Pennebaker et al. y utilizado para estudiar el significado psicológico de las palabras en [11], es un estándar en el análisis de texto que permite analizar de manera efectiva cómo el lenguaje que utilizamos revelan pensamientos, sentimientos, personalidad y motivaciones.

Las aplicaciones más comunes del análisis de datos de las redes sociales son el descubrimiento de tendencias, analíticas, opiniones y sentiment analysis con el objetivo de obtener feedback acerca de productos o servicios que entregan las organizaciones relacionadas, de esta manera se logran tomar decisiones de negocio en base al análisis de estas redes.

1.1.2. NLP para idioma español

Una limitante importante de los estudios sobre NLP es que la mayoría de las técnicas se enfocan principalmente en texto y publicaciones en inglés, lo cual acota la investigación a personas que puedan hablar este idioma, mientras que técnicas de clasificación y modelos del lenguaje en español aún son tema de estudio y desarrollo.

La presencia del español en el internet ha incrementado a lo largo de los años y las redes sociales se han convertido en una parte importante de la vida diaria de las personas. Trabajos como el de Anta et al[1] intentan encontrar una forma para incluir al español en análisis de texto utilizando técnicas de Machine Learning para el Sentiment Analysis y Topic Detection, pues recordemos que es común encontrar elementos, abreviaciones y modismos específicos del idioma.

1.2. Objetivos

1.2.1. Objetivos Generales

Creación de un modelo del lenguaje en redes sociales para la detección de la depresión en idioma español con el uso de redes neuronales secuenciales, que permiten evaluar cómo van evolucionando las publicaciones de las personas estudiadas.

1.2.2. Objetivos Específicos

Creación del dataset

Lo primero y más importante para el desarrollo de este trabajo es la construcción del dataset, el cual al final de este trabajo se planea publicar para su uso en investigaciones similares, teniendo siempre en cuenta la privacidad de los usuarios. Para ello las principales fuentes de datos son las redes sociales:

- **Facebook:** Facebook es una de las redes sociales más populares en todo el mundo y con usuarios diversos respecto al rango etario e idioma, lo cual la hace perfecta para obtener publicaciones en español. Se planea obtener los posts de usuarios que estén en grupos donde el tema principal sea la salud mental, por ejemplo, grupos y páginas de ayuda mutua para personas que sufran de depresión.
- **Reddit:** Reddit es actualmente una de las redes sociales más populares con más de 300 millones de miembros. Sin embargo a diferencia de otras redes, reddit se enfoca en comunidades llamadas **subreddits** en donde las personas postean texto, fotos, vídeos o links, además la comunidad puede comentar los posts para continuar la conversación. El problema que tiene esta red social para el desarrollo de esta memoria es que es principalmente habitada por usuarios que hablan en inglés, por lo que encontrar subreddits en español que traten sobre salud mental podría ser difícil.
- **Twitter:** Es una red social que funciona con “microblogs”, donde las personas pueden

postear por tópicos o tipos utilizando hashtags. Al igual que facebook, twitter es una red social que cuenta con publicaciones diarias en diferentes idiomas incluyendo español. Esto y teniendo en cuenta la accesibilidad de la red a la hora de obtener datos, la hace una fuente importante de información para el dataset que se quiere construir. Se realiza la recolección de tweets mediante web scrapping utilizando la librería de python **Twint**, la cual permite recolectar más twits que la API oficial de la red social accedida mediante **TwipPy**, la cual está limitada al 1 % de los tweets para queries muy amplias.

Arquitectura y modelo del lenguaje

Para desarrollar el modelo del lenguaje, son necesaria una serie de pasos para eliminar el ruido del dataset en caso que no se pueda conseguir uno ya construido como en [10], y entrenarlo sin que se produzca el problema de overfitting o underfitting.

- Desarrollar una arquitectura para la red que permita resolver el problema.
- Entrenar la red con los datos obtenidos y evaluar los resultados. Si los resultados no cumplen con la calidad suficiente, torcer parámetros y modificar la arquitectura para obtener mejores resultados.
- Comparar resultados con modelos del estado del arte para evaluar mejoras.

Capítulo 2

Estado del Arte

Uno de los primeros acercamientos en el análisis del lenguaje para la salud mental basado en redes sociales fue desarrollado por De Choudhury et al.[2], donde desarrollaron un clasificador para evaluar auto revelaciones (*self-disclosure*), es decir, posts que revelan información o experiencias personales a otros. Los datos son extraídos de la red social **reddit**, específicamente subreddits relacionados a la salud mental. Evalúan diferentes métodos de clasificación tales como, arboles de decisión, *k-Nearest Neighbor*, *Naive Bayes* y Perceptrones, este último siendo el que tuvo mejor desempeño.

Coppersmith et al. tiene varias publicaciones que tratan sobre la salud mental utilizando redes sociales para obtener los datos necesarios. En su trabajo publicado el 2014 [8], proponen un método que permite la clasificación y medición del Desorden de Estrés Post Traumático (PTSD), mediante el estudio de las diferencias del lenguaje utilizado por personas con PTSD e individuos aleatorios en la red social Twitter. Utilizan los post y comentarios de estas personas para entrenar un clasificador de tipo *Unigram Language Model (ULM)*, que examina palabras completas de manera individual, y un modelo basado en las categorías del LIWC. Los resultados muestran que utilizar solamente LIWC puede ser insuficiente para la clasificación de usuarios.

En otro trabajo de Coppersmith et al.[5] analizan el lenguaje de la salud mental mediante auto-diagnósticos. Obtienen sus datos de la red social Twitter, donde las condiciones que

las personas dicen tener son *Attention Deficit Hyperactivity Disorder (ADHD)*, *Generalized Anxiety Disorder (Anx)*, *Bipolar Disorder*, *Borderline Personality Disorder*, *Depression*, *Eating Dissorder* (anorexia, bulimia y otros no especificados), *Obsessive Compulsive Disorder (OCD)*, *PTSD* y Esquizofrenia. Utilizan para las mediciones LIWC, del cual se obtienen resultados interesantes, los cuales indican que si existe diferencias en el lenguaje utilizado por personas con diferentes trastornos. Dado que LIWC solo puede cubrir una fracción del lenguaje, utilizan también un acercamiento con vocabulario abierto CLM.

Sin embargo los estudios anteriores tienen como limitante que están basados en post exclusivamente escritos en inglés. Por ello, los estudios más importantes para esta memoria son aquellos desarrollados por Angela Leis et al. [10] y [9], donde desarrollan un *dataset* y métodos para detectar signos de depresión en Tweets en español. Analizan los comportamientos y aspectos lingüísticos. Analizando la distribución de Tweets a través del tiempo, los resultados muestran que las personas con depresión tienden a postear menos en la red social. Además modifican sus patrones lingüísticos a temas más negativos, utilizan más verbos, adverbios y pronombres, pero menos sustantivos. Se cree que la pobre estructura de las personas con la enfermedad se debe al poco interés en aquello que los rodea y se enfocan más en acciones y los temas que tratan están relacionados a la auto exposición, como lo ocurrido en [2].

Modelos pre-entrenados son efectivos para mejorar el rendimiento de varias tareas relacionadas al procesamiento de lenguaje natural. El 2018 Devlin et al.[6] desarrollan BERT: *Bidirectional Encoder Representations from Transformers*. BERT utiliza *masked language model (MLM)*, el cual enmascara de manera aleatoria algunos de los *tokens* de la entrada (figura 2.1) con el objetivo de predecir el id de vocabulario de la palabra enmascarada solamente en base al contexto, esto permite el entrenamiento de *Transformers* bidireccionales. BERT *fine-tuning* es evaluado con el *General Language Understanding Evaluation (GLUE) benchmark*, el cual consiste en una colección de varias tareas de NLP, entre las cuales se encuentran:

- **Natural Language Inference:** Dado un par de sentencias, donde la primera es llamada premisa y la segunda la hipótesis. La tarea consiste en predecir si la premisa respalda, contradice o es neutral a la hipótesis.

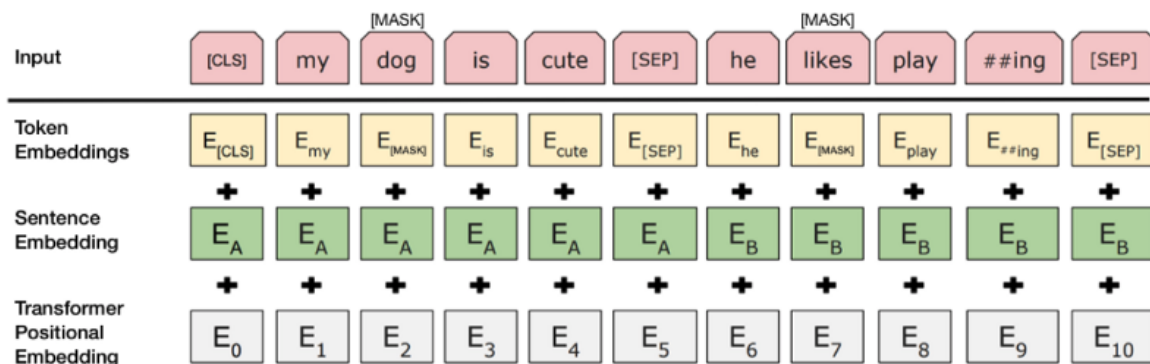


Figura 2.1: Fuente: Devlin 2018[6]. Representación de la entrada para BERT.

- **Paraphrasing:** Consiste en determinar si dos frases dadas son semanticamente equivalentes o no.
- **Named Entity Recognition:** Etiquetar el contenido de frases, reconociendo nombres de personas, organizaciones, lugares, cantidades, entre otros.
- **Part-of-Speech Tagging:** Categorías de palabras con propiedades gramaticales similares, tales como verbos, adjetivos, sustantivos, preposiciones y conjunciones.
- **Document Classification:** Tarea de clasificar documentos en categorías como industrial, economía, social, entre otros.
- **Dependency Parsing:** Construcción de un árbol de dependencia para una oración.
- **Question Answering:** Dado el contexto y una pregunta, la tarea consiste en encontrar la secuencia de palabras que permite responder a la pregunta.

Otra cosa importante a tener en cuenta son los modelos de lenguaje pre-entrenados para el español como es el caso del modelo BERT desarrollado por Cañete et al.[4], el cual es entrenado con un corpus que contiene texto de diferentes fuentes en español como lo es Wikipedia, artículos de la ONU, *TED talks*, subtítulos, entre otros.

Finalmente, un clasificador de texto reciente ha obtenido buenos resultados en las tareas propuestas en la eRisk, conferencia que se centra principalmente en temas de salud mental y detección de estas condiciones mediante técnicas de NLP. Sin embargo, poca atención se le

había puesto a este modelo hasta que Sergio Burdisso et al. presenta una implementación de carácter Open-Source para este modelo en lenguaje python. SS3 construye un diccionario para almacenar frecuencias de palabras usando todos los documentos de entrenamiento pertenecientes a la categoría. Este método de entrenamiento permite que SS3 soporte el aprendizaje online, ya que cuando se agregan nuevos documentos de capacitación, SS3 simplemente necesita actualizar los diccionarios utilizando el contenido de estos nuevos documentos, lo que hace que el entrenamiento sea incremental. Luego, usando las frecuencias de palabras almacenadas en los diccionarios, SS3 calcula un valor para cada palabra usando una función, $gv(w, c)$, para valorar palabras en relación a categorías. gv toma una palabra w y una categoría c y entrega un número entre $[0, 1]$ el cual representa la confianza con la cual se cree que w pertenece exclusivamente a la clase c [3]. El proceso de clasificación de este algoritmo se divide en dos fases:

- La entrada es dividida en múltiples bloques y estos son nuevamente seccionado en unidades más pequeñas hasta llegar a palabras. Creándose así una jerarquía de bloques a partir de un documento plano.
- En la segunda fase se aplica la función gv a cada palabra creando un set de vectores de confianza los cuales son reducidos para crear el vector de confianza para frases. Esta reducción se aplica recursivamente y es propagada a los bloques superiores hasta que solo queda un único vector de confianza para el documento.
- Finalmente, la clasificación es realizada aplicando una política basada en los valores de confianza para el vector final.

Capítulo 3

Metodología

3.1. Datos

Los datos son directamente sacados desde la red social Twitter¹. Las redes sociales proveen valiosa información pues proviene de personas comunes y corrientes con un formato no estructurado del cual se pueden rescatar patrones del lenguaje relacionados a la enfermedad. Las publicaciones de twitter son sacados directamente de la página utilizando la librería de Python *Twint*, la cual consiste en una herramienta para el *scraping* de twits sin la necesidad de utilizar la API de Twitter, lo cual la hace conveniente pues no tiene limitantes con la fecha de publicación ni cantidad de tweets. Para cumplir con los protocolos de privacidad, la identidad de los usuarios es anónima. El dataset está totalmente disponible para su uso en otras investigaciones.

3.1.1. Extracción de los datos

Para construir el dataset de entrenamiento con individuos diagnosticados con depresión se recolectan tweets que tengan las keywords “Diagnosticaron depresión”, pues esto hace altamente probable que el tema relacionado al post sea de cuando se diagnostica la enfermedad

¹<https://twitter.com/>

al individuo y relate sus experiencias al respecto. Luego, manualmente aquellos usuarios que hayan sido diagnosticados con depresión, de esta forma se filtran aquellos tweets que contengan comentarios de broma, hablen sobre vídeos musicales o el diagnostico de terceros. Otro dato a mencionar debido al contexto social, es que los tweets son obtenidos en el primer semestre de 2021, específicamente entre el primero de enero del 2021 y primero de junio del 2021, esto es importante pues durante el periodo de pandemia el tema de la salud mental se hizo más popular en redes sociales.

Para obtener los posts relacionados a la enfermedad de aquellos usuarios diagnosticados se usan como base las keywords obtenidas de [10].

Lista de palabras relacionadas con la depresión, obtenidas desde Leis et al.[10]:

- | | |
|----------------|-----------------|
| ■ agobiado/a | ■ desesperado/a |
| ■ agotado/a | ■ desmotivado/a |
| ■ angustiado/a | ■ insomnio |
| ■ ansiedad | ■ llorar |
| ■ cansado/a | ■ nervioso |
| ■ decaído/a | ■ preocupado/a |
| ■ depresión | ■ solo/a |
| ■ depresivo/a | ■ triste |
| ■ deprimido/a | ■ vacío/a |
| ■ desanimado/a | |

Todos los tweets obtenidos mediante scraping con estas keywords son considerados y marcados como “depresivos”. Al mismo tiempo los usuarios autores de estos tweets también son considerados como “depresivos”. Esto servirá para el entrenamiento de las redes neuronales utilizadas en este trabajo.

El volumen total de los datos corresponden a 35896 tweets catalogados como depresivos y

un subset del mismo tamaño de los datos recolectados para los no depresivos, obtenidos de manera aleatoria de un total de 1213056 tweets catalogados como no depresivos. Este subset se crea debido al problema del balanceo de datos, pues los usuarios con depresión fueron elegidos de forma manual, lo cual provoca que la cantidad de posts de usuarios depresivos sea mucho menor que para usuarios no depresivos.

Para el grupo considerado como “No depresivo” se utiliza la ley de Zipf para scrapear los tweets en español de manera aleatoria. Con las palabras en español con un largo mayor a tres caracteres más comunes en ley de Zipf se logra obtener un dataset con menor ruido provocado por palabras o tweets completos en otro idioma como el inglés.

Palabras en español con un largo mayor a tres caracteres más comunes en ley de Zipf:

- | | |
|-----------|-----------|
| ■ tambien | ■ vez |
| ■ fue | ■ puede |
| ■ habia | ■ todos |
| ■ era | ■ asi |
| ■ muy | ■ tiene |
| ■ hasta | ■ parte |
| ■ desde | ■ donde |
| ■ esta | ■ ahora |
| ■ porque | ■ despues |
| ■ solo | ■ otro |
| ■ han | ■ aunque |
| ■ hay | ■ hace |

Todos los tweets del grupo no depresivo son marcados como tal. Dado que los tweets son obtenidos de manera aleatoria, el autor del tweet no es marcado directamente como no depresivo, si no que se realiza un preprocesamiento a grupos de 100 tweets los cuales serán considerados como un usuario.

3.1.2. Preprocesado de los datos

La data extraída desde twitter debe ser limpiada debido al uso de emojis, caracteres y palabras no pertenecientes al idioma. Para ello se utiliza la librería Cucco ², con la cual se eliminan todas las palabras con caracteres ajenos al lenguaje español con el objetivo de minimizar palabras de idiomas parecidos como lo es el portugués o el italiano y tweets de personas que ponen palabras en idiomas como el japonés o chino.

Para el procesamiento de los datos se necesitan dos formas de pre-procesar la data: “Agrupada” y “No Agrupada” por usuario y es importante la anonimización de los datos para ambas formas.

La data agrupada por usuario se refiere a que la catalogación del tweet depende de si el usuario ha sido catalogado como depresivo, esto con el objetivo de tener un conjunto de tweets que correspondan a alguien diagnosticado con la enfermedad, independiente si el contenido del post corresponde con las características generales de un tweet depresivo o no. Por otro lado en la data no agrupada por usuario, los tweets en cuestión están catalogados como depresivos o no, en base a si el usuario que lo postea ha sido diagnosticado con depresión y si este contiene las keywords obtenidas del estudio realizado por Leis et al [10]. Esta agrupación se realiza debido a que la LSTM propuesta en este trabajo utiliza el **conjunto de tweets** publicados por el usuario como input, mientras que el resto utiliza el **tweet por si solo**, independiente del usuario.

Anonimización de los datos

Para anonimizar la data agrupada se elimina toda referencia a usuarios incluso en el mismo tweet, reemplazando cada uno por “@usuario”. Además se eliminan los hashtags de la data. Además se elimina la columna que contiene el nombre de usuario e identificador.

Para la data agrupada por usuario es también necesario anonimizar a los usuarios creadores del post. Esto se logra eliminando del dataset su nombre de usuario y se deja solo su identificador, el cual pasa por la función de hashing MD5. Finalmente a cada hash se reemplaza por

²<https://github.com/davidmogar/cucco>

un identificador numérico partiendo desde el 0. De esta forma se hace fácil la agrupación de estos y se vuelve imposible trackear al usuario en base al identificador.

Dado que los tweets no depresivos son obtenidos de manera aleatoria, se realiza una agrupación de 100 tweets, lo cual representa un usuario en el dataset no depresivo, es decir, no son usuarios como tal si no que grupos de tweets a los que se les ha catalogado como tweets de un usuario.

Para los usuarios depresivos se realiza un preprocesamiento similar, subdividiendo los tweets de cada usuario en grupos de 100 y estableciendo cada grupo como un usuario único. Si la cantidad de tweets del usuario no es un múltiplo de 100, como por ejemplo, 616 tweets, dicho usuario quedaría dividido como 7 usuarios, 6 usuarios de 100 tweets y un usuario con 16 tweets.

Limpieza de Datos

Pre-procesamiento de Data No Agrupada

Para los datos no agrupados existen dos datasets, “Depresivos” y “No depresivos” que contienen simplemente la data no agrupada y anonimizada. Además contienen un tag que permite saber si el tweets es de carácter depresivo o no depresivo. Luego ambos datasets se juntan en un único dataset y se realiza un shuffle a los tweets, de tal manera que queden mezclados los tweets de ambos grupos.

Finalmente los datos no agrupados se subdividen en tres datasets: **Train**, **Validation** y **Test** en una relación 70:15:15, los cuales serán utilizados para entrenar las redes neuronales utilizadas en este trabajo.

Pre-procesamiento de Data Agrupada

En los datos agrupados también existen dos datasets, “Depresivos” y “No depresivos”. Sin embargo estos datasets es de simplemente tweets y tags, si no que también incluye el id

numérico del usuario al que corresponde cada tweet. Además el tag no representa si el tweet es depresivo o no, si no que si el usuario que lo escribió está catalogado como depresivo o no depresivo. De esta manera es más fácil agruparlos por usuario y dividir la data por usuario en vez de por tweet. Se juntan ambos datasets y se realiza un shuffle por usuario (a diferencia del anterior que era por tweet).

Finalmente los datos agrupados se subdividen en tres datasets: **Train**, **Validation** y **Test** en una relacion 70:15:15. Este dataset está dividido por usuario en vez de tweets, el cual será utilizado para entrenar la red neuronal de tipo Long Short-Term Memory utilizada en este trabajo.

3.2. Arquitectura

Para vectorizar la data se utilizan los Spanish Word Embeddings³ resultantes de entrenar con el dataset Spanish Unannotated Corpora, el cual contiene 3 mil millones de palabras disponible en GitHub⁴, utilizando los parametros por defecto de fastText para la tarea Skipgram exceptuando las epochs, las cuales están definidas para 20 en vez de 5⁵. Sea $x_0...x_m$ las palabras de un tweet, para poder entrenar los modelos debemos construir un tensor utilizando los Spanish Word Embeddings. Los vectores de dicho tensor se construyen de la siguiente manera:

Sea $V_0...V_n$ los vectores del tensor y sea W_{emb} la función que transforma la palabra en vector de 300 dimensiones, el vector resultante es

$$v_i = \sum_{k=0}^m W_{emb}(x_k), \quad i \in 0...n$$

$$V_i = \frac{v_i}{||v_i||}, \quad i \in 0...n$$

³<https://github.com/dccuchile/spanish-word-embeddings>

⁴<https://github.com/josecannete/spanish-corpora>

⁵<https://github.com/BotCenter/spanishWordEmbeddings>

De esta manera, el vector resultante del tweet corresponde a la sumatoria de los vectores de cada palabra incluida en el tweet dividido por la norma del vector. El tensor corresponde al conjunto de todos los tweets del usuario vectorizados.

Para el entrenamiento se vectorizan todos los tweets de cada usuario y se guardan en el tensor.

3.3. Entrenamiento del Modelo

Para entrenar los modelos, primero los tweets en los dataset deben ser vectorizados utilizando los word-embeddings entrenados con Spanish Unannotated Corpora [4] con los siguientes parámetros:

- Implementación: FastText con Skipgram
- Parameters:
- min subword-ngram = 3
- max subword-ngram = 6
- minCount = 5
- epochs = 20
- dim = 300
- el resto de parametros están por defecto

Modelo	Densidad de la Capa oculta	Epochs	Dropout
Perceptron Multicapa	32	150	0.5
Regresión Logística	-	-	-
Support Vector Machine	-	-	-
Long Short-Term Memory	8	150	0.5

Para los primeros tres modelos la entrada corresponde al total de tweets vectorizados no agrupados por usuario y el label correspondiente a si el tweet es de carácter depresivo o no-depresivo.

Por otro lado, para el modelo Long Short-Term Memory (LSTM), la entrada corresponde al vector de los tweets agrupados por usuario y el label correspondiente a si el usuario en cuestión está diagnosticado con depresión o no.

Para el entrenamiento del Perceptron Multicapa y la LSTM, se utiliza la *binary cross entropy loss function*, de tal manera que los resultados digan si el tweet o el usuario en cuestión es de carácter depresivo o no-depresivo.

3.3.1. Método Experimental

Debido al tamaño del dataset, los modelos pueden ser entrenados utilizando el ambiente básico de Google Colab. Se realizan experimentos con cuatro distintos modelos para comparar los resultados se toman los resultados con menor *val loss* para cada uno de los modelos.

Para los experimentos cada uno de los modelos recibe como entrada un subset de los datos que no fueron utilizados en el entrenamiento, correspondiente al 15 % de los tweets totales para el caso del perceptron multicapa (PM), regresión logística (LR) y la Support Vector Machine (SVM).

Para el caso de la LSTM, el modelo recibe como entrada un subset de los datos no utilizados en el entrenamiento, el cual corresponde al 15 % de los usuarios totales.

3.4. Métrica de validación

Para validar y comparar el rendimiento y los resultados entregados por los modelos y las instancias utilizadas para el testing se utiliza el F1-Score, el cual combina la precisión y la exhaustividad (*recall*) en una sola métrica calculando la media armónica. Se espera que los modelos con buen rendimiento obtengan un F1-score cercano a 1.

Sea P , R la precisión y la exhaustividad del modelo respectivamente, el F1-Score de un modelo de clasificación se calcula de la siguiente manera:

$$F1\text{-score} = 2 \cdot \frac{P * R}{P + R}$$

Además durante los entrenamientos de los modelos de perceptron multicapa y Long Short-Term Memory, se espera un comportamiento en el gráfico de val loss como la figura 3.1

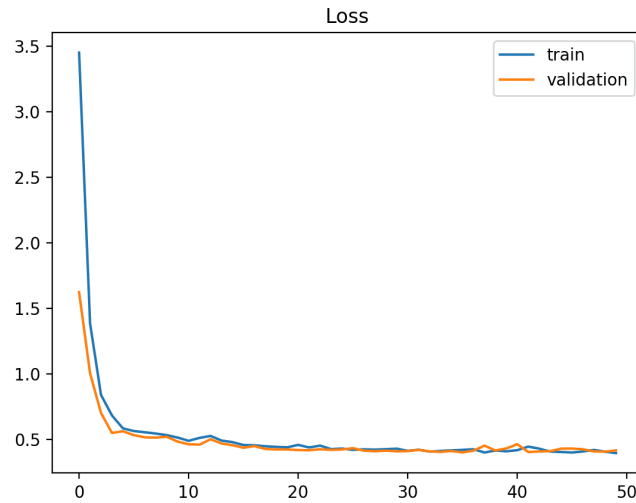


Figura 3.1: Ejemplo de un gráfico de entrenamiento para el val loss en una red con un buen ajuste.

Capítulo 4

Resultados

En este capítulo se evalúa el rendimiento de cada uno de los modelos en cuestión en su habilidad de clasificar los tweets o los usuarios como depresivos o no-depresivos en base al *val loss*. Luego se discuten los resultados comparando los modelos con mejor y peor resultado y se realizan experimentos con distintos datasets de prueba agrupados según cualidades de los tweets.

4.1. Comparación de los modelos entrenados

Se muestra en la tabla 4.1 los resultados obtenidos del entrenamiento de cada uno de los modelos utilizados. Se comparan cuatro modelos distintos y la implementación en código de pyss3 [3], los cuales sirven como benchmark para la LSTM, modelo seleccionado debido a que es superior en las tres métricas utilizadas en la comparación respecto a los demás modelos.

Para la comparación con Pyss3[3] con la LSTM se presenta la tabla 4.2 donde se observan los resultados para la precisión, la exhaustividad y el f1 score de cada modelo entrenado. Los hiperparámetros utilizados para el modelo SS3 son los provistos por defecto:

- σ (suavidad): 0,45. Este valor permite suavizar la relación entre frecuencia y el valor

Modelo	accuracy	val loss	f1 score
Perceptron Multicapa	0.9307	0.1966	0.9313
Regresión Logística	0.9015	-	0.9005
Support Vector Machine	0.9043	-	0.9035
Long Short-Term Memory	0.9869	0.0629	0.9926

Cuadro 4.1: Resultados obtenidos del entrenamiento para la precisión, val loss y f1-score para los distintos modelos

final asignado a la palabra.

- λ (importancia): 0,5. Controla que tanto el valor local de una categoría debe diferenciarse del valor local de otra categoría para que la palabra sea considerada importante para esa categoría.
- ρ (Penalización): 1,0. Permite disminuir el valor global en relación a la cantidad de categorías en las que una palabra es relevante. Este valor controla que tan sensible es la penalización.

Además se modifica el parámetro α para obtener un mejor resultado con este modelo. Este controla el valor mínimo global que una palabra debe tener para ser relevante y no ser ignorada por el modelo. Con esto se evita que se le ponga atención a palabras que no pertenezcan al lenguaje, es decir, palabras en otro idioma o jerga propia de alguna comunidad o nacionalidad.

Para el caso de la LSTM se calcula la precisión y la exhaustividad utilizando los datos obtenidos en la matriz de confusión (figura 4.1) con la siguiente fórmula:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

Donde TP , FP y FN corresponden a los Verdaderos positivos, Falsos positivos y Falsos negativos de la matriz de confusión respectivamente.

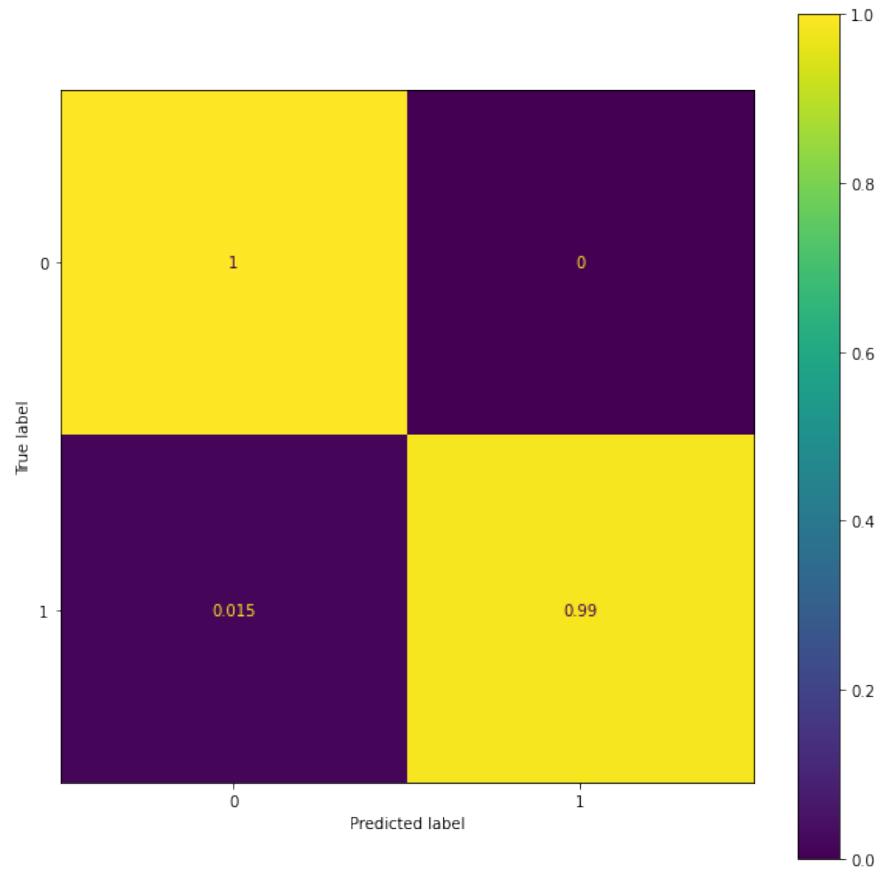


Figura 4.1: Matriz de confusión obtenida a partir del subset de validación para la LSTM

Se observa en la tabla 4.2 que la LSTM es superior al modelo entrenado utilizando Pyss3 respecto a las 3 métricas anteriores incluso con el hiperparámetro α modificado, por lo que se sigue evaluando el desempeño de este modelo para los distintos casos propuestos.

Se presenta además en las Figura 4.2 y 4.3 el comportamiento del modelo respecto a la precisión y val loss durante el entrenamiento, las cuales representan curvas de aprendizaje bien ajustadas notando que en la LSTM la curva es ligeramente mejor debido a que esta alcanza un mayor accuracy, menor val loss y las curvas se ajustan apropiadamente al final del entrenamiento. Notar además que la LSTM necesita una menor cantidad de epochs para alcanzar un ajuste apropiado en comparación al perceptrón multicapa.

Modelo	Precision	Recall	F1 Score
LSTM	1.0000	0.9852	0.9926
SS3 ($\alpha = 0,1$)	0.8629	0.8532	0.8580
SS3	0.7733	0.9366	0.8471

Cuadro 4.2: Comparación de resultados para la precisión, la exhaustividad y el f1-score entre el modelo entrenado con Pyss3 y la LSTM propuesta.

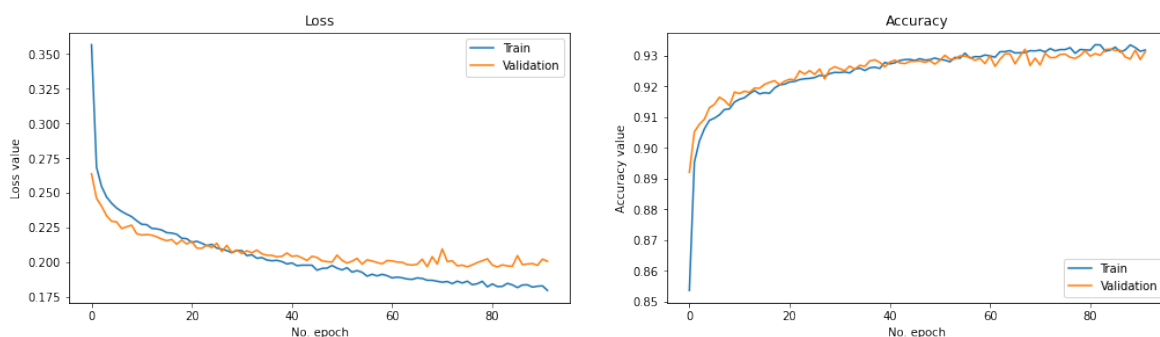


Figura 4.2: Gráfico del entrenamiento para el perceptron multicapa dónde se muestra el accuracy y el val loss

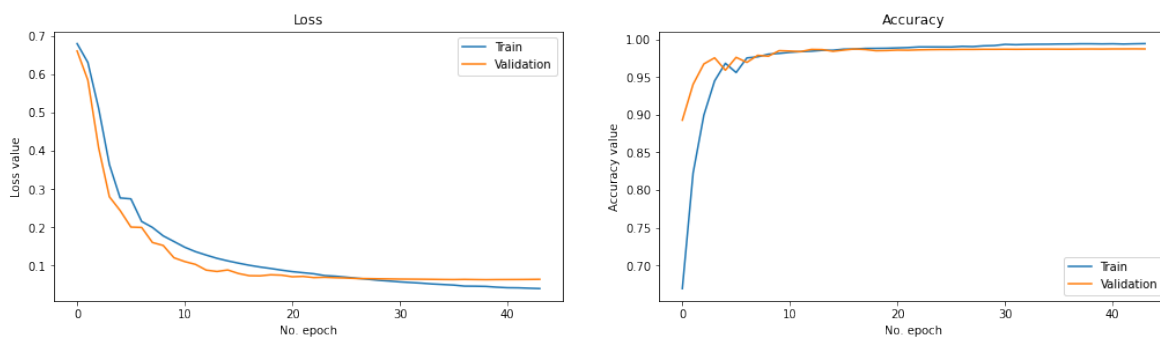


Figura 4.3: Gráfico del entrenamiento para la LSTM dónde se muestra el accuracy y el val loss

4.2. Resultados de las pruebas

A continuación se presentan los resultados obtenidos para cada uno de los datasets de pruebas contruidos a partir de las características de los tweets recopilados, siendo estas características el Largo (cantidad de tokens o palabras por tweet), el sentiment score y la cantidad de tweets publicados por usuario.

4.2.1. Tweets Largos vs Tweets Cortos

Modelo (cortos)	accuracy	val loss	f1 score
Perceptron Multicapa	0.9370	0.1726	0.9360
Regresión Logistica	0.9089	-	0.9067
Support Vector Machine	0.9106	-	0.9084
Long Short-Term Memory	0.9950	0.0194	1.0000

Cuadro 4.3: Resultados obtenidos del entrenamiento para la precisión, val loss y f1-score para los distintos modelos para el dataset con los tweets de entrenamiento que poseen la menor cantidad de tokens o palabras

La tabla 4.3 muestra los resultados para los modelos cuando la entrada tiene la mitad de los tweets más cortos del dataset de entrenamiento. Los resultados muestran que la LSTM es nuevamente la que mejor se desempeña en las tres métricas, seguido del perceptrón multicapa. Además los resultados muestran que los modelos se mantienen consistentes respecto a la tabla 4.1 puesto que la diferencia en desempeño no es significativa siendo esta aproximadamente $< 0,005$ para la mayoría de los modelos, siendo la excepción la regresión logística el cual presenta una diferencia de 0,0062 respecto a la tabla 4.1, sin embargo esto sigue sin ser un desempeño significativamente mayor.

Para el caso de la tabla 4.4 se muestran los resultados de los modelos cuando la entrada corresponde a la mitad del set de entrenamiento que contiene los tweets con la mayor cantidad de tokens. Nuevamente la LSTM es la que mejor se desempeña tanto en accuracy, val loss y

Modelo (largos)	accuracy	val loss	f1 score
Perceptron Multicapa	0.9275	0.2036	0.9287
Regresión Logística	0.8982	-	0.8969
Support Vector Machine	0.8984	-	0.8970
Long Short-Term Memory	0.9787	0.0924	0.9836

Cuadro 4.4: Resultados obtenidos del entrenamiento para la precisión, val loss y f1-score para los distintos modelos para el dataset con los tweets de entrenamiento que poseen la mayor cantidad de tokens o palabras

f1-score. Cabe destacar que el desempeño baja para todos los modelos, siendo la LSTM la más afectada, pues tiene una diferencia para el f1-score de 0,0090 con respecto a los resultados en la tabla 4.1.

Con esto, los resultados muestran que el largo o cantidad de tokens en los tweets no afecta de manera significativa en el desempeño de los modelos. Las matrices de confusión obtenidas para estos datasets se pueden encontrar en el Anexo, figuras 6.6, 6.7, 6.8, 6.9, 6.10, 6.11.

4.2.2. Sentiment Score, más negativos vs más positivos

En la tabla se comparan los resultados obtenidos comparando el desempeño de los modelos con datasets con los tweets con sentiment score más alto y más bajo respectivamente, obtenido a partir de la librería para python 3 sentiment analysis spanish creada por Hugo Bello¹.

De la tabla 4.5 se observa nuevamente una mejora en los f1-scores de todos los modelos > 0,005, siendo el más notable el score para la LSTM siendo este de 1,0000 con una mejora respecto a los resultados de 4.1 de 0,0074. Esto nos indica que para los modelos es más fácil predecir si el usuario en cuestión tiene depresión o no cuando los tweets tienen un sentiment score bajo, es decir, cuando los tweets son de índole negativa. Las matrices de confusión para este dataset de pruebas se encuentran en el Anexo, figuras 6.14, 6.16, 6.18

¹<https://github.com/sentiment-analysis-spanish/sentiment-spanish>

Modelo (negativos)	accuracy	val loss	f1 score
Perceptron Multicapa	0.9280	0.2029	0.9364
Regresión Logística	0.8977	-	0.9076
Support Vector Machine	0.8983	-	0.9079
Long Short-Term Memory	0.9977	0.0198	1.0000

Cuadro 4.5: Resultados obtenidos del entrenamiento para la precisión, val loss y f1-score para los distintos modelos para el dataset con los tweets de entrenamiento que poseen el sentiment score más bajo

Si se comparan los resultados con la tabla 4.6 se observa un deterioro notable en el rendimiento de los modelos, especialmente la LSTM donde el f1-score baja su puntuación 0,0769 puntos. Esto nos indica que los modelos tienen una gran dificultad prediciendo la depresión del usuario cuando las publicaciones tienen un sentiment score alto, es decir, son de carácter en su mayoría positivos. Sin embargo, cabe destacar que la precisión del modelo LSTM sigue siendo mayor al resto a pesar del bajo f1-score.

Las diferencias de resultados para ambos datasets de prueba son esperables puesto que, en general, las personas con depresión tienden a publicar, en su mayoría, tweets que presentan palabras con un sentiment score negativo, por lo que si las publicaciones tienen un sentiment score positivo debiera ser más difícil detectar la depresión en base al lenguaje utilizado produciendo falsos negativos tal y como se aprecia en las figuras 6.13, 6.15, 6.17 y 6.19 del Anexo.

4.2.3. Cantidad de tweets por usuario

A continuación se muestran los resultados del desempeño para la LSTM para el dataset de pruebas cuando este se divide en función de la cantidad de tweets que se tiene por usuario. El dataset en este caso está dividido a la mitad, pero no respecto a los tweets, si no más bien respecto a la cantidad de usuarios en este estudio. Sin embargo, se muestra en la figura 6.22 la distribución en la cantidad de tweets que existen por usuarios, mostrando que la mayoría

Modelo (positivos)	accuracy	val loss	f1 score
Perceptron Multicapa	0.9364	0.1733	0.9270
Regresión Logística	0.9094	-	0.8943
Support Vector Machine	0.9107	-	0.8961
Long Short-Term Memory	0.9760	0.0920	0.9231

Cuadro 4.6: Resultados obtenidos del entrenamiento para la precisión, val loss y f1-score para los distintos modelos para el dataset con los tweets de entrenamiento que poseen sentiment score más alto

entre los 0 y 100 tweets, habiendo outliers que tienen publicado entre 3300 y 3450 tweets.

Es por lo anterior que se toma la decisión de crear subusuarios a partir de aquellos que tienen más de 100 tweets. Para esto, cada 100 tweets se considera que es un usuario distinto y lo que sobra es considerado como un usuario con menos de 100 tweets. Para efectos prácticos ayuda a que el modelo entrene con usuarios que tienen una cantidad consistentes de tweets y se evitan problemas de memoria a la hora de realizar el padding a los usuarios con pocos tweets, pero además crea una limitación a la hora de evaluar el desempeño de la LSTM, pues como resultado, la mitad del dataset con usuarios con mayor cantidad de tweets, está compuesto por usuarios con exclusivamente 100 tweets. Mientras que la otra mitad está compuesto por usuarios que tienen una cantidad de tweets ≤ 100 .

Dataset	accuracy	val loss	f1 score
Usuarios con 100 tweets	0.9942	0.0193	1.0000
Usuarios con ≤ 100 tweets	0.9795	0.0925	0.9846

Cuadro 4.7: Resultados para la LSTM con los dataset de pruebas que divide a los usuarios según cantidad de tweets publicados.

Los resultados experimentales de la tabla 4.7 muestran el desempeño de la LSTM para los datasets que posee exclusivamente usuarios con 100 tweets y el dataset que posee usuarios con ≤ 100 tweets. Se observa que la LSTM tiene un mejor desempeño cuando los usuarios postean una gran cantidad de tweets, mostrando diferencias significativas respecto a aquellos

que tienen 100 o menos publicaciones puesto que la precisión difiere por 0,0147, el val loss en 0,0732 y el f1-score en 0,0154 puntos. Esto es esperable, puesto que para la entrada de la LSTM se debe realizar un padding previo de tal manera que todos los usuarios tengan la misma cantidad de tweets vectorizados, esto se logra relleno con vectores nulos el tensor.

4.3. Ejemplos ilustrativos

A continuación se presentan como ejemplo los 5 mejores y peores resultados para la LSTM para el dataset de validación, con el objetivo de mostrar las características de los tweets de aquellos usuarios con los que la red predijo con mucha certeza y aquellos con los que el modelo tuvo dificultades para clasificar.

Estos resultados se obtienen del valor absoluto entre la resta del promedio del número entre $[0, 1]$ que entrega el modelo LSTM para cada uno de los tweets de un usuario y el valor real esperado. De esta forma, valores cercanos a 0 representan aquellos en los que la red tuvo menos problemas para clasificar, y aquellos con valores cercanos a 1 son los valores en que la red tuvo mayores problemas.

Cada usuario se presenta en conjunto a su Id en el dataset, el valor obtenido de la diferencia entre el valor real de la clasificación (entre 0 o 1 si el usuario tiene depresión o no) y el valor obtenido del modelo en la sigmoide, y el resultado esperado del modelo.

4.3.1. Mejores 5 resultados

UserId: 576, **Diferencia con Sigmoide:** 0,0050, **Depresión:** True

- Estoy en una carrera que de verdad no es de mi agrado de un poquito y hoy una amiga me preguntó esto y les juro que estaba a nada de llorar no saben la impotencia que siento al estar aquí un lugar que no quiero una carrera que no me gusta el venir a la escuela me hace sentir mal httpstcoWUzBrBvIci

- Mi mamá me llamó por teléfono desesperada que si en donde estaba y yo así de estoy en mi cuarto vino a comer y se durmió y jamás se enteró que yo estaba en la casa nmms jajaja
- Puse que me quería morir en una historia en CF e Instagram me puso esto nadie se había preocupado tanto por mi httpstcoswDBhkgLN9
- No llevo ni 20 minutos hablando con él y ya me hizo llorar httpstcod4QYAYbILQ
- En cuanto empiecen a cantar The Hanging Tree voy a llorar
- que he llegado a tener cada día por sentirme inútil cada día al llegar de la escuela al cruzar la puerta de mi casa es como si todo en mi se apagara la risa la alegría todo todo se apaga al punto de que la depresión me come por dentro
- user Casi siempre me toca el mismo conductor y es el único que cuando le digo Gracias me dice Que tenga un muy bonito día me dan ganas de llorar lt3

UserId: 609, **Diferencia con Sigmoide:** 0,0067, **Depresión:** True

- Perdí una semana de estudio por andar triste A veces odio tener sentimientos
- Estoy harta de estar preocupada por entregar un tp a horario harta de preparar finales harta de estudiar para parciales harta de estar estresada y harta de renegar con profesores En fin harta de la facultad A esta altura ya no aguanto
- El novio de mi amiga está tan preocupado por el parto que ya me preguntó el grupo sanguíneo por si llegara a necesitar que alguien le done sangre jajaja
- Desde que el psicólogo me dijo que tengo permitido llorar cada vez que lo sienta lloro por cualquier cosita jaaja No saben lo liberador que es
- Mi mamá ya notó que estoy al borde de la depresión y me pidió que hable igual no puedo hablar
- No quiero estudiar quiero estar en la cama tirada y deprimida escuchando música bien triste

- Todas y todos preocupados por que no llegan al verano y yo aqui preocupada por que no llego a los finales

UserId: 612, Diferencia con Sigmoide: 0,0067, Depresión: True

- Hoy me acordé que me decías Princesa Grumosa y me puse a llorar poquito en el metro Hasta aquí mi reporte
- No puedo dejar de llorar algún día acaba esta miseria llamada MI VIDA
- user SOBREVIVO POR PURA ANSIEDAD CON UN NUDO EN LA GARGANTA Y ES QUE NO TE DEJO DE PENSAAAAAR
- Nunca he comprado boletos en preventa y me pone nerviosa perder los de morrissey mañana por estar TODA IDIOTA
- Me está empezando a gustar alguien que al parecer también le gusto y siento que no me hace tanto caso pero luego sí y me pone nerviosa HELP
- Tener depresión o ansiedad no es estar loquito no se quita solito no es solo para llamar la atención
- Yo hoy va a ser un gran día miraré todo con filosofía Que hermoso el cielo hola señor Sol Ansiedad y depresión Hahahaha LOL NO

UserId: 665, Diferencia con Sigmoide: 0,0070, Depresión: True

- La ansiedad me está matando y estoy que me pego el pique al aeropuerto para esperar a los niños de Foster
- Estaba todo bien hasta que abrió la boca Juro que estoy a punto de gritarle a alguien o de ponerme a llorar inserte emoji de pistola
- Estoy triste malhumorada chata de todo Ojalá el semestre se terminara ahora
- Llegue a mi casa tan echa pico que ahora solo quiero llorar hasta que todos los virus de mi cuerpo salgan de mí en puras lágrimas

- Justo el día que estoy preocupada de hacer lo de métodos el cable da dos de mis películas favoritas BTTF III y Gremlins
- Que triste es ver los comerciales de Movistar en el cable donde Alexis dice que está en modo mundial Ya no campeón ya no
- Se acabaron las solemnes y no se si reír o llorar Creo que la segunda es una mejor opción

UserId: 573, **Diferencia con Sigmoid:** 0,0071, **Depresión:** True

- creo que el peor síntoma de la ansiedad es ese nudo horrible en la garganta que no te deja tragar y el dolor de pecho como si se te saliera el corazón lpm que asco
- no estoy pudiendo dormir estoy teniendo ataques de ansiedad re seguidos y no hay quien me hable sin que me moleste y me sienta peor lpm
- ni siquiera me animo a nombrarla no me sale me inunda el dolor en el alma me duele el pecho la necesito no puedo recordarla sin sentirme vacía me hace tanta falta
- Jjajajaja me cago de risa con el tweet del cigarrillo JAJAJJAJAJAJAJAAJA es que si re triste [httpstcoST9zLOuKD1](http://stcoST9zLOuKD1)
- user DIOS NO ME ALCANZA EL TIEMPO MUJER DEJAME HABLAR NO VES QUE ESTOY DESESPERADA ahre bueno basta
- mi psicóloga está tan acostumbrada a que en todas las sesiones me cague llorando que compró un paquete grande de pañuelos descartables Y CLARO si sabia que iba a llorar por decirme que tengo un estado emocional dependiente espantoso que ni siquiera yo puedo controlar
- estoy feliz porque está re fresco y llueve pero seguro esa felicidad me dure 2min hasta que alguien me rompa las bolas y con lo sensible que estoy me ponga a llorar todo el día como de costumbre

4.3.2. Peores 5 resultados

UserId: 674, **Diferencia con Sigmoide:** 0,9765, **Depresión:** True

- user Mi insomnio y yo tenemos una relación sólida y de momento no nos queremos separar
- No sé si estuve más nervioso en el examen de estadística y microeconomía o ahora viendo el sorteo del mundial
- user tan preocupada estai si por eso hay gente q repite voluntariamente pa subir el nem
- user user user si quiero llorar mejor me pego una patada en el dedo guaton mas rapido ajajajajajajajajja
- Mañana es la psu Ayer todos preguntaban estai nervioso la verdad es que casi nada tan nervioso como cuando das una prueba comun P
- Primero los ronquidos luego las pesadillas episódicas y ahora el insomnio U de mierda yo estaba sano cuando entré
- Hoy me diagnosticaron depresión Igual era obvio pero alguien me tenía que decir

UserId: 667, **Diferencia con Sigmoide:** 0,0968, **Depresión:** True

- Si no he estado xon depresión he estado enferma o con insomnio quitándome las energías
- Después de recientes acontecimientos y de poner mi cerebro a pensar si era ansiedad u otra cosa A parte de asexual soy sexrepulsed NICE Bueno gracias a esta introspección me va a bajar el nivel de ansiedad respecto a ese tema uwu
- Más que nada porque me he echado tantas horas con ataques de ansiedad por no ser capaz de memorizar los temas que no me he dado cuenta de que también tenía que leer los libros Ole yo Suspenso automático Pa qué estudiar más

- Qué buena mañana se ha quedado para Ignorar completamente el horario de estudio que hice ayer y seguir con mi racha de depresión y ansiedad
- El ataque de ansiedad que me dio después es cosa mía que él obviamente no podía saber de mi enfermedad pero DDD
- Me siento tontísima por no saber hacer los deberes y las ganas de llorar con las que llevo dos semanas no ayudan y llevo desde que me levanté con ganas de vomitar <http://pstcoyPT4f8bWWU>
- Es que cuán vacía está vuestra vida para tener que ir en masa a insultar a autores y artistas por hacer su trabajo

UserId: 584, **Diferencia con Sigmoide:** 0,0718, **Depresión:** True

- user Encuentro tan penca todo lo que rodea a su historia la familia dice q andaba triste y débil y nadie la ayudaba Le decían q no tenían tiempo pa ella La hmna decia q hablaban todos los días y aun así no tiene idea de la hermana
- Yo tenía la esperanza de que la encontraran viva que mala en verdad Que triste
- Últimamente lamentablemente user ha decaído ene Ahora un loco me acepto el viaje y esta como a 20 minutos de distancia Por qué Porque el otro uber me canceló al ver q iba a valpo
- Que chucha tienen que ver los exámenes si yo me siento inexplicablemente triste y sin ganas pero dsps me dijo q era para descartar otras cosas la cosa es que dsps de tener los exámenes en mano y unas cuantas sesiones más mi diagnostico fue claro depresión severa
- Yo ese día asumí q llevaba muchas horas en el turno cachai de hecho le dije si estas cansada dime Yo voy a entender y me dice filo filo filo chao y yo
- Hoy estoy super cansada no me hubiese levantado
- En mi edificio un perrito llora todos los días y a toda hora aun no puedo saber que piso es pero es cuatico y triste

UserId: 569, Diferencia con Sigmoid: 0,0609, Depresión: False

- Me parece una locura que se juegue el partido ahí cuando hay una realidad tremenda afuera Los de la conmebol están enfermos
- No soporto más esta mufa encima q tengo 21 años de mala leche la puta q me parió
- Este tipo choco a mis padres en las calles Mendez y 50 bis escapo del lugar por favor el municipio de user y la policía se hagan cargo de buscarlo Azucena Cordoba está en la sala 28 del hospital zonal
- Giovanna é a maior fofoqueira de todos os tempos gente KKKKKKKKKK
- user Meu sonho era que ela fosse no estilo 20102011 Mas tá indo no efeito Mod Sun Infelizmente as vezes a vibe dela vai de acordo com os namorados
- user La estupidez es algo contra lo que no se puede batallar
- user user user user user user la llamada CLASE MEDIA jamás se a vendido por un bono los llamados vulnerables si lo han echo pero por el hambre de sus hijos si hemos tenido la oportunidad de retirar nuestro dinero es por que no queremos más afps y el bono universal seria solo la devolución de lo que deben

UserId: 598, Diferencia con Sigmoid: 0,0402, Depresión: True

- Mañana a medianoche no habrá ningún regalo para mí debajo del árbol Que triste y decepcionante es mi vida
- Iré a ver la grabación que tengo del World Stage de Coldplay y a llorar
- Nada más triste que ir al cine sola
- user user yo ellos caería en depresión es decir si YO YA ESTOY EN DEPRESIÓN
- No ha pasado NADA y yo estoy super nerviosa
- Mi ansiedad por aprender idiomas nunca para Ya me inscribí en alemán y francés

- La vida es injusta luego para llorar con el live de Harry Potter

De los ejemplos anteriores es interesante destacar el caso del usuario 674 quien tiene una diferencia promedio con el output del modelo de 0,9765 lo cual indica que el modelo tuvo problemas para clasificar la gran mayoría de los tweets de este usuario. Se cree que esto es debido al tipo de tópicos que trata el usuario en comparación a los mostrados en los 5 mejores resultados, quienes se observa tienden a publicar contenido de carácter personal y sus estados anímicos. Las diferencias en el lenguaje utilizado por este usuario, y en general para los falsos negativos mostrados en los 5 peores resultados, hace que el modelo propuesto tenga dificultades en clasificarlo como un usuario con depresión. Esto concuerda con lo establecido en [10], donde se observa que los usuarios con depresión tienden a publicar contenidos asociados a la auto-exposición.

Es interesante además analizar el caso para el usuario 569 quien presenta un ejemplo para un resultado de falso positivo donde se observa una gran cantidad de términos en otros idiomas como el portugués y palabras y abreviaciones pertenecientes propias a la jerga popular de cada país o idioma. Se conjetura que esto afecta al modelo de manera negativa impidiendo obtener resultados limpios debido al ruido producido por la diversidad de tweets y palabras. Esto muestra una de las limitantes del modelo estudiado y de la data presentada.

4.3.3. Resultados adicionales

UserId: 650, **Diferencia con Sigmoide:** 0,0079, **Depresión:** False

- pls nunca m dejes niisan eres muy válido y vales muxho para mi3
- cuántos hombres están cayéndose a pedazos por lo del millón de dólares y la mujer más linda del mundo no puede ser que ninguno haya respondido bien
- Ha Sido una semana super pesada que no acaba desde el martes de la semana pasada Ya ganamos el primer round vamos por el segundo
- Los compadres se quitan las caretas y se juntan para el baile lastimosamente el ganador

de la fiesta no se da cuenta que le pueden boicotear la misma Los que se daban golpes en el pecho contra el Correísmo y su corrupción donde están

- Yo ya hacía esto hasta en presencial una visionaria
- user todos nós amamos essa tia
- Besties hay que seducirnos en Instagram o se fresean Este es mi perfil I vote for on user

UserId: 579, Diferencia con Sigmoide: 0,0080, Depresión: True

- user esta en rehabI por depresión y cansancio mental creo
- user yo ayer ni sabia I andaba sin internet y supe por una amiga esta mañana Fuck que depresión me entra D
- Es difícil tratar de pensar en otra cosa que no sea Dougie está en rehab por depresión
- Hayley Williams la mujer mas cariñosa y preocupada que puede existir
- Tengo hambre mareo ganas de ir al baño dolor de cabeza y ansiedad Que mas me pueden pedir
- Mañana será como el 1008 mi mama se despide se pone a llorar y yo normal
- Estaré en depresión hasta que vea un tweet o alguna información de Dougie I

UserId: 571, Diferencia con Sigmoide: 0,0099, Depresión: False

- user jugando jaja fuiste muy bueno
- user Me parece bien que te expreses Me gusta Pero una cosa importante dejemos de repetir lo que dice pap y mamá Revelate se combativa ya habrá tiempo de tener 50 años cuando tengas 50 años Ah y sali a la calle a caminar preocúpate por el otro y vas a disfrutar más Saludos

- user Hola user por qué candidato a concejal recomiendas votar Saludos y que te vaya muy bien este finde
- No se si lo veas ferchis pero sólo quiero recordarte que te amo muchísimo y que siempre voy a estar pa ti hasta mi último día y que estoy muy orgullosa de verte cumplir cada proyecto y meta nunca dudes ni olvides cuanto te amo user
- user user user user user Y tú cerebro No hay necesidad de que te metas con su cuerpo enferma de mrd
- user user user Exacto mira nomás tu foto con un chingo de filtros y ni así te ves bien
- user Exactamente asi se divide caballito flores y parque chacabuco

Capítulo 5

Conclusiones

Este trabajo tuvo por objetivo la construcción de un modelo de clasificación para la detección de depresión en twitter en base a las publicaciones hechas por un usuario, utilizando técnicas NLP y un modelo LSTM en conjunto con word-embeddings que permiten potenciar el modelo cuando se trabaja en español. El principal desafío de este es la construcción de un dataset apropiado para el entrenamiento del modelo y la validación de resultados.

De todos los modelos presentados, el propuesto por este trabajo muestra ser el que presenta mejores resultados, esto se conjetura puede ser debido al sistema de olvido que tienen las LSTM a la hora de construir el clasificador, lo cual minimiza los efectos de slangs y patrones del lenguaje que resultan ser propios de cada usuario.

El uso de publicaciones hechas por usuarios presenta desafíos a la hora de construir y limpiar el dataset, puesto que lenguajes como el español comparten palabras y términos similares con el portugués y el inglés, lo cual hace difícil el filtrado de estos idiomas en el dataset. Sin embargo, el uso de word-embeddings en conjunto con algún modelo clasificador de texto muestra ser apropiado para minimizar los efectos negativos que pueden tener apariciones de palabras en idiomas que no se quieren estudiar, pues resultados muestran ser incluso mejores cuando lo comparamos con modelos en el estado del arte, específicamente SS3, que presentan resultados favorables en conferencias como la eRisk.

Capítulo 6

Trabajo A Futuro

Las principales limitaciones de este trabajo vienen de la construcción y limpiado del dataset obtenido, donde palabras e incluso tweets completos en idiomas como el inglés y portugués ensucian la data propuesta. Esto puede ser logrado implementando técnicas que permitan detectar y filtrar idiomas no correspondiente al que se intenta estudiar.

Otro gran problema que existe en NLP, y que aún debe sr estudiado para idiomas como el español, es la inclusión de slangs y términos utilizados propios de alguna comunidad, nacionalidad o grupo en particular. El eliminar estos tokens puede comprometer la data pues se pierde información contextual valiosa a la hora de analizar patrones del lenguaje. Un problema importante a destacar son aquellas palabras que pueden tener multiples significados dependiendo de la nacionalidad del usuario posteador.

El manejar la data en español o idiomas similares ayudaría en la construcción de modelos que reconozcan patrones del lenguaje e incluso nacionalidad del usuario. Permitiendo una temprana detección de enfermedades mentales como lo es la depresión.

Anexo

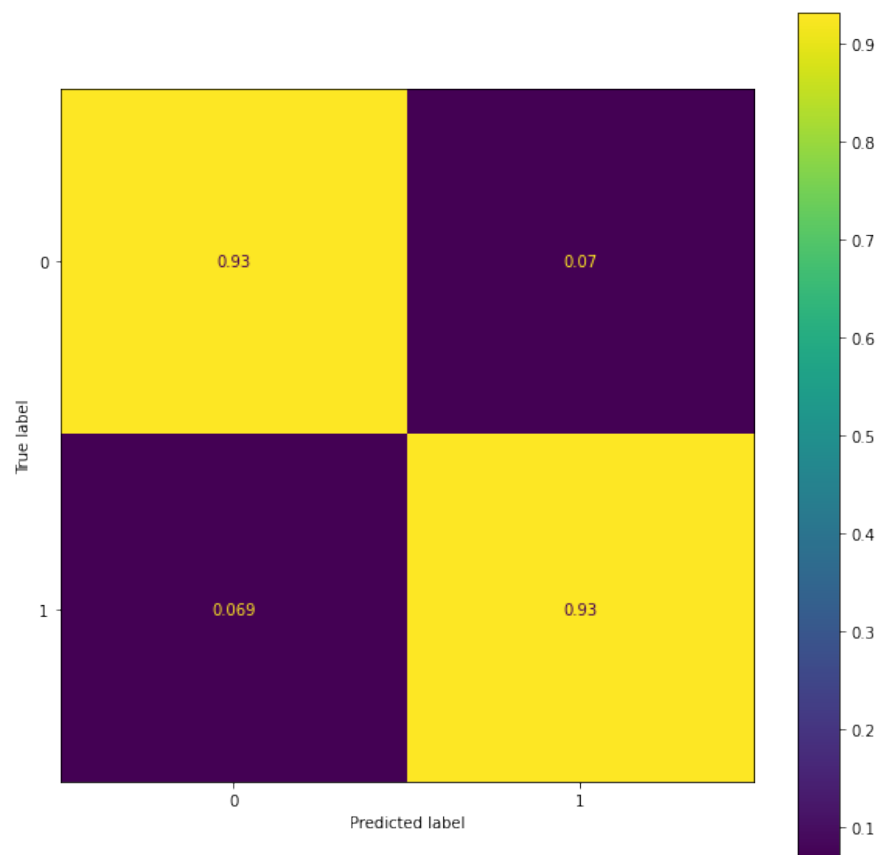


Figura 6.1: Matriz de confusión obtenida a partir del subset de validación para el perceptron multicapa

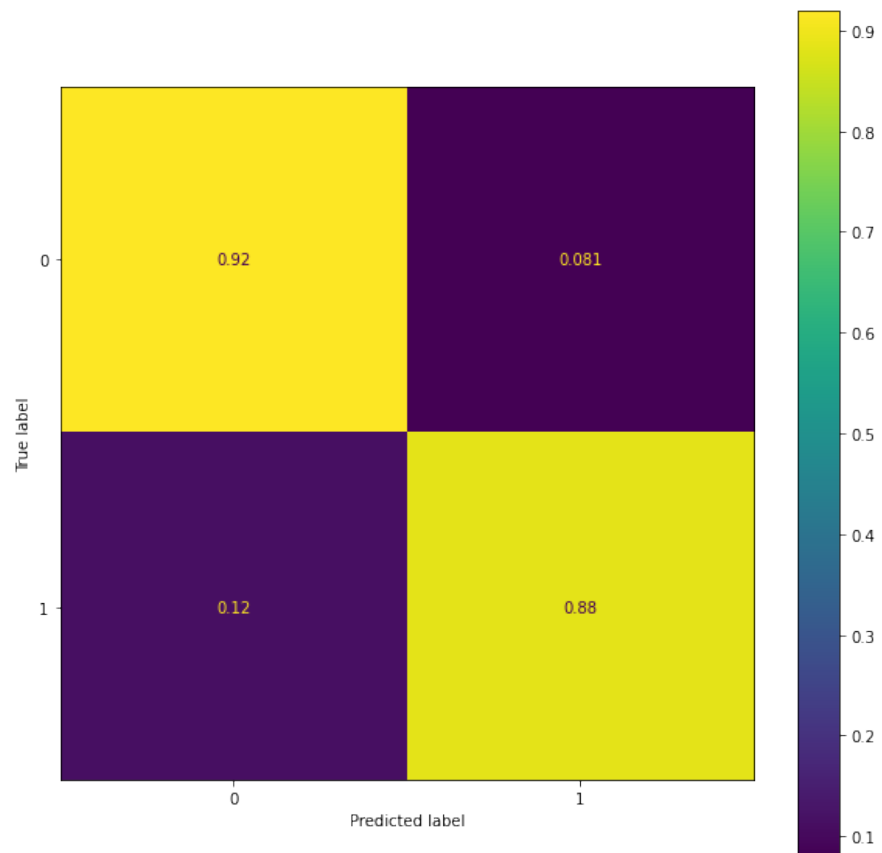


Figura 6.2: Matriz de confusión obtenida a partir del subset de validación para el modelo de regresión logística

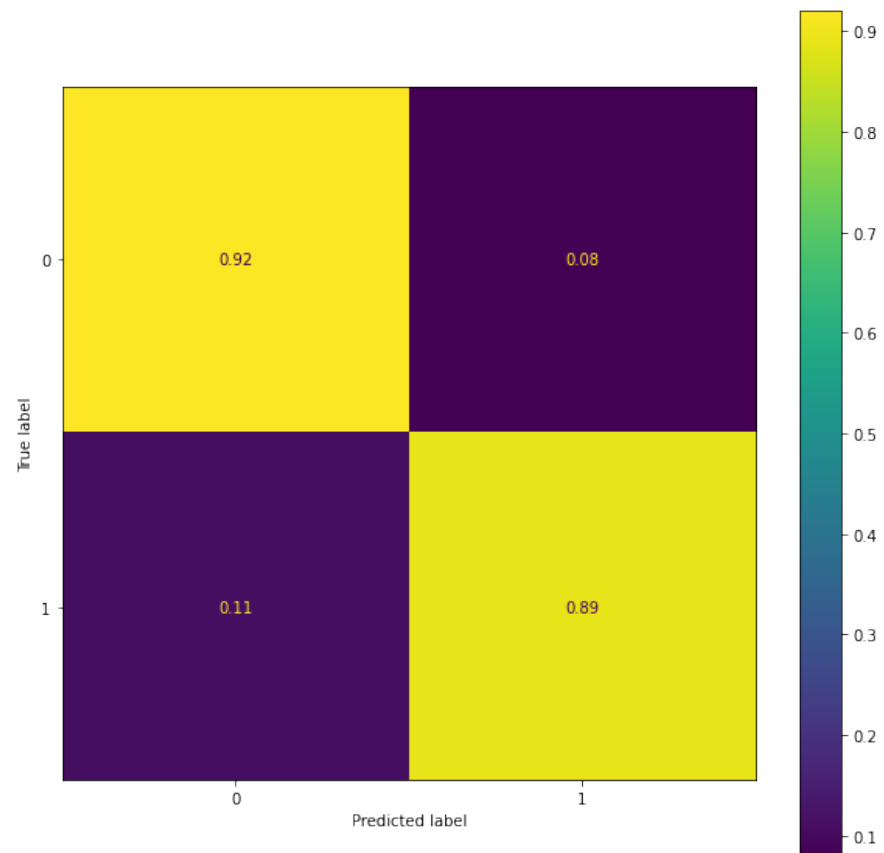


Figura 6.3: Matriz de confusión obtenida a partir del subset de validación para el modelo de SVM

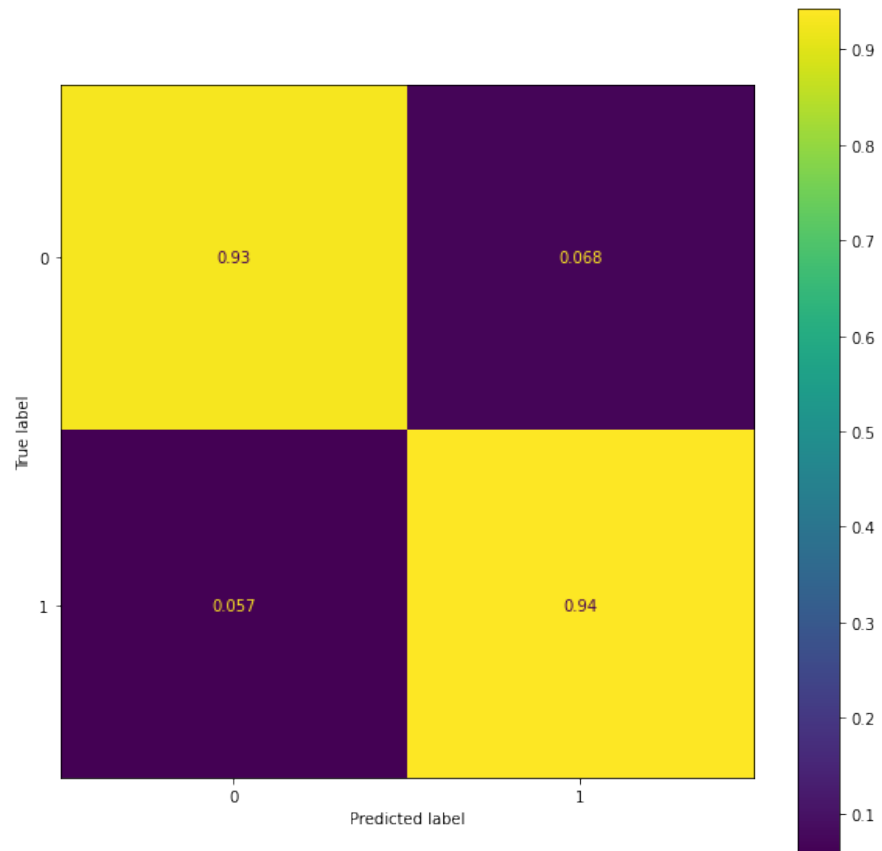


Figura 6.4: Matriz de confusión obtenida a partir del subset de test con tweets con menor cantidad de tokens para el modelo de perceptron multicapa

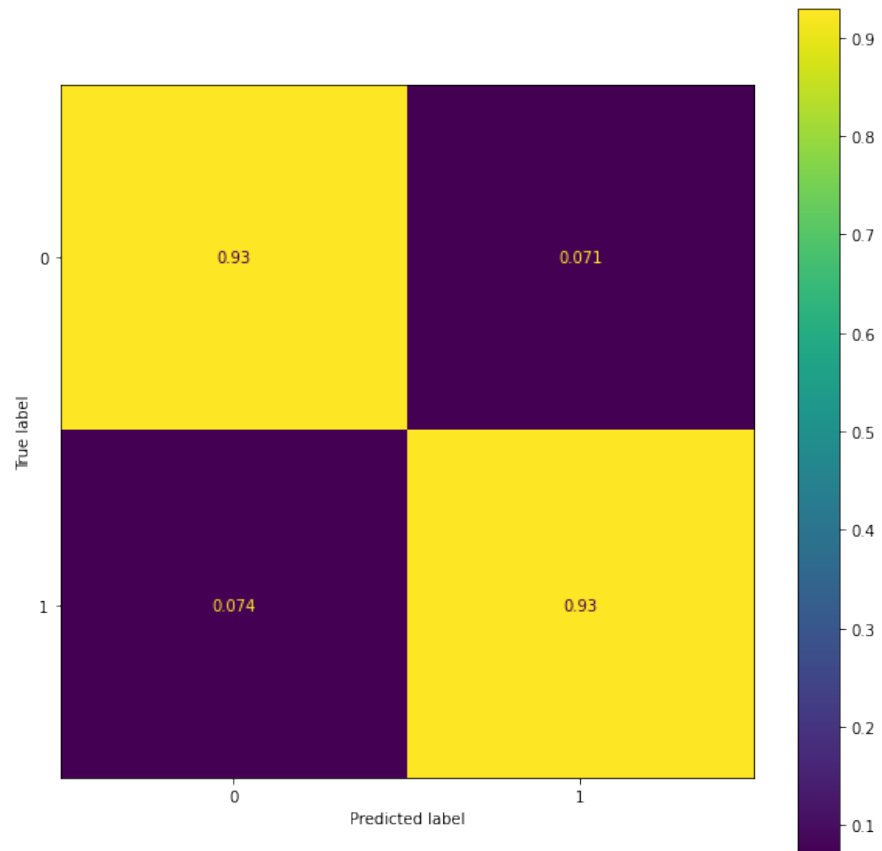


Figura 6.5: Matriz de confusión obtenida a partir del subset de test con tweets con mayor cantidad de tokens para el modelo de perceptron multicapa

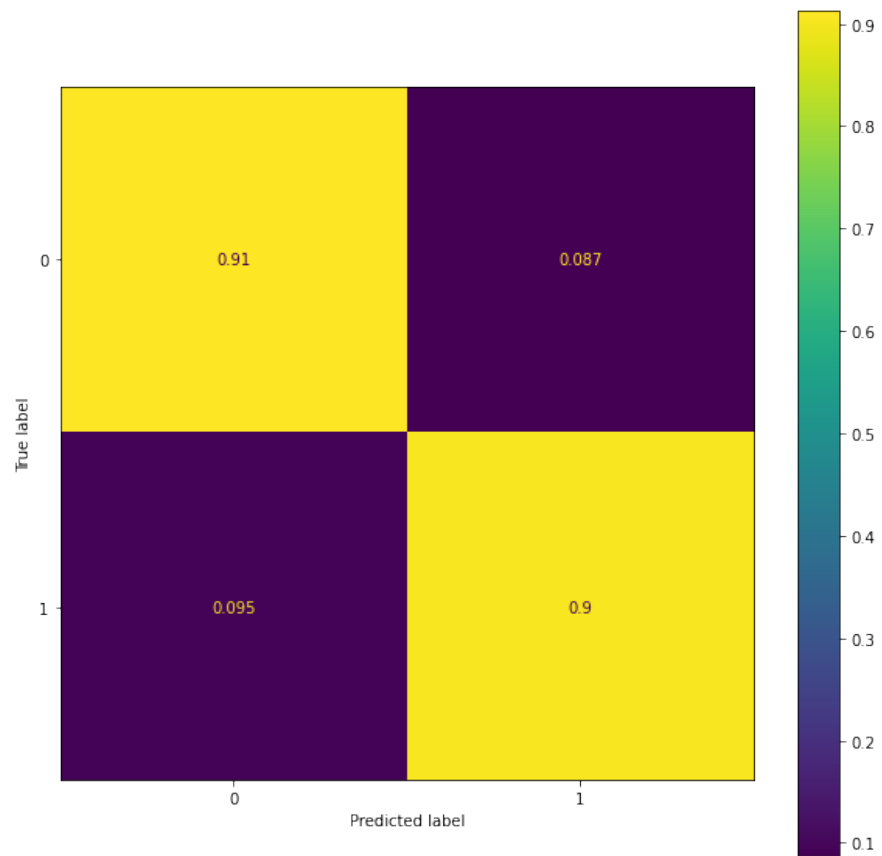


Figura 6.6: Matriz de confusión obtenida a partir del subset de test con tweets con menor cantidad de tokens para el modelo de regresión logística

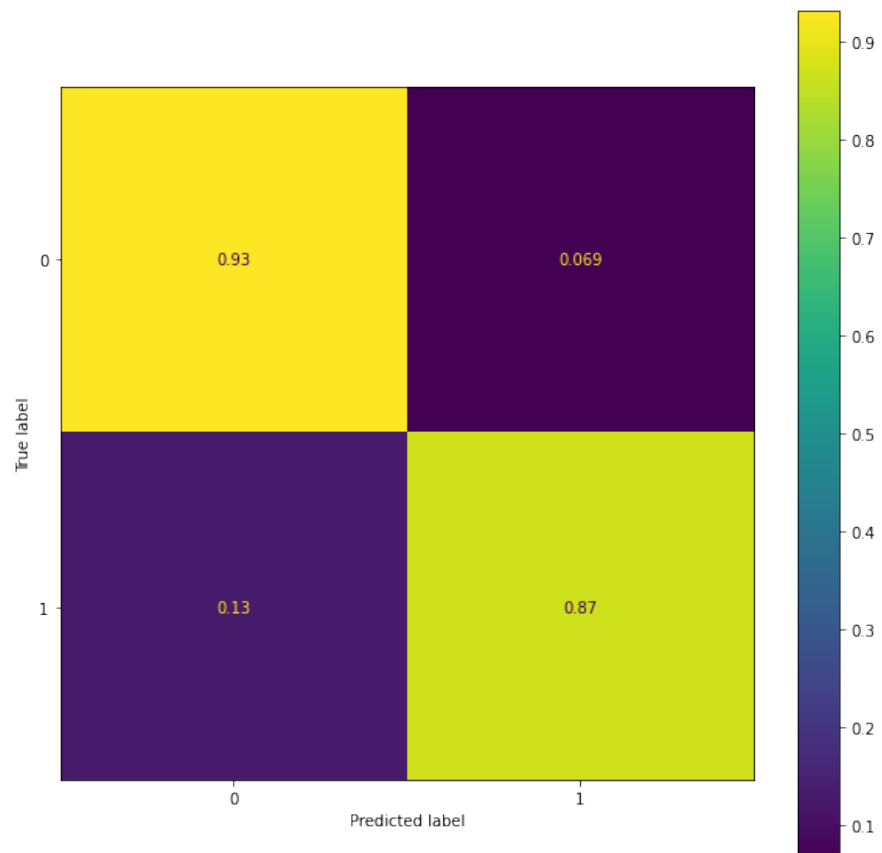


Figura 6.7: Matriz de confusión obtenida a partir del subset de test con tweets con mayor cantidad de tokens para el modelo de regresión logística

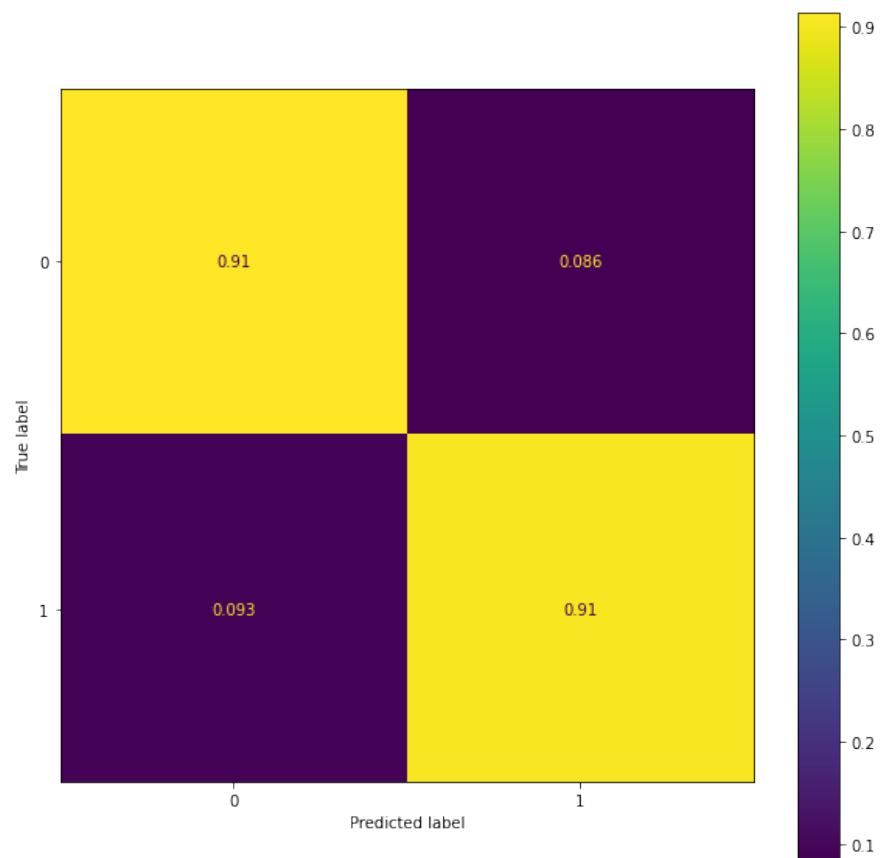


Figura 6.8: Matriz de confusión obtenida a partir del subset de test con tweets con menor cantidad de tokens para el modelo SVM

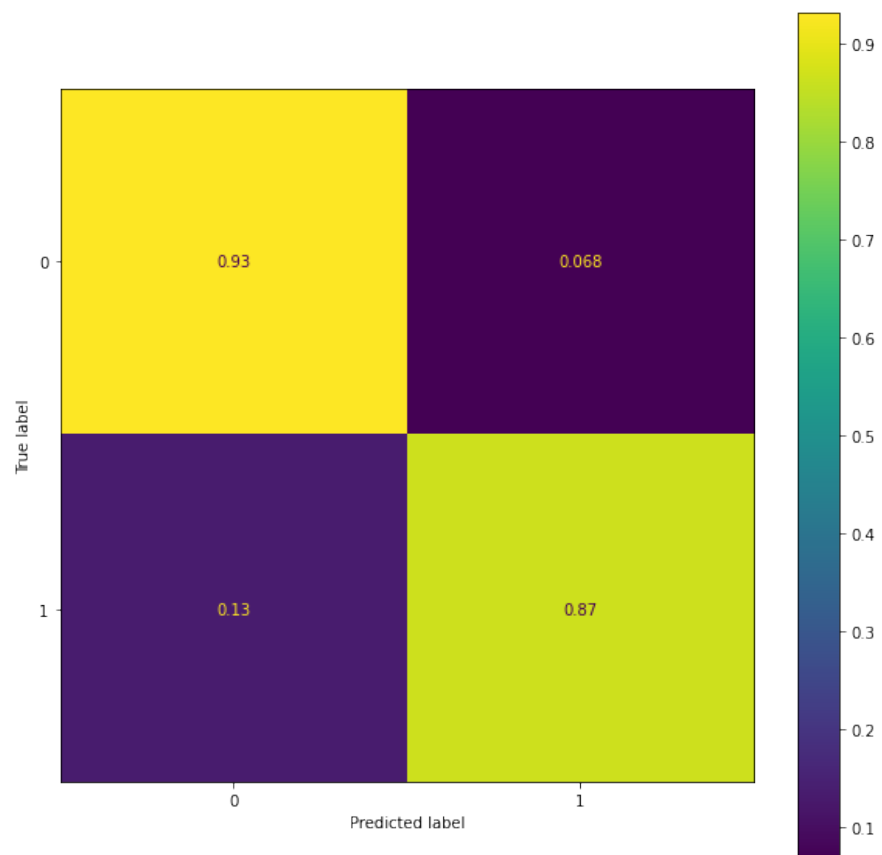


Figura 6.9: Matriz de confusión obtenida a partir del subset de test con tweets con mayor cantidad de tokens para el modelo SVM

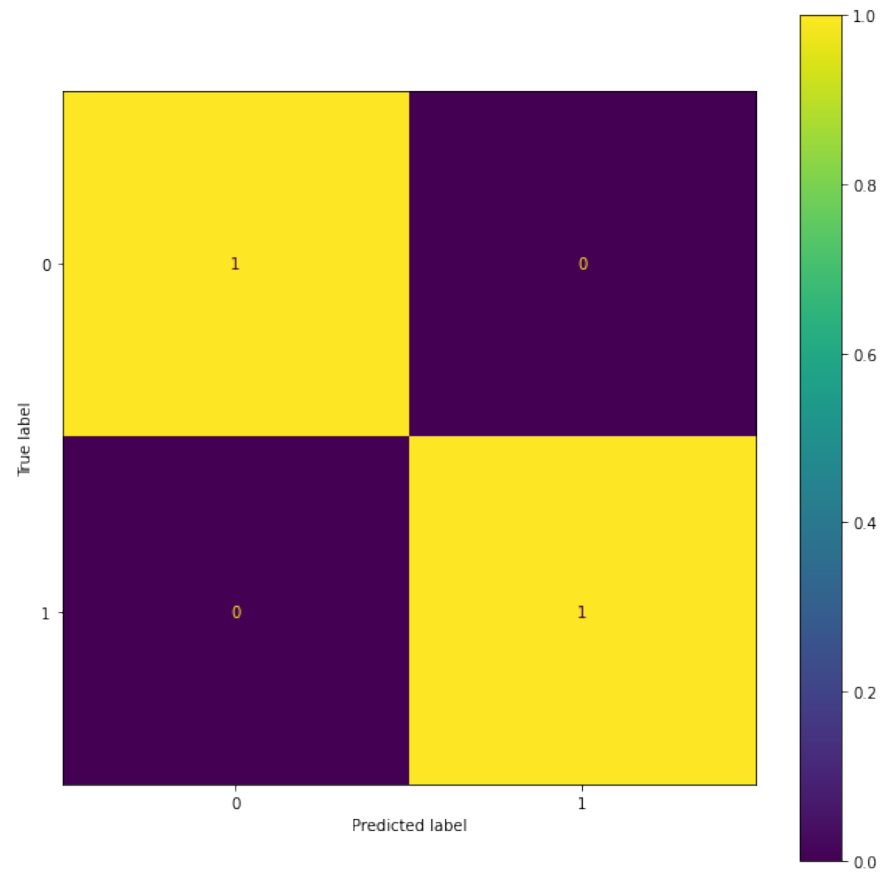


Figura 6.10: Matriz de confusión obtenida a partir del subset de test de usuarios con tweets con menor cantidad de tokens promedio para el modelo LSTM

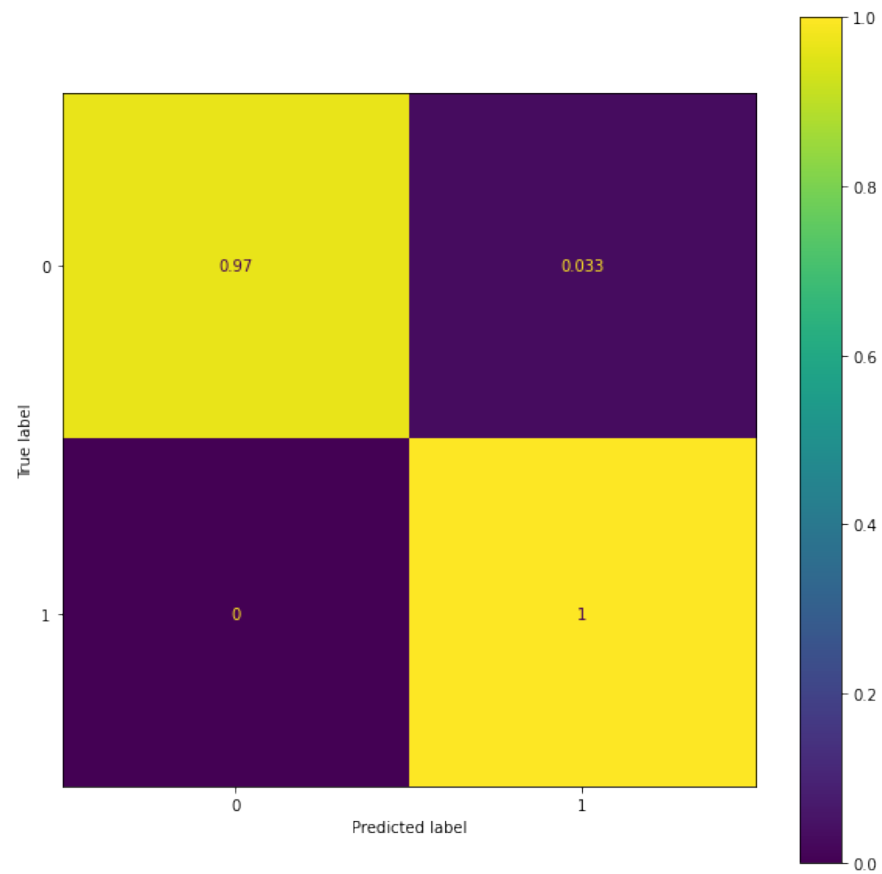


Figura 6.11: Matriz de confusión obtenida a partir del subset de test de usuarios con tweets con mayor cantidad de tokens promedio para el modelo LSTM

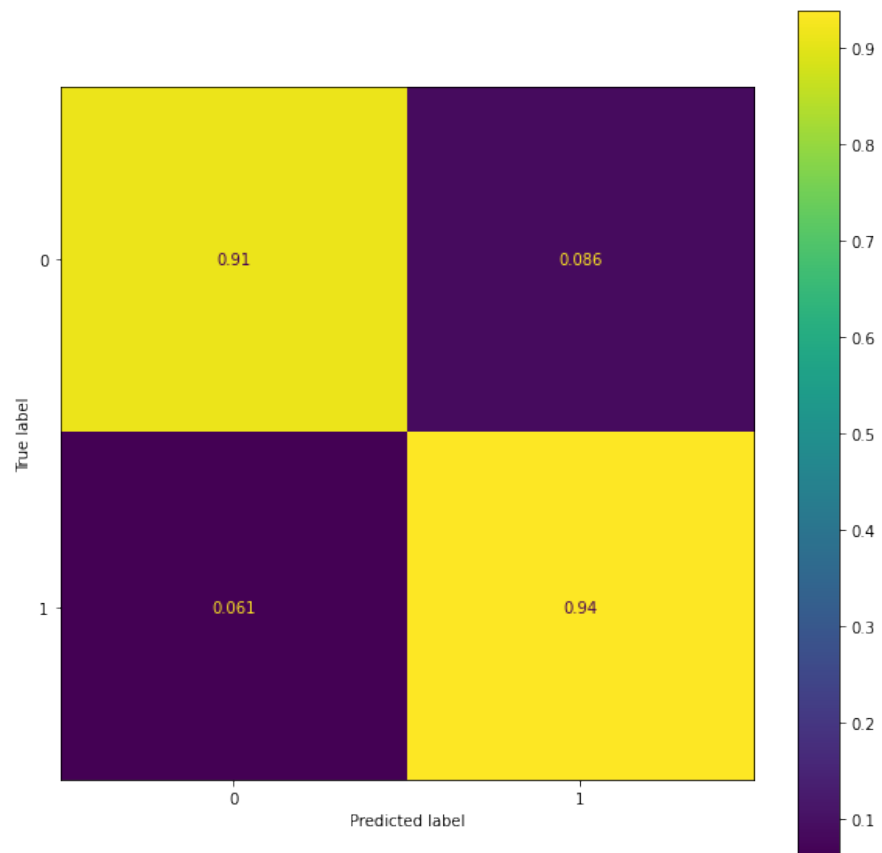


Figura 6.12: Matriz de confusión obtenida a partir del subset de test con tweets con menor sentiment score para el modelo de perceptron multicapa

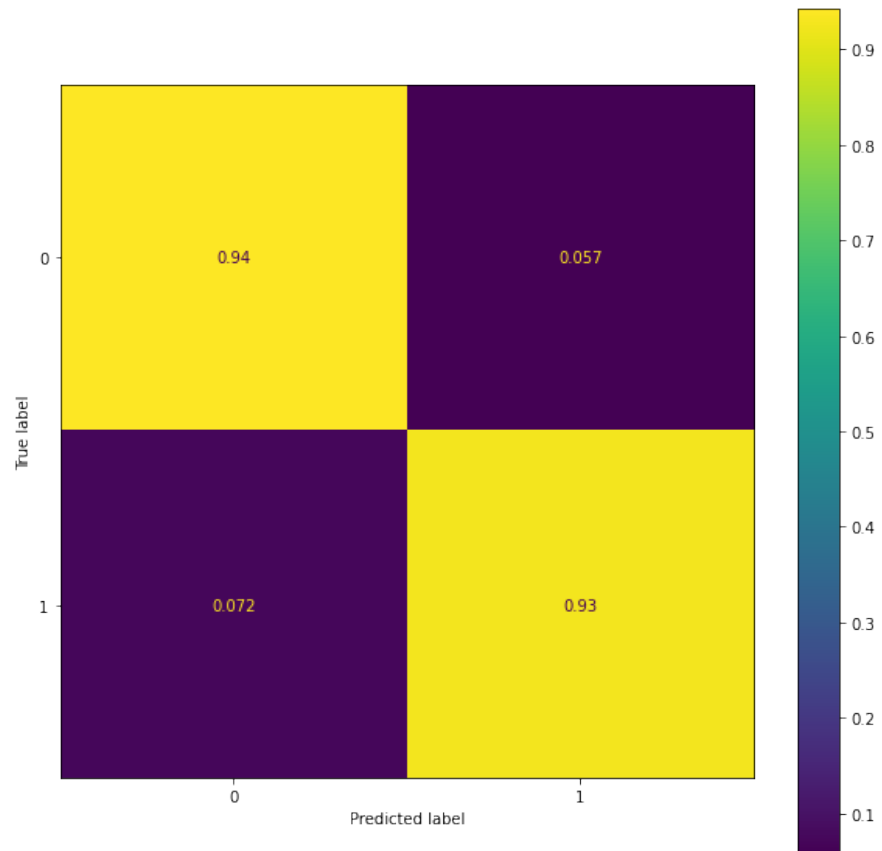


Figura 6.13: Matriz de confusión obtenida a partir del subset de test con tweets con mayor sentiment score para el modelo de perceptron multicapa

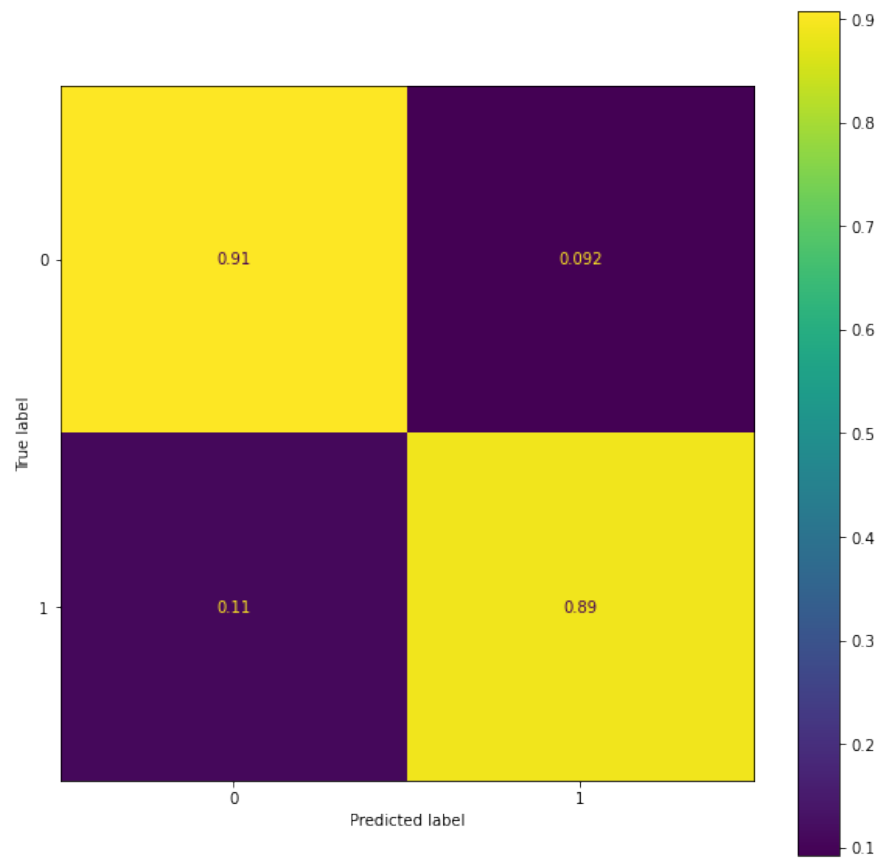


Figura 6.14: Matriz de confusión obtenida a partir del subset de test con tweets con menor sentiment score para el modelo de regresión logística

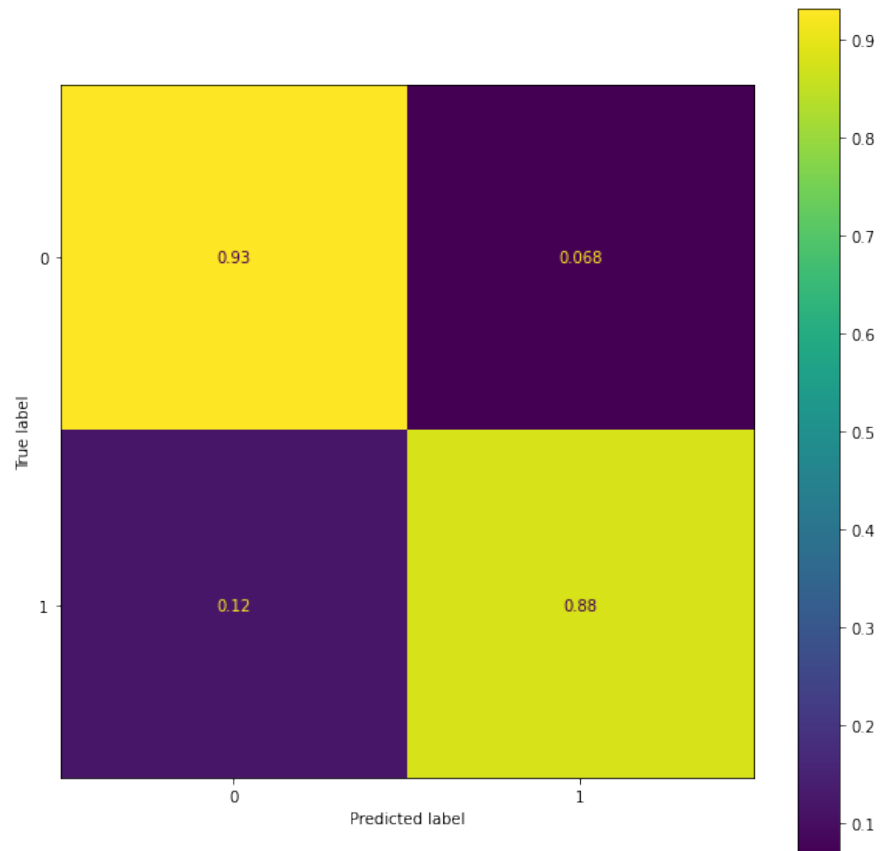


Figura 6.15: Matriz de confusión obtenida a partir del subset de test con tweets con mayor sentiment score para el modelo de regresión logística

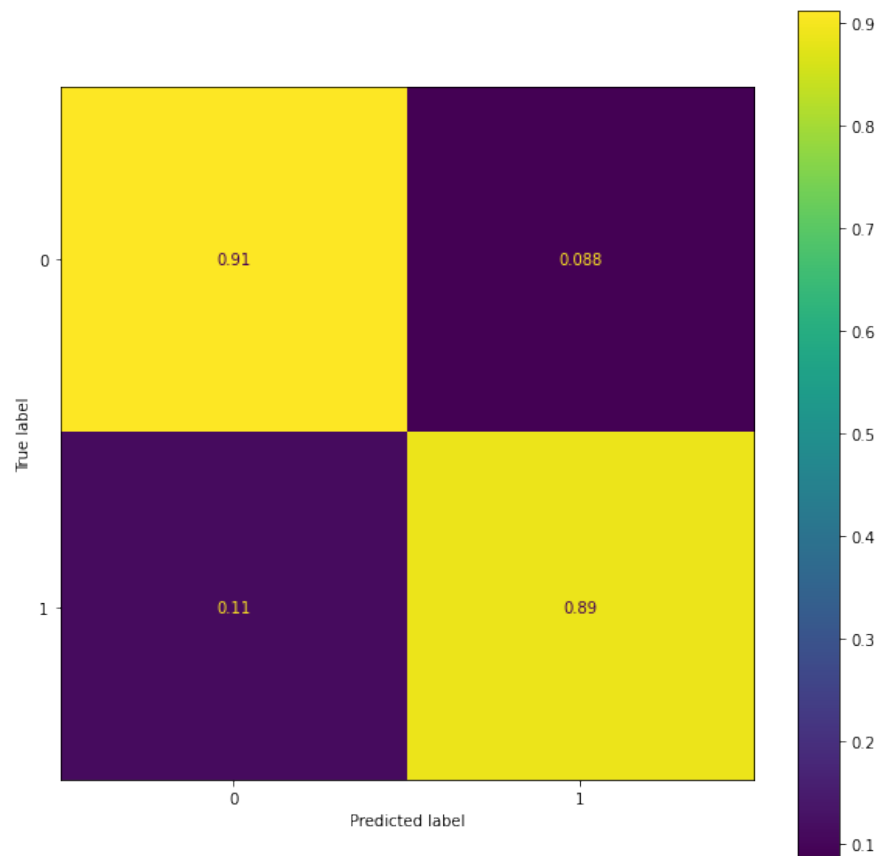


Figura 6.16: Matriz de confusión obtenida a partir del subset de test con tweets con menor sentiment score para el modelo SVM

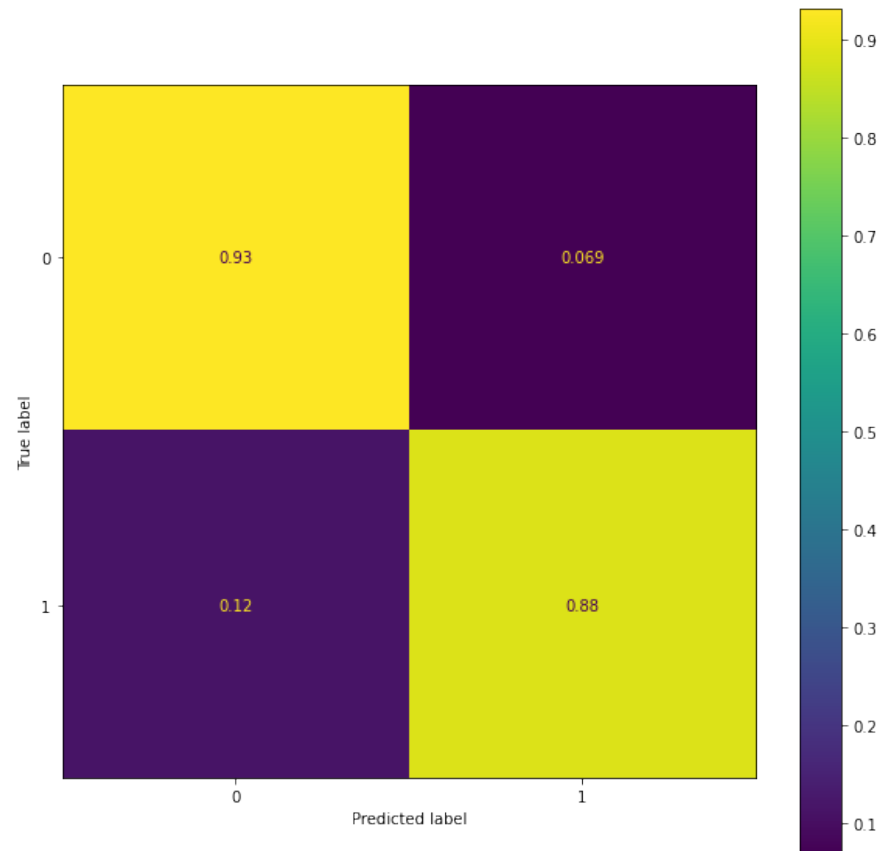


Figura 6.17: Matriz de confusión obtenida a partir del subset de test con tweets con mayor sentiment score para el modelo SVM

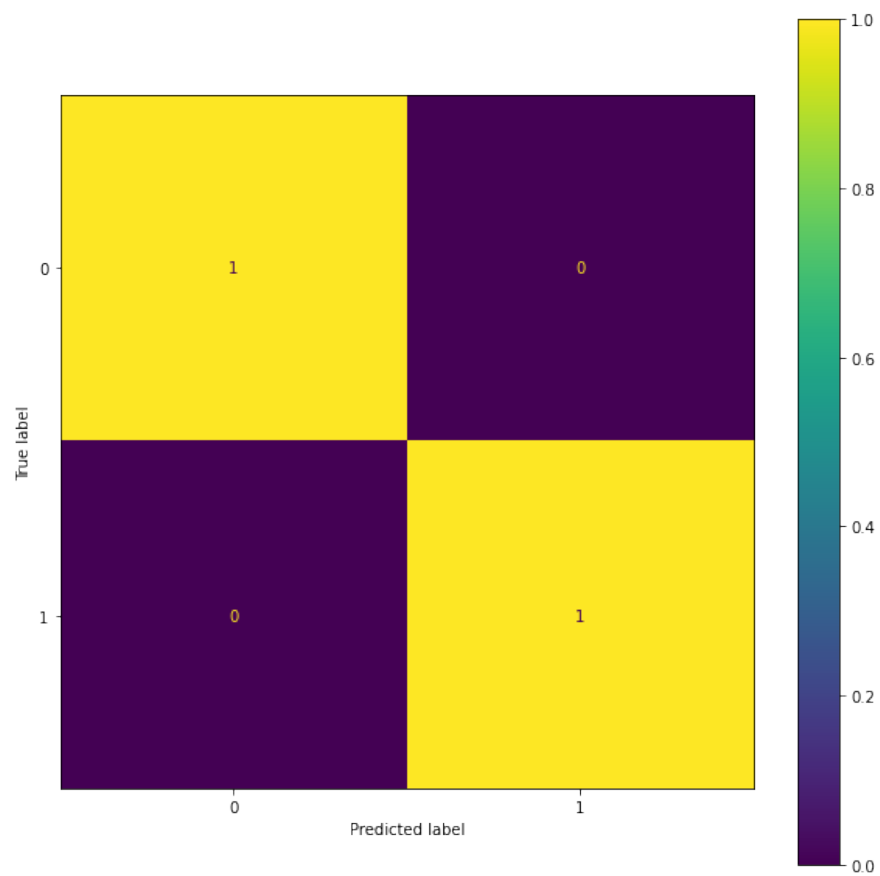


Figura 6.18: Matriz de confusión obtenida a partir del subset de test de usuarios con tweets con menor sentiment score promedio para el modelo LSTM

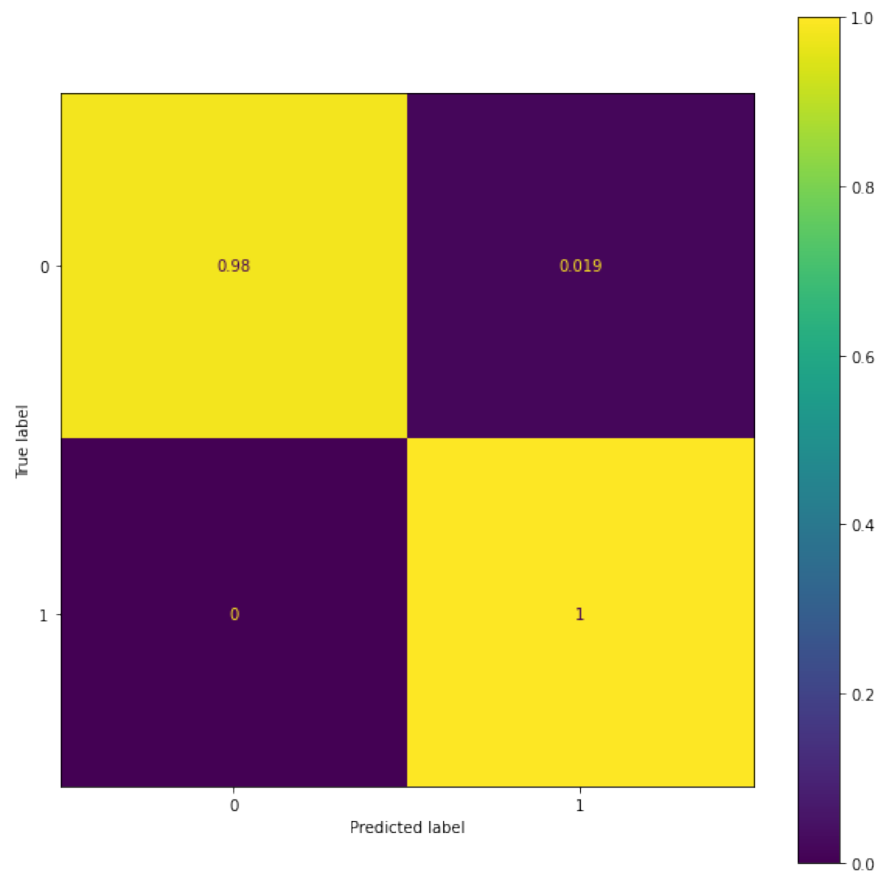


Figura 6.19: Matriz de confusión obtenida a partir del subset de test de usuarios con tweets con mayor sentiment score promedio para el modelo LSTM

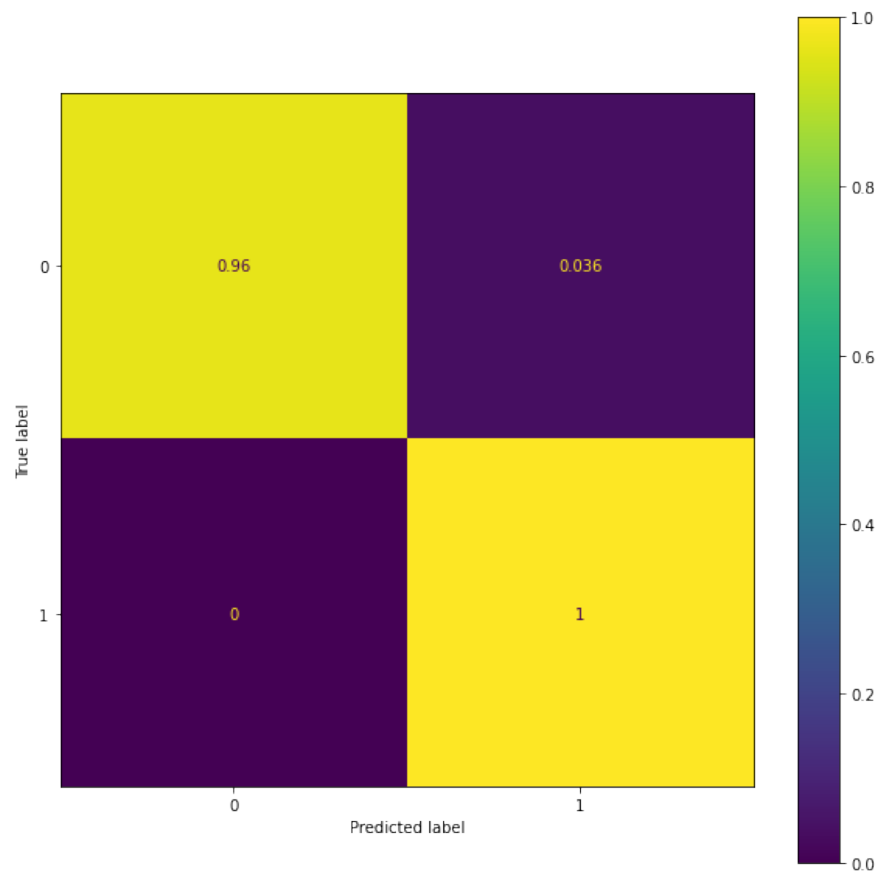


Figura 6.20: Matriz de confusión obtenida a partir del subset de test de usuarios con menor cantidad de tweets para el modelo LSTM

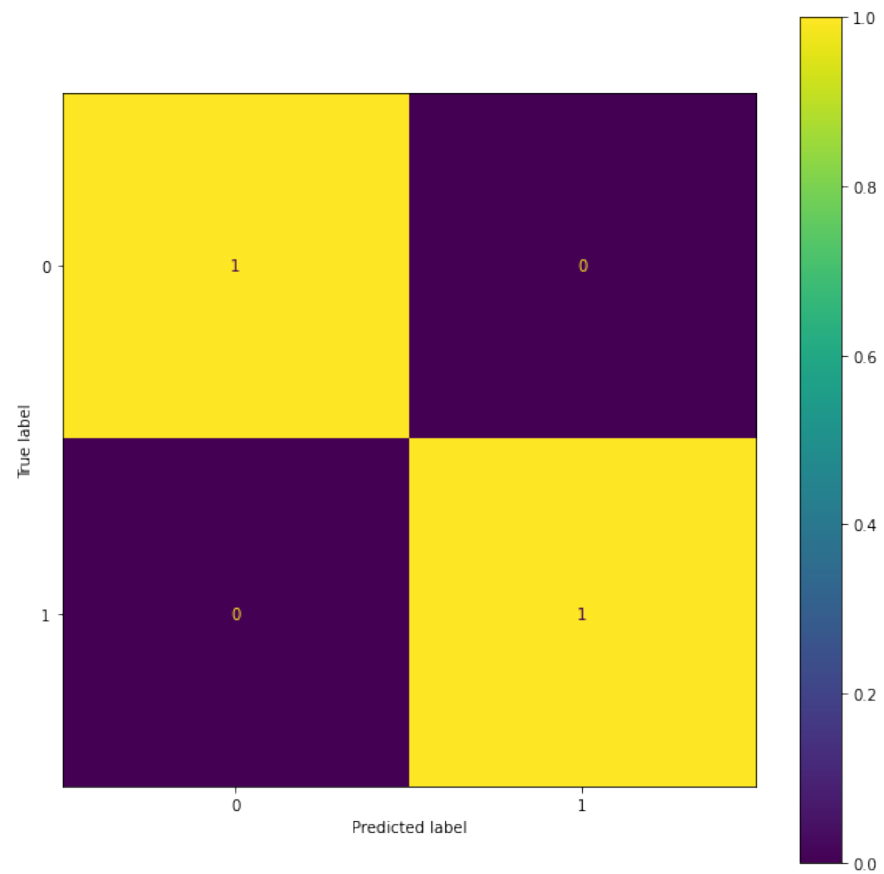


Figura 6.21: Matriz de confusión obtenida a partir del subset de test de usuarios con mayor cantidad de tweets para el modelo LSTM

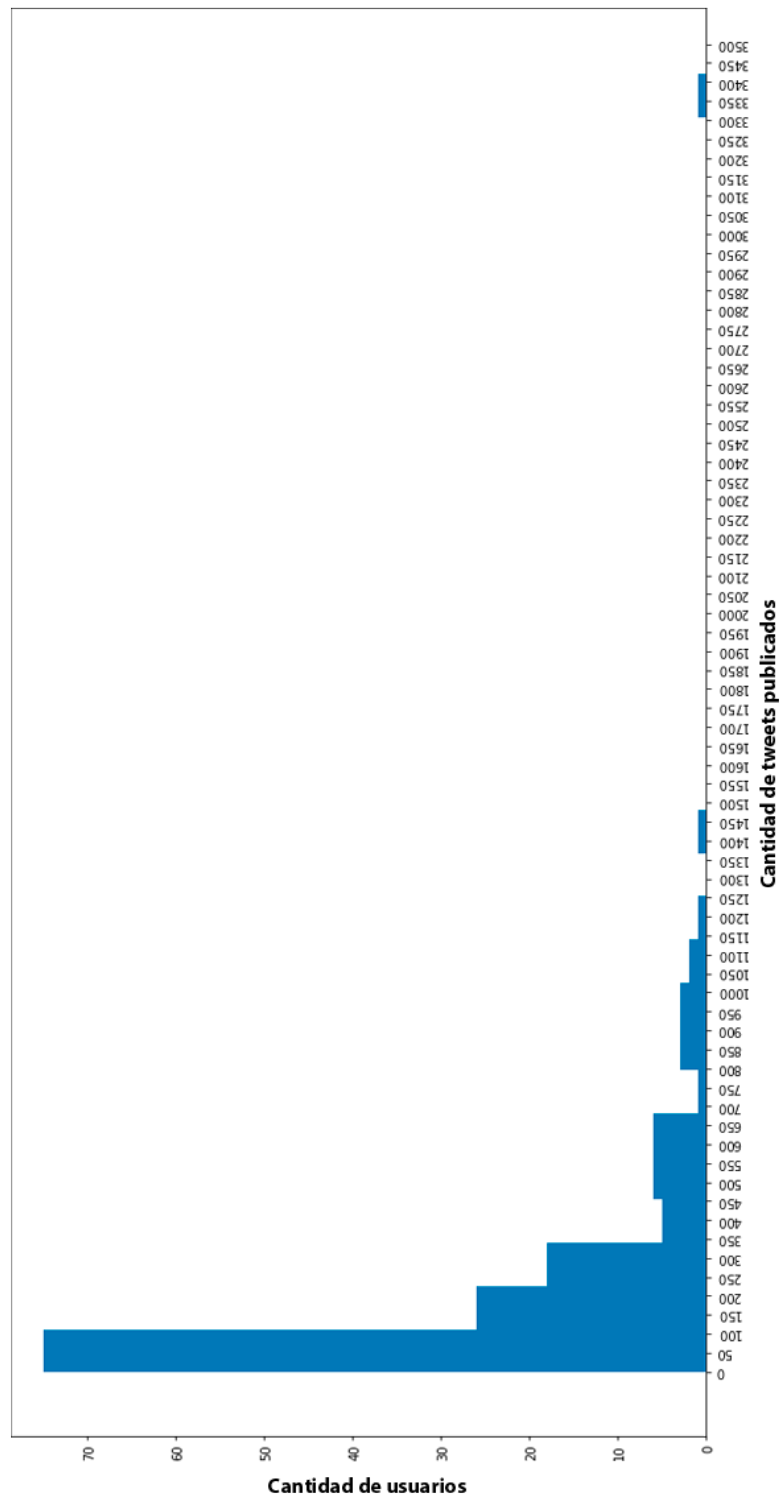


Figura 6.22: Histograma que muestra distribución en la cantidad de usuarios que publican cierta cantidad de tweets

Bibliografía

- [1] Anta, Antonio Fernández, Luis Núñez Chiroque, Philippe Morere y Agustín Santos: *Sentiment analysis and topic detection of Spanish tweets: A comparative study of of NLP techniques*. *Procesamiento del lenguaje natural*, 50:45–52, 2013.
- [2] Balani, Sairam y Munmun De Choudhury: *Detecting and characterizing mental health related self-disclosure in social media*. En *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, páginas 1373–1378, 2015.
- [3] Burdisso, Sergio G, Marcelo Errecalde y Manuel Montes-y Gómez: *PySS3: A Python package implementing a novel text classifier with visualization tools for Explainable AI*. arXiv preprint arXiv:1912.09322, 2019.
- [4] Cañete, José, Gabriel Chaperon, Rodrigo Fuentes, Jou Hui Ho, Hojin Kang y Jorge Pérez: *Spanish Pre-Trained BERT Model and Evaluation Data*. En *PML4DC at ICLR 2020*, 2020.
- [5] Coppersmith, Glen, Mark Dredze, Craig Harman y Kristy Hollingshead: *From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses*. En *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, páginas 1–10, 2015.
- [6] Devlin, Jacob, Ming Wei Chang, Kenton Lee y Kristina Toutanova: *Bert: Pre-training of deep bidirectional transformers for language understanding*. arXiv preprint arXiv:1810.04805, 2018.
- [7] Ghani, Norjihan Abdul, Suraya Hamid, Ibrahim Abaker Targio Hashem y Ejaz Ahmed: *Social media big data analytics: A survey*. *Computers in Human Behavior*, 101:417–428, 2019.
- [8] Harman, GACCT y Mark H Dredze: *Measuring post traumatic stress disorder in Twitter*. In *ICWSM*, 2014.

- [9] Leis, Angela, Miguel Angel Mayer, Francesco Ronzano, Marta Torrens, Claudio Castillo, Laura I Furlong y Ferran Sanz: *Clinical-Based and Expert Selection of Terms Related to Depression for Twitter Streaming and Language Analysis*. Studies in health technology and informatics, 270:921–925, 2020.
- [10] Leis Machin, Angela, Francesco Ronzano, Miguel Ángel Mayer, Laura I Furlong y Ferran Sanz: *Detecting signs of depression in tweets in spanish: behavioral and linguistic analysis*. J Med Internet Res. 2019; 21 (6): e14199, 2019.
- [11] Tausczik, Yla R y James W Pennebaker: *The psychological meaning of words: LIWC and computerized text analysis methods*. Journal of language and social psychology, 29(1):24–54, 2010.