



UNIVERSIDAD AUSTRAL

TESIS DE MAESTRÍA

**Clasificación de los dígitos escritos
en los telegramas de las elecciones
legislativas en Santa Fe mediante
técnicas de adaptación de dominio.**

Autor:
Franco LIANZA

Supervisor:
Dr. Leandro BUGNON

4 de diciembre de 2022

UNIVERSIDAD AUSTRAL

Resumen

Facultad de Ingeniería

Magister en Explotación de Datos y Gestión del Conocimiento

Clasificación de los dígitos escritos en los telegramas de las elecciones legislativas en Santa Fe mediante técnicas de adaptación de dominio.

by Franco LIANZA

The Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too...

Reconocimientos

The acknowledgments and the people to thank go here, don't forget to include your project advisor...

Índice general

Resumen	III
Reconocimientos	v
1. Introducción	1
1.1. Contexto	1
1.2. Motivación	2
1.3. Objetivos	3
2. Marco teórico	5
2.1. Reconocimiento de dígitos	5
2.2. Redes Neuronales	5
2.2.1. Redes densas	6
2.2.2. Convolucionales	6
2.2.3. Arquitecturas	6
LeNet	7
ResNet	7
2.3. Adaptación de Dominio	7
2.3.1. Domain Adversarial Neural Network	8
2.3.2. Adversarial Discriminative Domain Adaptation	9
2.3.3. AFN	10
2.3.4. MDD	11
3. Metodología	13
3.1. Descripción de los datos	13
3.2. Extracción de dígitos de los telegramas	13
3.2.1. Enderezado	13
3.2.2. Extracción de la grilla de votos	14
3.2.3. Detección de líneas de la grilla de votos	14
3.2.4. Segmentación de dígitos	16
3.3. Análisis del dataset	16
3.4. Diseño experimental	18
3.4.1. Metodología de entrenamiento	18
3.4.2. Selección y optimización del modelo	19
3.5. Métricas de evaluación	19
3.5.1. Accuracy	20
3.5.2. F_1	20
3.5.3. Intersección sobre unión	21
3.5.4. Distancia \mathcal{A}	21
3.5.5. Discrepancia Máxima Promedio	22

4. Análisis de resultados	23
4.1. Análisis de métricas	23
4.2. Comparación de espacios latentes	24
4.3. Análisis de errores	25
5. Conclusiones	29
5.1. Conclusiones	29
5.2. Trabajos Futuros	29
A. Anexo: Telegramas	31
A.1. Ejemplo de telegrama	32
A.2. Ejemplo de telegrama mal cargado	33
A.3. Ejemplo de telegrama con números sin espacio entre ellos . . .	34
A.4. Ejemplo de telegrama con caracteres distintos a números . . .	35
A.5. Ejemplo de telegrama ceros a la izquierda de la cantidad de votos	36
Bibliografía	37

Índice de figuras

1.1. Proceso eleccionario. TODO: cambiar para que se pueda leer o eliminar directament?	2
2.1. Esquema de las <i>DANN</i> . Los supra indices s y t indican si los datos son del origen o destino respectivamente. Los círculos con líneas rayadas son salidas de los modelos mientras que los círculos de líneas sólidas son datos conocidos.	8
2.2. Esquema de las <i>ADDA</i> . Los supra indices s y t indican si los datos son del origen o destino respectivamente. Los círculos con líneas rayadas son salidas de los modelos mientras que los círculos de líneas sólidas son datos conocidos. La primera fase consta de un pre-entrenamiento con una red generadora \mathcal{G}_s con los datos origen y la segunda fase de adaptación consta de otra red generadora \mathcal{G}_t que debe aprender a generar el mismo espacio latente que \mathcal{G}_s con los datos de destino para confundir a \mathcal{D} .	10
3.1. Enderezamiento de un telegrama utilizando OpenCV.	14
3.2. Grilla de votos extraída buscando el contorno de mayor área.	14
3.3. Proyecciones de los ejes de la grilla. En rojo se marca el umbral de corte.	15
3.4. Grilla detectada (en verde) utilizando el umbral por sobre las proyecciones y su posterior agrupamiento con clustering jerárquico.	15
3.5. Primer registro extraído de la grilla de votos.	16
3.6. Dígitos detectados en el primer bloque de votos.	16
3.7. Histogramas de proporciones de píxeles blancos por voto.	17
3.8. Dataset TDS de dígitos.	18
3.9. Ejemplos de distribuciones de dominios.	22
4.1. Distribución de aciertos de cantidad de votos por cada par técnica AD y modelo.	24
4.2. Histogramas de la métrica <i>IoU</i> promedio por telegrama por cada par técnica AD y modelo.	26

Índice de cuadros

3.1. Ejemplo de registros del dataset. Cada fila representa un voto.	16
3.2. Estadísticos descriptivos del mínimo y máximo tamaño de las imágenes de los votos.	17
3.3. Rango de hiperparámetros optimizados. *LR refiere a "learning rate". **T-O refiere a "trade-off".	19
4.1. Métricas de los experimentos realizados. Entre paréntesis se encuentra la posición que ocupa dentro del top 3 de la columna.	23
4.2. IoU promedio y cantidad promedio de aciertos al aplicar los modelos a cada telegrama.	24

Capítulo 1

Introducción

La intro la podes pensar en 3 subsecciones: descripción de elecciones y contabilización de las actas hoy, la oportunidad que se presenta con la digitalización (y los desafíos que aparecen) y luego los objetivos. Tiene que quedar claro que problema ves y qué propones hacer a grandes rasgos. También es importante que estos objetivos se vean reflejados en los resultados y conclusiones (es decir, prometer algo que después cumplis al final del doc)

1.1. Contexto

En Argentina se celebran elecciones cada 2 años a excepción de las presidenciales que se realizan cada 4 años. Existen, principalmente, tres tipos de elecciones:

- Elecciones nacionales, para elegir a las autoridades federales del país: el Poder Ejecutivo, constituido por el Presidente y el vicepresidente y el Congreso Nacional, formado por Senadores y Diputados.
- Elecciones provinciales y de la Ciudad de Buenos Aires o locales, para elegir a las autoridades de cada provincia: los poderes ejecutivos de las provincias y sus legislaturas.
- Elecciones municipales, regidas por las leyes y procedimientos de cada provincia.

Si bien emitir el sufragio es diferente en cada una de ellas, generalmente consta de ingresar a un cuarto oscuro, elegir el candidato que se desea y depositar el voto en una urna. Al finalizar la jornada, las autoridades de mesas recuentan los votos y llenan una planilla a mano alzada donde se resume la cantidad de votos obtenidos por cada candidato o partido político. Dicha planilla es escaneada y enviada a través de un telegrama del correo argentino al centro de cómputo para su procesamiento. Una vez allí, se contabilizan en un sistema informático a partir de un grupo de personas. Este proceso cuenta con una etapa de digitalización y otra de validación (TODO: esto lo vi en el diagrama de las elecciones de cordoba. no encontre nada oficial al respecto.).



FIGURA 1.1: Proceso eleccionario. TODO: cambiar para que se pueda leer o eliminar directamente?

Para que el proceso sea lo mas rápido y eficaz posible, se requiere a una gran cantidad personas destinadas al centro de cómputo. Tal solución hace que el proceso sea altamente ineficiente en cuanto a tiempos y costos se refiere. En las elecciones legislativas del 2021 se gastaron unos \$17.000 millones de pesos de los cuales \$4.000 millones de pesos fueron destinados a sueldos para el personal¹.

1.2. Motivación

Digitalizar los telegramas de manera automática supondrá un ahorro considerable en el presupuesto de las elecciones, agilizará la obtención de los resultados y aportará transparencia al proceso en general. Es posible entrenar un modelo de clasificación de dígitos a un costo extremadamente menor al actual y utilizarlo al momento de la contabilización de los votos.

Contar con una digitalización automática permitirá bajar los costos debido a que se necesitará un grupo menor de personas en el centro de cómputo. Además, el trabajo a realizar será mas simple ya que sólo constará del proceso de validación.

La clasificación de números es un problema que, si bien parece resuelto con LeCun et al., 1998 y la creación del dataset MNIST, no debe ser tomada a la ligera. No existe una única forma de escribir y año a año cambian las personas que son los jefes de mesa encargados de completar los telegramas. Las características de los números escritos a mano difiere entre cada elección. Cuando la distribución de los datos de entrenamiento difiere a la de los datos de aplicación, se está ante un *corrimiento de dominio* (*domain shift* o *data drift* en inglés). Esto implica que un modelo que se entrene en algún dataset estático como el MNIST que fue creado en el 1998, hará malas clasificaciones los dígitos de las elecciones.

Cuando los dominios de entrenamiento (origen) y aplicación (destino) son distintos, se debe a que hay un *sesgo* en los datos. Las técnicas de entrenamiento clásicas suponen que, si bien existe el sesgo existe, éste será el mismo entre origen y destino. Cuando el supuesto no se cumple, se debe recurrir a técnicas de entrenamiento más complejas donde se intenta que el

¹Fuente: [El cronista](#)

modelo aprenda a adaptar un dominio a otro. Hoy en día no hay trabajos relacionados a la digitalización de telegramas en Argentina que compare cuál es la mejor técnica de adaptación de dominio.

1.3. Objetivos

La presente tesis enfocará en el desarrollo de un modelo que permita digitalizar los telegramas de las elecciones en Argentina utilizando las legislativas de la provincia de Santa Fe del año 2021 como comparativa. Los telegramas son públicos y se encuentran subidos en la [página oficial del estado argentino](#). En el anexo A.1 se adjunta un ejemplo de uno de ellos.

Para poder llevarlo a cabo, se procede a:

- Extraer los números escritos a mano en los telegramas.
- Evaluar técnicas de adaptación de dominio para entrenar modelos destinados a la clasificación de números de los telegramas.
- Analizar los espacios latentes de los modelos con el fin de obtener *insights* respecto a la capacidad de generalización de los modelos.

Capítulo 2

Marco teórico

En este capítulo se presenta una visión general de la bibliografía relacionada al objetivo del presente trabajo. En la primera sección se detalla una revisión de trabajos relacionados. En ...

2.1. Reconocimiento de dígitos

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

2.2. Redes Neuronales

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

2.2.1. Redes densas

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

2.2.2. Convolucionales

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

2.2.3. Arquitecturas

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

LeNet

Lore ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

ResNet

Lore ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

2.3. Adaptación de Dominio

El *pre-entrenamiento* y el *fine-tuning* han mejorado considerablemente el estado del arte para diversos problemas y aplicaciones del machine learning, incluso las redes profundas pre-entrenadas pueden adaptarse fácilmente a tareas donde se cuenta con una pequeña cantidad de datos etiquetados. Sin embargo, en muchos escenarios prácticos, no hay datos de entrenamiento etiquetados y, por lo tanto, existe la necesidad de transferir el aprendizaje de una red profunda desde un dominio de origen en el que se dispone de datos de entrenamiento etiquetados a un dominio de destino en el que solo existen datos sin etiquetar (Glorot, Bordes y Bengio, 2011). En esta situación, los modelos profundos siguen sufriendo degradaciones de rendimiento debido al cambio de distribución (Quinonero-Candela et al., 2008). Por lo tanto, se propone la *adaptación de dominio* para reducir el cambio de distribución entre los dominios de entrenamiento y de prueba (Jiang et al., 2022).

Se han propuesto muchos métodos para la adaptación de dominios, ya sea ponderando o seleccionando muestras del dominio de origen (Sugiyama

et al., 2007) o buscando una transformación de la distribución de origen a la distribución de destino (Gong, Grauman y Sha, 2013). Esta tesis analizará algunos de los métodos más actuales del ultimo caso.

2.3.1. Domain Adversarial Neural Network

Un gran hito al momento de modelar distribuciones es la Red Generativa Adversaria *GAN* (Goodfellow et al., 2020). Inspirada en estas arquitecturas *GAN*, la Red Neural Adversaria de Dominio *DANN* (Ganin et al., 2016) dos redes en la adaptación del dominio (figura 2.1). La primera es la discriminadora de dominios \mathcal{D} entrenada para distinguir las características de origen de las de destino y la segunda es una generadora de características \mathcal{G} que se entrena para confundir a \mathcal{D} y simultáneamente ayudar a la red clasificadora \mathcal{C} .

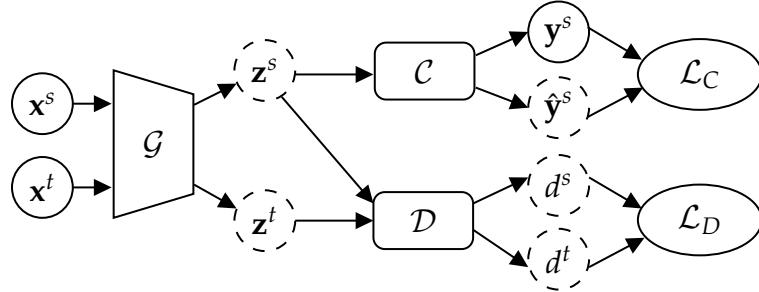


FIGURA 2.1: Esquema de las *DANN*. Los supra indices s y t indican si los datos son del origen o destino respectivamente. Los círculos con líneas rayadas son salidas de los modelos mientras que los círculos de líneas sólidas son datos conocidos.

Las redes *DANN* poseen una función de pérdida \mathcal{L}_G que depende de \mathcal{L}_C y \mathcal{L}_D . \mathcal{L}_C mide el error que posee la red en la clasificación de los datos de origen y viene dada por el *cross-entropy* \mathcal{L}_{CE} aplicado a la salida de la red \mathcal{C} . El objetivo de \mathcal{D} será la descripta en la ecuación 2.1:

$$\begin{aligned} \min_{\mathcal{C}} \mathcal{L}_C(\mathbf{x}^s, \mathbf{y}^s) &= \mathcal{L}_{CE}(C(\mathcal{G}(\mathbf{x}^s)), \mathbf{y}^s) \\ &= \mathcal{L}_{CE}(C(\mathbf{z}^s), \mathbf{y}^s) \\ &= \mathcal{L}_{CE}(\hat{\mathbf{y}}^s, \mathbf{y}^s) \end{aligned} \quad (2.1)$$

La función de pérdida de \mathcal{D} viene dada por la ecuación 2.2, donde $\mathbb{E}_{\mathbf{x}^s \sim \hat{\mathcal{S}}}$ y $\mathbb{E}_{\mathbf{x}^t \sim \hat{\mathcal{T}}}$ representan la proporción esperada de datos de origen $\hat{\mathcal{S}}$ y destino $\hat{\mathcal{T}}$ respectivamente.

$$\begin{aligned} \max_{\mathcal{G}, \mathcal{D}} \mathcal{L}_D(\mathbf{x}^s, \mathbf{x}^t) &= \mathbb{E}_{\mathbf{x}^s \sim \hat{\mathcal{S}}} \log[\mathcal{D}(\mathcal{G}(\mathbf{x}^s))] + \mathbb{E}_{\mathbf{x}^t \sim \hat{\mathcal{T}}} \log[1 - \mathcal{D}(\mathcal{G}(\mathbf{x}^t))] \\ &= \mathbb{E}_{\mathbf{x}^s \sim \hat{\mathcal{S}}} \log[\mathcal{D}(\mathbf{z}^s)] + \mathbb{E}_{\mathbf{x}^t \sim \hat{\mathcal{T}}} \log[1 - \mathcal{D}(\mathbf{z}^t)] \\ &= \mathbb{E}_{\mathbf{x}^s \sim \hat{\mathcal{S}}} \log[d^s] + \mathbb{E}_{\mathbf{x}^t \sim \hat{\mathcal{T}}} \log[1 - d^t] \end{aligned} \quad (2.2)$$

Por lo tanto, el \mathcal{L}_G viene dado por la ecuación 2.3. Donde λ es un hiperparámetro a optimizar que regula el trade-off entre la clasificación y la discriminación adversaria. La minimización de \mathcal{L}_C dará lugar a representaciones discriminables, mientras que la disminución de \mathcal{L}_D dará lugar a representaciones transferibles.

$$\min_{\mathcal{G}} \mathcal{L}_G(\mathbf{x}^s, \mathbf{x}^t) = \mathcal{L}_C(\mathbf{x}^s, \mathbf{y}^s) + \lambda \mathcal{L}_D(\mathbf{x}^s, \mathbf{x}^t) \quad (2.3)$$

2.3.2. Adversarial Discriminative Domain Adaptation

Las redes Adversarial Discriminative Domain Adaptation *ADDA* surgen a partir de las *DANN* debido a que su estrategia de optimización podría no funcionar bien en la práctica debido a la desvanecimiento del gradiente, que también es un problema importante en el entrenamiento de GANs. Por ejemplo, cuando el espacio latente de destino generado \mathbf{z}^t es extremadamente distingible del de origen tal que $\mathcal{D}(\mathbf{z}^t) = 0$, el gradiente es pequeño y vice versa. Esto provoca que la optimización de \mathcal{G} sea extremadamente difícil.

Las redes *ADDA* separan la optimización de \mathcal{G} y de \mathcal{D} en dos objetivos separados (figura 2.2) siendo el primero el pre-entrenamiento de \mathcal{G}_s y \mathcal{C} en los datos de origen y el otro siendo la adaptación adversaria de una \mathcal{G}_t utilizando a \mathcal{G}_s pre-entrenada y \mathcal{D} .

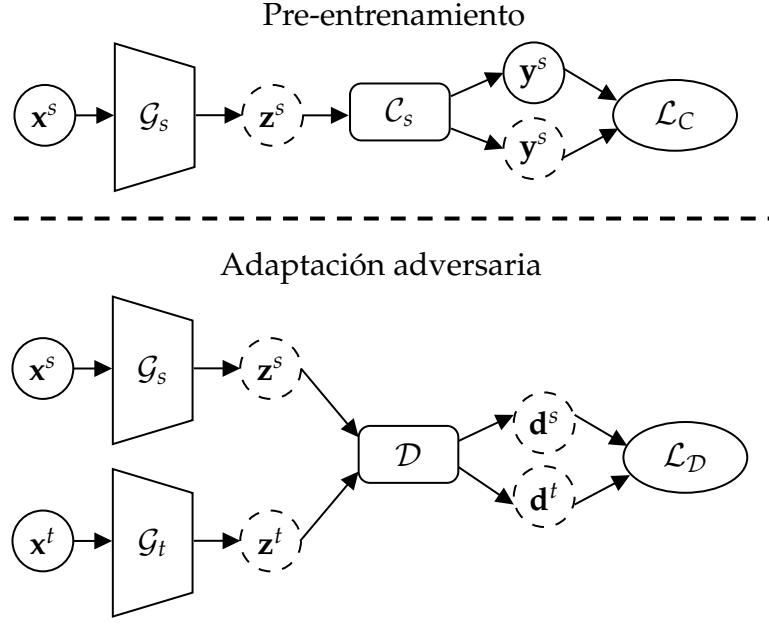


FIGURA 2.2: Esquema de las ADDA. Los supra indices s y t indican si los datos son del origen o destino respectivamente. Los círculos con líneas rayadas son salidas de los modelos mientras que los círculos de líneas sólidas son datos conocidos. La primera fase consta de un pre-entrenamiento con una red generadora \mathcal{G}_s con los datos origen y la segunda fase de adaptación consta de otra red generadora \mathcal{G}_t que debe aprender a generar el mismo espacio latente que \mathcal{G}_s con los datos de destino para confundir a \mathcal{D} .

No obstante del cambio, los objetivos son similares a ADDA excepción de \mathcal{G}_t . La ecuación 2.4 contiene \mathcal{L}_C que es análoga a 2.1. La ecuación 2.5 contiene \mathcal{L}_D que es análoga a 2.2. Finalmente, la ecuación 2.6 contiene $\mathcal{L}_{\mathcal{G}_s}$ que asigna gradientes pequeños a registros de destino que sean similares a los de origen y gradientes mas grandes para los otros registros de destino.

$$\min_{\mathcal{G}_s, \mathcal{C}} \mathcal{L}_C(\mathbf{x}^s, \mathbf{y}^s) = \mathcal{L}_{CE}(C(\mathcal{G}_s(\mathbf{x}^s)), \mathbf{y}^s) \quad (2.4)$$

$$\max_{\mathcal{D}} \mathcal{L}_D(\mathbf{x}^s, \mathbf{x}^t) = \mathbb{E}_{\mathbf{x}^s \sim \hat{\mathcal{S}}} \log[\mathcal{D}(\mathcal{G}_s(\mathbf{x}^s))] + \mathbb{E}_{\mathbf{x}^t \sim \hat{\mathcal{T}}} \log[1 - \mathcal{D}(\mathcal{G}_t(\mathbf{x}^t))] \quad (2.5)$$

$$\min_{\mathcal{G}_t} \mathcal{L}_{\mathcal{G}_t}(\mathbf{x}^t) = -\mathbb{E}_{\mathbf{x}^t \sim \hat{\mathcal{T}}} \log[\mathcal{D}(\mathcal{G}_t(\mathbf{x}^t))] \quad (2.6)$$

2.3.3. AFN

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida

mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

2.3.4. MDD

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Capítulo 3

Metodología

En este capítulo se detallarán los procesos abordados para la extracción y clasificación de los dígitos escritos en los telegramas de las elecciones. Primero se describirá el proceso de obtención, limpieza y transformación de los datos a fin de obtener un dataset con el cual entrenar. Luego, se describirán los procesos a ser llevados a cabo para el diseño experimental de los modelos.

3.1. Descripción de los datos

Los telegramas fueron descargados desde la [página oficial del estado argentino](#). Presentan un formato estándar en forma de grilla donde en cada renglón se encuentra el partido político y los votos obtenidos para diputados y senadores. En la página también se puede descargar un CSV donde se encuentra el *id* de cada telegrama y los valores digitalizados oficialmente.

3.2. Extracción de dígitos de los telegramas

Los telegramas de las elecciones de Santa Fe presentan un formato tabular, donde cada fila representa un partido político y los votos obtenidos abiertos por senadores y diputados. Se debieron ejecutar múltiples pasos de extracción, transformación y carga (ETL por sus siglas) de los mismos antes de poder armar un dataset con el cual entrenar los modelos. Se utilizó la librería OpenCV (Bradski, 2000) para manipular las imágenes y poder llevar a cabo el proceso de ETL.

3.2.1. Enderezado

Como los telegramas son escaneados a mano, el primer paso consiste en enderezarlos. Este proceso puede realizarse buscando el rectángulo de mayor área, calculando el ángulo de rotación y rotando la imagen completa con la función `getRotationMatrix2D` de OpenCV.

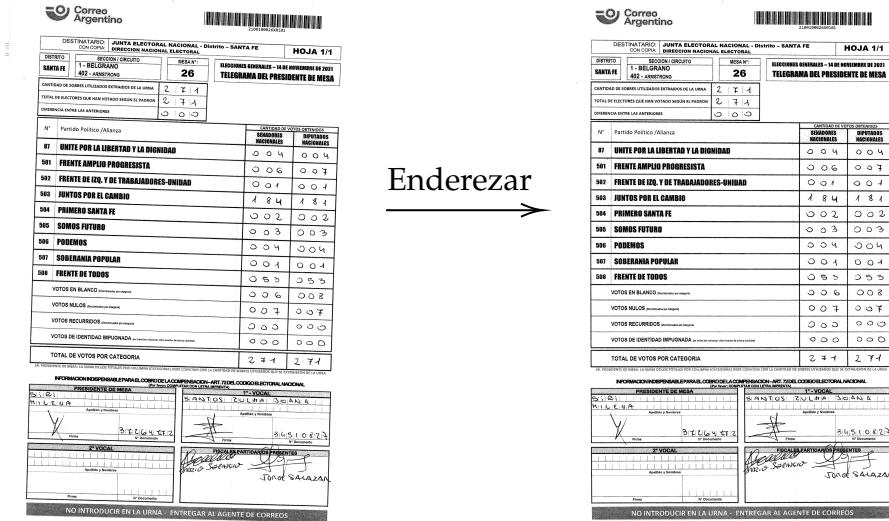


FIGURA 3.1: Enderezamiento de un telegrama utilizando OpenCV.

3.2.2. Extracción de la grilla de votos

El siguiente paso consiste en poder seleccionar la grilla de los votos y poder extraerla para seguir trabajando en ella. Esto puede realizarse utilizando la función `getContours` de OpenCV y quedándose con el contorno de mayor área.

Nº	Partido Político /Alianza	CANTIDAD DE VOTOS OBTENIDOS	
		SOCIALES NACIONALES	DEPARTAMENTOS NACIONALES
07	UNITE POR LA LIBERTAD Y LA DIGNIDAD	0 0 4	0 0 4
501	FRENTE AMPLIO PROGRESISTA	0 0 6	0 0 7
502	FRENTE DE IZQ. Y DE TRABAJADORES-UNIDAD	0 0 1	0 0 1
503	JUNTOS POR EL CAMBIO	1 8 4	1 8 4
504	PRIMERO SANTA FE	0 0 2	0 0 2
505	SOMOS FUTURO	0 0 3	0 0 3
506	PODEMOS	0 0 4	0 0 4
507	SOBERANIA POPULAR	0 0 1	0 0 1
508	FRENTE DE TODOS	0 5 5	0 5 5
VOTOS EN BLANCO (descartados por voto nulo)		0 0 6	0 0 2
VOTOS NULOS (descartados por voto nulo)		0 0 7	0 0 7
VOTOS RECURRIDOS (descartados por voto nulo)		0 0 0	0 0 0
VOTOS DE IDENTIDAD IMPUGNADA (no se ha informado sobre la votación)		0 0 0	0 0 0
TOTAL DE VOTOS POR CATEGORÍA		2 4 1	2 7 4

FIGURA 3.2: Grilla de votos extraída buscando el contorno de mayor área.

3.2.3. Detección de líneas de la grilla de votos

Teniendo la grilla de votos separada del telegrama, se procede a detectar las líneas para poder extraer cada registro de la misma. Si bien OpenCV posee funciones para la detección de líneas, se optó por utilizar un enfoque simplificado. El mismo es similar al utilizado en (Lamagna, 2016). Debido a que la grilla se encuentra enderezada, se puede utilizar proyecciones de los colores sobre los ejes x e y para detectarlas. Al sumar todos los colores por eje, se pueden encontrar los picos donde se encuentran las líneas horizontales y verticales. Posteriormente, se selecciona un umbral de corte por cada

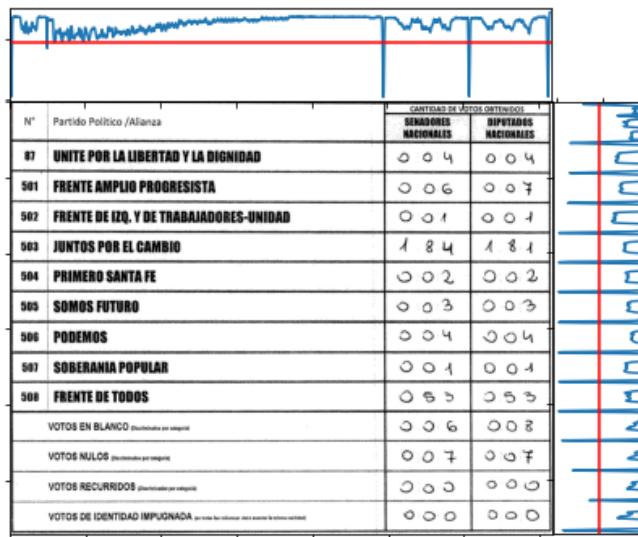


FIGURA 3.3: Proyecciones de los ejes de la grilla. En rojo se marca el umbral de corte.

eje, siendo el promedio menos dos desvíos estándar. La figura 3.3 muestra las proyecciones obtenidas con los umbrales por cada eje.

Sin embargo, aunque se seleccione aquellos píxeles que se encuentren por debajo del umbral, por cada pico existe mas de un pixel. Siguiendo con el ejemplo, en el eje x se obtienen los siguientes píxeles que se encuentran en los "picos":

```
array([    3,     4,     5,     6,     7,   100,   119,   120,   988,
        989,   990,   991,   992,  1215,  1216,  1217,  1218,  1219,
       1423,  1424,  1425,  1426,  1427])
```

Con la estrategia del umbral no se obtienen las 4 líneas que se desean detectar sobre el eje x . No obstante, se puede aplicar un clustering jerárquico sobre los índices obtenidos y luego calcular el índice promedio de cada cluster para tomarlo como único representante. El proceso se repite de la misma manera sobre el eje y . Esta última modificación sobre el trabajo realizado en (Lamagna, 2016) asegura tener un único pixel por línea.

FIGURA 3.4: Grilla detectada (en verde) utilizando el umbral por sobre las proyecciones y su posterior agrupamiento con clustering jerárquico.

Nº	Partido Político /Alianza	CANTIDAD DE VOTOS DETERMINADOS	
		SEÑADORES NACIONALES	DIPUTADOS NACIONALES
07	UNITE POR LA LIBERTAD Y LA DIGNIDAD	0 0 4	0 0 4
501	FRENTE AMPLIO PROGRESISTA	0 0 6	0 0 7
502	FRENTE DE IZQ. Y DE TRABAJADORES-UNIDAD	0 0 1	0 0 1
503	JUNTOS POR EL CAMBIO	1 8 4	1 8 4
504	PRIMERO SANTA FE	0 0 2	0 0 2
505	SOMOS FUTURO	0 0 3	0 0 3
506	PODEMOS	0 0 4	0 0 4
507	SOBERANIA POPULAR	0 0 1	0 0 4
508	FRENTE DE TODOS	0 5 5	0 5 5
	VOTOS EN BLANCO (descartados por voto nulo)	0 0 6	0 0 8
	VOTOS NULOS (descartados por voto nulo)	0 0 7	0 0 7
	VOTOS RECURRIDOS (descartados por voto nulo)	0 0 0	0 0 0
	VOTOS DE IDENTIDAD IMPUGNADA (en votos que no cumplieron con las normas establecidas)	0 0 0	0 0 0

FIGURA 3.4: Grilla detectada (en verde) utilizando el umbral por sobre las proyecciones y su posterior agrupamiento con clustering jerárquico.

3.2.4. Segmentación de dígitos

Una vez obtenida la grilla, se itera por cada registro obteniendo los rectángulos que contienen los votos, ignorando el primer renglón que contiene los títulos de la grilla.



FIGURA 3.5: Primer registro extraído de la grilla de votos.

Para separar cada dígito en una única imagen, se utiliza un análisis de componentes conectados (connectedComponents de OpenCV) y se recortan aquellos contornos que posean de área entre el 5 % y el 70 % de los pixeles totales de la imagen. Los valores fueron obtenidos mediante experimentación y sirven para descartar ruido que pueda detectar el análisis de componentes conectados.



FIGURA 3.6: Dígitos detectados en el primer bloque de votos.

Por último, los dígitos son encuadrados y se guardan junto al partido político y al tipo de voto (senadores o diputados) en un dataset, cuyo análisis será abordado en la siguiente sección.

3.3. Análisis del dataset

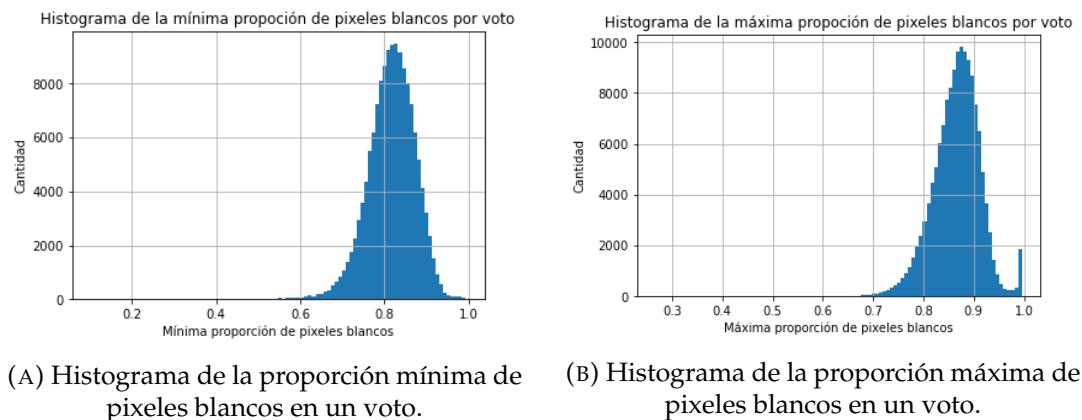
Con la extracción descripta en la sección anterior, se procede a hacer un análisis exploratorio de los datos en búsqueda de posibles errores. La tabla 3.1 muestra algunos registros del dataset que vincula los dígitos extraídos de los telegramas con los partidos políticos.

Telegrama	Partido	Tipo	Dígitos	# Dígitos
2100100026X	Unite	Diputados		3
2100100026X	Unite	Senadores		4
2100100067X	Frente de Todos	Diputados		4
2100100067X	Frente de Todos	Senadores		3

CUADRO 3.1: Ejemplo de registros del dataset. Cada fila representa un voto.

El lector podrá darse cuenta que existen errores en el cuadro anterior. En primer lugar, en la columna donde sólamente deben haber las imágenes de

los dígitos existe una con una línea en el segundo registro. Por otro lado, también existen dígitos mal escritos como los que se encuentran en los dos últimos registros. Para limpiar estas extracciones incorrectas, se calculan dos columnas las cuales contienen las proporciones mínimas y máximas de pixeles blancos en las imágenes por cada voto. La figura 3.7 muestra los histogramas generales de las mismas.



(A) Histograma de la proporción mínima de pixeles blancos en un voto. (B) Histograma de la proporción máxima de pixeles blancos en un voto.

FIGURA 3.7: Histogramas de proporciones de pixeles blancos por voto.

La cola derecha de la distribución de 3.7b presenta un pico anormal en el extremo cercano a 1. Al eliminar aquellas imágenes de dígitos que posean mas de un 95 % de pixeles blancos, se logran descartar los casos de líneas como la vista en el cuadro 3.1. Por el contrario, al eliminar aquellas imágenes que posean menos del 50 % de pixeles blancos, se logran descartar los casos de escritura incorrecta.

El siguiente punto a analizar es el tamaño de las imágenes. Como fue mostrado en el cuadro 3.1, las imágenes no fueron estandarizadas en un tamaño estándar. De manera análoga a la variable de proporción de pixeles blancos, se calculan las variables mínimo y máximo tamaño por voto. La tabla 3.2 muestra los estadísticos descriptivos de las mismas.

Variable	Promedio	Desvío	Mínimo	Mediana	Máximo
Tamaño mínimo	38.35	9.61	13	37	621
Tamaño máximo	44.80	10.73	18	43	621

CUADRO 3.2: Estadísticos descriptivos del mínimo y máximo tamaño de las imágenes de los votos.

Al analizar los valores extremos, se encuentran telegramas que fueron cargados de forma incorrecta (ver anexo A.2) o con caligrafía en la cual los números se encuentran unidos y no pueden ser separados por el análisis de componentes conectados de OpenCV (ver anexo A.3). Se eliminan estos casos filtrando por aquellos registros que se encuentren por fuera de la media ± 4 desvíos estándar.

Concluida la etapa de análisis y limpieza, se exportan los dígitos en un formato similar al dataset *MNIST* con la etiqueta que oficialmente fue registrada en las elecciones del 2021. El mismo será nombrado como *TDS* (dataset de telegramas) de aquí en adelante. El mismo posee un total de 170.718 imágenes de dígitos cuadradas con sus dimensiones orginales, es decir, no se encuentran estandarizados en una dimensión específica. Esto último es adrede ya que puede servir para mejorar el dataset en futuros trabajos.



FIGURA 3.8: Dataset TDS de dígitos.

3.4. Diseño experimental

Los experimentos fueron programados en Python mediante la librería PyTorch (Paszke et al., 2019) y TLLIB para aplicar las técnicas de adaptación de dominio.

Se entrenaron redes ResNet-50 (He et al., 2016) y LeNet-8 (LeCun et al., 1998) por cada una de las técnicas de adaptación de dominio descriptas en el capítulo 2. Como modelo baseline, se entrenaron ambas redes sin ninguna técnica de adaptación.

3.4.1. Metodología de entrenamiento

Para entrenar modelos mediante adaptación de dominio, se precisan dos datasets: el de origen que se encuentra etiquetado y el de destino sin etiquetar. Los experimentos llevados a cabo utilizan el *MNIST* como origen y el *TDS* como destino. En cada iteración de entrenamiento, se provee de un batch de 1024 imágenes del dataset origen y otro del de destino. Se realizaron 3 particiones de cada uno:

- Entrenamiento (70 % de cada dataset): utilizado para entrenar la ResNet-18 en cada época.

- Validación (15 % de cada dataset): utilizado para calcular los mejores hiperparámetros del modelo.
- Test (15 % de cada dataset): utilizado para calcular las métricas y gráficos finales del modelo entrenado.

Los modelos se entrenaron utilizando Stochastic Gradient Descent (SGD) (Sutskever et al., 2013) utilizando nesterov, 0.9 como momento y 0.001 de weight decay.

3.4.2. Selección y optimización del modelo

La optimización de hiperparámetros se realizó utilizando la librería Optuna (Akiba et al., 2019). Se ejecutaron 15 rondas de optimización con 10 épocas de entrenamiento por cada experimento. En cada uno de los experimentos, se optimizaron los hiperparámetros para minimizar el loss de cada modelo. El cuadro 3.3 muestra los rangos búsquedas.

	ADDA	AFN	DANN	MDD	Sin AD
LR*	[1e-4, 0.1]	[1e-4, 0.1]	[1e-4, 0.1]	[1e-4, 0.1]	[1e-4, 0.1]
Pretrain LR*	[1e-4, 0.1]	-	-	-	-
Pretrain Epochs	[5, 10]	-	[0.5, 2]	[0.5, 2]	-
T-O** Losses	[0.5, 2]	-	[0.5, 2]	[0.5, 2]	-
T-O** Norm	-	[0.001, 0.1]	-	-	-
T-O** Entropy	-	[0.0, 0.1]	-	-	-
Delta	-	[0.01, 5]	-	-	-
# Blocks	-	[1, 4]	-	-	-
Dropout	-	[0.3, 0.7]	-	-	-
Margin	-	-	-	[1, 10]	-

CUADRO 3.3: Rango de hiperparámetros optimizados. *LR refiere a "learning rate". **T-O refiere a "trade-off".

Se implementó una disminución del learning rate en cada iteración de cada época, mediante la fórmula 3.1. Se mantuvieron constantes los valores de $\gamma = 0.001$ y $\theta = 0.25$.

$$\eta_t = \eta_0 * (1 + \gamma * t)^{-\theta} \quad (3.1)$$

3.5. Métricas de evaluación

Los modelos optimizados serán evaluados con distintas métricas en tiempo de test, descriptas en las siguientes sub-secciones del capítulo. Las mismas pretenden evaluar qué tan buenos son los modelos respecto a la tarea de clasificación y qué capacidad de adaptación de dominio poseen.

3.5.1. Accuracy

La métrica de *accuracy* permite identificar qué tan cerca o lejos un conjunto de observaciones se encuentra respecto a los valores reales. Es el ratio de predicciones correctas sobre las totales.

$$\text{Accuracy}(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} 1(\hat{y}_i = y_i) \quad (3.2)$$

Donde:

- n : es la cantidad de observaciones totales.
- y_i : es el valor de la i -ésima observación correspondiente al real.
- \hat{y}_i : es el valor predicho para la i -ésima observación.
- $1(x)$: es la función indicador.

3.5.2. F_1

La métrica F_1 es una media armónica de otras dos: *precisión* y *recall*. De manera simplificada, la primera muestra la capacidad del modelo de no etiquetar como positivo una obsevación que es negativa y la segunda muestra la capacidad del modelo de encontrar todas las observaciones positivas.

La *precisión* consiste en calcular el ratio de predicciones positivas correctas de la clase respecto al total de predicciones positivas de la clase.

$$\text{Precision}(y, \hat{y}) = \frac{TP}{TP + FP} \quad (3.3)$$

Donde:

- TP : es la cantidad de verdaderos positivos.
- FP : es la cantidad de falsos positivos.

Por otro lado, el *recall* consiste en calcular el ratio de predicciones positivas correctas de la clase respecto al total de predicciones correctas de la clase.

$$\text{Recall}(y, \hat{y}) = \frac{TP}{TP + FN} \quad (3.4)$$

Donde:

- TP : es la cantidad de verdaderos positivos.
- FN : es la cantidad de falsos negativos.

Las dos métricas mencionada anteriormente pueden ser combinadas en una sola denominada F_β . El valor de β permite asignar un peso distinto a la precisión o al recall dentro del promedio armónico. Cuando $\beta = 1$, ambas poseen el mismo peso.

$$F_\beta(y, \hat{y}) = (1 + \beta^2) \times \frac{\text{precision}(y, \hat{y}) \times \text{recall}(y, \hat{y})}{\beta^2 \times \text{precision}(y, \hat{y}) + \text{recall}(y, \hat{y})} \quad (3.5)$$

El rango de valores es de $[0, 1]$, donde 1 corresponde a un clasificador que funciona sin errores.

3.5.3. Intersección sobre unión

Otra forma de medir el clasificador es mediante la *intersección sobre unión* (IoU por sus siglas en inglés). Consta de calcular la cantidad de dígitos únicos predichos por sobre la cantidad de dígitos únicos reales. Por ejemplo, para un telegrama y un partido el clasificador predice 189 votos cuando lo real es 180. Entonces, el *IoU* viene dado por:

$$\begin{aligned} \text{IoU}(\{1, 8, 0\}, \{1, 8, 9\}) &= \frac{|\{1, 8, 0\} \cap \{1, 8, 9\}|}{|\{1, 8, 0\} \cup \{1, 8, 9\}|} \\ &= \frac{|\{1, 8\}|}{|\{1, 8, 0, 9\}|} \\ &= \frac{2}{4} \\ &= 0.5 \end{aligned} \quad (3.6)$$

De forma general: sea y_i el dígito i -ésimo del voto real y de longitud n , \hat{y}_j el dígito j -ésimo del voto predicho por el modelo \hat{y} de longitud m , entonces la métrica viene dada por:

$$\text{IoU}(y, \hat{y}) = \frac{|\{y_i\} \cap \{\hat{y}_j\}|}{|\{y_i\} \cup \{\hat{y}_j\}|} \forall i \in \{1, \dots, n\}, \forall j \in \{1, \dots, m\} \quad (3.7)$$

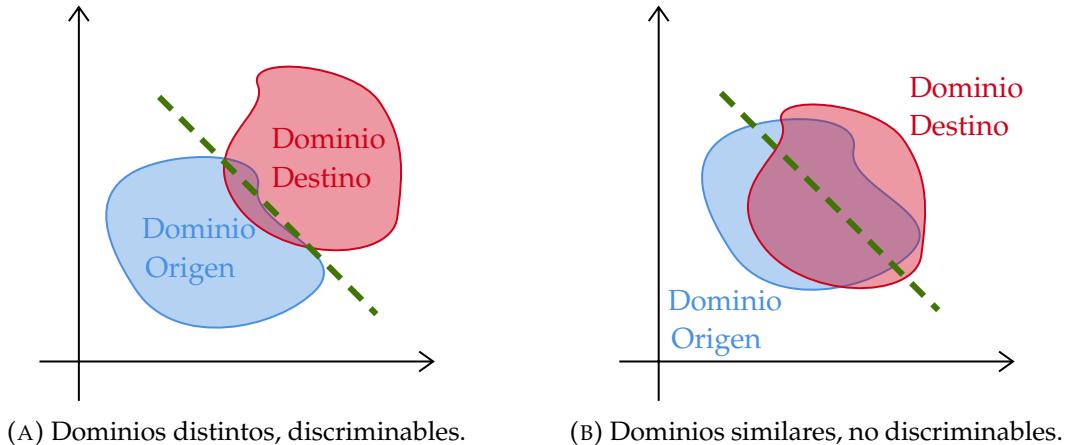
3.5.4. Distancia \mathcal{A}

La distancia \mathcal{A} mide la similaridad entre dos distribuciones. Puede ser utilizada para analizar los espacios latentes de clasificadores en problemas de adaptación de dominio (Ben-David et al., 2006). La métrica viene dada por:

$$\text{dist}_{\mathcal{A}} = 2(1 - 2\epsilon) \quad (3.8)$$

Donde ϵ es el error en test de un clasificador entrenado para discriminar el dominio de origen del de destino. Cuando el error de clasificación es bajo, significa que hay diferencias significativas entre las dos distribuciones, haciendo que $\text{dist}_{\mathcal{A}}$ sea grande y vice versa.

En la figura 3.9 se muestran posibles casos de distribuciones de dominios. En la sub-figura 3.9a las distribuciones de dominios son significativamente diferentes entre si, por lo tanto $\text{dist}_{\mathcal{A}} \approx 2$. Por el contrario, en la sub-figura 3.9b las distribuciones de dominios son similares entre si, por lo que se espera que $\text{dist}_{\mathcal{A}} \approx 0$.



Durante el proyecto se utiliza como clasificador una regresión logística a fin de calcular la distancia \mathcal{A} .

3.5.5. Discrepancia Máxima Promedio

Gretton et al., 2012 introduce un test estadístico que permite identificar si una distribución es distinta a otra llamado Discrepancia Máxima Promedio (MMD por sus siglas en inglés). El test permite analizar qué tan bien el modelo permite adaptar el aprendizaje del dataset de origen con el de destino. Cuando el test resulte no significativo se puede concluir que los espacios latentes generados por el modelo para ambos datasets difieren y no se logró una buena adaptación.

De manera formal, dado un espacio \mathbb{R}^d y muestras IID $X_i \in \mathbb{R}^d, i = 1, \dots, N_X$ de la distribución $X \sim P_X$ y $Y_i \in \mathbb{R}^d, i = 1, \dots, N_Y$ de la distribución $Y \sim P_Y$, se quiere demostrar que P_X es distinto de P_Y . El test planteado es:

$$\begin{aligned} H_0) P_X &\neq P_Y \\ H_1) P_X &= P_Y \end{aligned} \tag{3.9}$$

Se utilizó la librería **TorchDrift** para implementarlo en la evaluación de los modelos. La misma utiliza una variación de test MMD utilizando un kernel RBF para aproximar P_X y P_Y .

TODO: Completar como se calcula el score del test

Capítulo 4

Análisis de resultados

A lo largo del presente capítulo, se analizarán los resultados obtenidos a partir de los experimentos planteados. Se detallarán las métricas obtenidas y los resultados de aplicar los modelos a los telegramas.

4.1. Análisis de métricas

Los experimentos realizados arrojaron los resultados mostrados en el cuadro 4.1. Las métricas referidas a la capacidad de clasificación (Accuracy, F_1) son evaluadas sobre la partición de test del dataset de origen donde se conocen las etiquetas a ciencia cierta (*MNIST*). Por otro lado, las métricas de adaptación (MMD , Dist. \mathcal{A}) son evaluadas sobre los espacios latentes que los modelos generaron para las particiones de test de ambos datasets. El mejor modelo es el que posea mejor capacidad de clasificación y de adaptación.

AD	Modelo	Acc.	F_1	MMD	Dist. \mathcal{A}
-	ResNet	(1) 98.73	(1) 0.9873	0.0595	1.9443
	LeNet	98.14	0.9814	0.0498	1.9313
MDD	ResNet	98.58	0.9858	0.0569	1.9031
	LeNet	(3) 98.62	(3) 0.9862	0.0363	1.7259
DANN	ResNet	97.52	0.9752	0.0153	1.6383
	LeNet	98.14	0.9814	(2) 0.0119	(3) 1.6231
AFN	ResNet	(2) 98.68	(2) 0.9868	(1) 0.0037	(1) 1.0047
	LeNet	98.58	0.9857	(3) 0.0138	(2) 1.5920
ADDA	ResNet	89.18	0.8918	0.0225	1.8691
	LeNet	97.49	0.9749	0.0288	1.8112

CUADRO 4.1: Métricas de los experimentos realizados. Entre paréntesis se encuentra la posición que ocupa dentro del top 3 de la columna.

Como era de esperarse, los modelos que fueron entrenados sin adaptación de dominio son los que mejores métricas de clasificación poseen. No obstante, sus valores de adaptación son los peores provocando que no puedan ser aplicados en *TDS*.

El modelo que mejor combinación de clasificación y adaptación es la ResNet utilizando AFN. Cabe destacar que los modelos LeNet entrenados con DANN y AFN obtuvieron buenas métricas en general, lo que deja en evidencia que *no siempre un modelo más complejo es mejor*.

Es posible aplicar los modelos sobre todos los telegramas y calcular el *IoU* promedio por telegrama con el cálculo descripto en el capítulo 3 y la cantidad de aciertos promedio por telegrama utilizando como etiquetas lo transcripto en el centro de cómputo suponiendo que existen pocos errores en ellos.

AD	Modelo	<i>IoU</i> prom.	# aciertos prom.	% aciertos prom.
-	ResNet	0.4494	4	22 %
	LeNet	0.4715	6	33 %
MDD	ResNet	0.5451	8	44 %
	LeNet	0.5801	9	50 %
DANN	ResNet	0.6941	12	67 %
	LeNet	0.7024	12	67 %
AFN	ResNet	0.7486	13	72 %
	LeNet	0.6493	11	61 %
ADDA	ResNet	0.6763	11	61 %
	LeNet	0.6406	10	56 %

CUADRO 4.2: *IoU* promedio y cantidad promedio de aciertos al aplicar los modelos a cada telegrama.

El cuadro 4.2 confirma la elección del mejor modelo realizada anteriormente. Independientemente de qué técnica de adaptación se utilice, todas presentan mejores porcentajes de aciertos promedio que los modelos que fueron entrenados únicamente con MNIST.

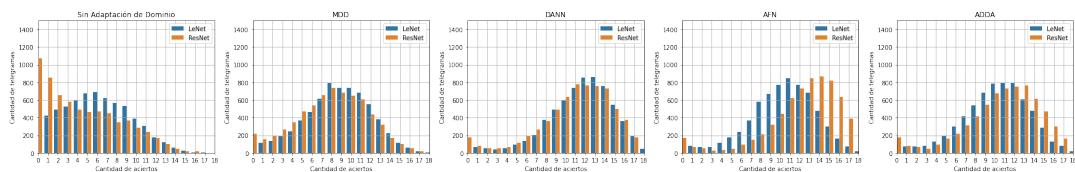


FIGURA 4.1: Distribución de aciertos de cantidad de votos por cada par técnica AD y modelo.

4.2. Comparación de espacios latentes

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices.

Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

4.3. Ánalisis de errores

Los errores de predicción de los modelos pueden reflejarse en las distribución de la métrica *IoU* para cada uno de ellos.

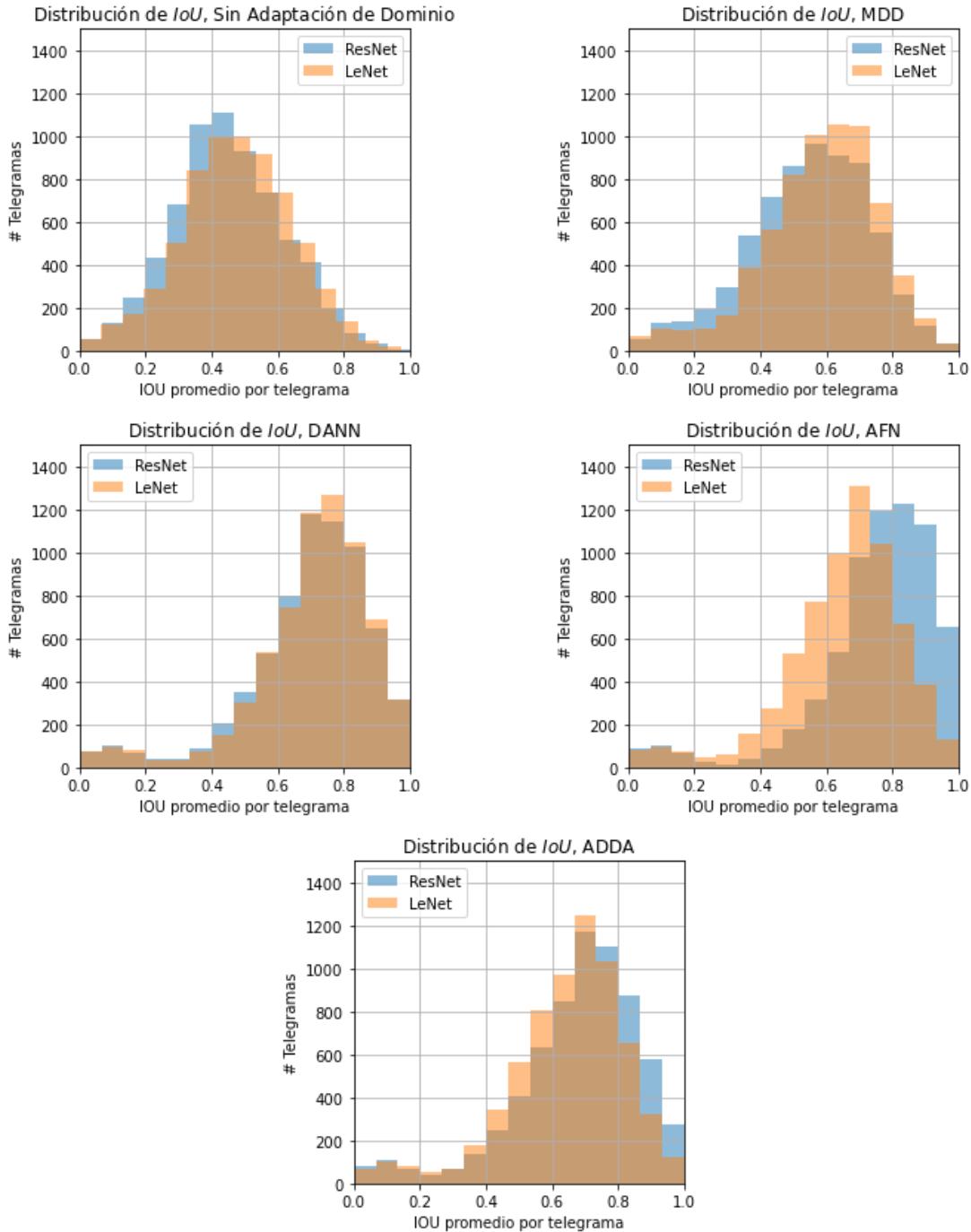


FIGURA 4.2: Histogramas de la métrica IoU promedio por telegrama por cada par técnica AD y modelo.

Resulta interesante mencionar existen telegramas que son más difíciles de analizar que otros. Esto puede evidenciarse en los histogramas de la figura 4.2 donde se pueden observar un conjunto de observaciones que contienen valores entre $[0, 0.2]$ en todos los experimentos realizados. Luego de analizar cada uno de estos casos, se detectaron las siguientes situaciones:

- Telegramas cargados de forma errónea: ver ejemplo en el anexo A.2.

- Telegramas correctos pero por alguna cuestión la lógica de extracción de dígitos no funciona correctamente: ver ejemplo en el anexo A.3.
- Telegramas donde existen otros caracteres distintos a números: al ser un cuadro de texto libre sin formato, los jefes de mesa pueden escribir lo que deseen. Ver ejemplo en el anexo A.4 donde se representa el 0 a la izquierda con X.
- Telegramas de mesas donde la mayor cantidad de votos se las lleva un único partido y completan los votos a la izquierda con 0: al agregar los ceros a la izquierda, aumenta la probabilidad de que el modelo se equivoque con esos ceros que no aportan al número final. Ver ejemplo en el anexo A.5.

En los primeros dos puntos se describen problemas que fueron detectados en el proceso de ETL del capítulo 3. Estandarizar los telegramas agregando un casillero por cada dígito junto a mejorar el proceso de extracción de los mismos, supondrá una mejora considerable en las capacidades predictivas de los modelos.

El tercer punto presenta un problema dentro de la adaptación de dominio. La misma supone que, si bien los datasets de origen y destino son diferentes pero representan lo mismo, debe existir la misma cantidad de clases entre origen y destino. Al agregar uno o varios caracteres adicionales en *TDS* (como es el ejemplo donde representaban el 0 con una X), se está incumpliendo este supuesto.

El cuarto punto aumenta la probabilidad de error en los modelos debido a que el 0 a la izquierda no aporta significado alguno al número de la cantidad de votos que se desea predecir.

Estandarizar la carga de los telegramas por parte de los jefes de mesa mediante alguna capacitación permitiría reducir los errores de los puntos tres y cuatro en elecciones futuras.

Capítulo 5

Conclusiones

5.1. Conclusiones

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

5.2. Trabajos Futuros

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Apéndice A

Anexo: Telegramas

A.1. Ejemplo de telegrama

Correo Argentino

DESTINATARIO: JUNTA ELECTORAL NACIONAL - Distrito - SANTA FE
CON COPIA: DIRECCION NACIONAL ELECTORAL

HOJA 1/1

DISTRITO SANTA FE	SECCION / CIRCUITO 1 - BELGRANO 402 - ARMSTRONG	MESA N°: 1	ELECCIONES GENERALES - 14 DE NOVIEMBRE DE 2021 TELEGRAMA DEL PRESIDENTE DE MESA
CANTIDAD DE SOBRES UTILIZADOS EXTRAIOS DE LA URNA		2 8 3	
TOTAL DE ELECTORES QUE HAN VOTADO SEGUN EL PADRON		2 8 3	
DIFERENCIA ENTRE LAS ANTERIORES		0 0 0	

Nº	Partido Político /Alianza	CANTIDAD DE VOTOS OBTENIDOS	
		SENADORES NACIONALES	DIPUTADOS NACIONALES
87	UNITE POR LA LIBERTAD Y LA DIGNIDAD	4	5
501	FRENTE AMPLIO PROGRESISTA	15	16
502	FRENTE DE IZQ. Y DE TRABAJADORES-UNIDAD	8	8
503	JUNTOS POR EL CAMBIO	135	135
504	PRIMERO SANTA FE	2	2
505	SOMOS FUTURO	7	7
506	PODEMOS	1	1
507	SOBERANIA POPULAR	6	6
508	FRENTE DE TODOS	94	92
VOTOS EN BLANCO (determinados por categoria)		5	5
VOTOS NULOS (determinados por categoria)		6	6
VOTOS RECURRIDOS (determinados por categoria)		—	—
VOTOS DE IDENTIDAD IMPUGNADA (en todas las columnas debe asentar la misma cantidad)		—	—
TOTAL DE VOTOS POR CATEGORIA		283	283

SR. PRESIDENTE DE MESA: LA SUMA DE LOS totales por columna (CATEGORIA) DEBE COINCIDIR CON LA CANTIDAD DE SOBRES UTILIZADOS QUE SE EXTRAJERON DE LA URNA

INFORMACION INDISPENSABLE PARA EL COBRO DE LA COMPENSACION-ART.72 DEL CODIGO ELECTORAL NACIONAL
(Por favor: COMPLETAR CON LETRA IMPRESA)

PRESIDENTE DE MESA		1º - VOCAL	
Apellido y Nombres		Apellido y Nombres	
	3 8 0 8 7 4 6 1 8		3 7 2 2 8 1 2 8
Firma	Nº Documento	Firma	Nº Documento
2º VOCAL		FISCALES PARTIDARIOS PRESENTES	
Apellido y Nombres		Apellido y Nombres	
	4 2 3 2 8 1 1 7		3 0 2 5 0 4 9 5
Firma	Nº Documento	Firma	Nº Documento

NO INTRODUCIR EN LA URNA - ENTREGAR AL AGENTE DE CORREOS

A.2. Ejemplo de telegrama mal cargado

A.3. Ejemplo de telegrama con números sin espacio entre ellos



2106200153X0101

DESTINATARIO: CON COPIA:	JUNTA ELECTORAL NACIONAL - Distrito - SANTA FE DIRECCION NACIONAL ELECTORAL		HOJA 1/1																																																																
DISTRITO SANTA FE	SECCION / CIRCUITO 2 - CASEROS 367 - BIGAND	MESA N°: 153	ELECCIONES GENERALES - 14 DE NOVIEMBRE DE 2021 TELEGRAMA DEL PRESIDENTE DE MESA																																																																
CANTIDAD DE SOBRES UTILIZADOS EXTRAIDOS DE LA URNA		262																																																																	
TOTAL DE ELECTORES QUE HAN VOTADO SEGUN EL PADRON		262																																																																	
DIFERENCIA ENTRE LAS ANTERIORES		000																																																																	
<table border="1"> <thead> <tr> <th>Nº</th> <th>Partido Político /Alianza</th> <th colspan="2">CANTIDAD DE VOTOS OBTENIDOS</th> </tr> <tr> <th></th> <th></th> <th>SENADORES NACIONALES</th> <th>DIPUTADOS NACIONALES</th> </tr> </thead> <tbody> <tr> <td>87</td> <td>UNITE POR LA LIBERTAD Y LA DIGNIDAD</td> <td>006</td> <td>005</td> </tr> <tr> <td>501</td> <td>FRENTE AMPLIO PROGRESISTA</td> <td>027</td> <td>027</td> </tr> <tr> <td>502</td> <td>FRENTE DE IZQ. Y DE TRABAJADORES-UNIDAD</td> <td>003</td> <td>003</td> </tr> <tr> <td>503</td> <td>JUNTOS POR EL CAMBIO</td> <td>100</td> <td>101.</td> </tr> <tr> <td>504</td> <td>PRIMERO SANTA FE</td> <td>008</td> <td>008</td> </tr> <tr> <td>505</td> <td>SOMOS FUTURO</td> <td>005</td> <td>005</td> </tr> <tr> <td>506</td> <td>PODEMOS</td> <td>001</td> <td>001</td> </tr> <tr> <td>507</td> <td>SOBERANIA POPULAR</td> <td>011</td> <td>011</td> </tr> <tr> <td>508</td> <td>FRENTE DE TODOS</td> <td>087</td> <td>086</td> </tr> <tr> <td colspan="2">VOTOS EN BLANCO (discriminados por categoría)</td> <td>009</td> <td>009</td> </tr> <tr> <td colspan="2">VOTOS NULOS (discriminados por categoría)</td> <td>011</td> <td>012</td> </tr> <tr> <td colspan="2">VOTOS RECURRIDOS (discriminados por categoría)</td> <td>000</td> <td>000</td> </tr> <tr> <td colspan="2">VOTOS DE IDENTIDAD IMPUGNADA (en todas las columnas debe aparecer la misma cantidad)</td> <td>000</td> <td>000</td> </tr> <tr> <td colspan="2">TOTAL DE VOTOS POR CATEGORIA</td> <td>262</td> <td>262</td> </tr> </tbody> </table>			Nº	Partido Político /Alianza	CANTIDAD DE VOTOS OBTENIDOS				SENADORES NACIONALES	DIPUTADOS NACIONALES	87	UNITE POR LA LIBERTAD Y LA DIGNIDAD	006	005	501	FRENTE AMPLIO PROGRESISTA	027	027	502	FRENTE DE IZQ. Y DE TRABAJADORES-UNIDAD	003	003	503	JUNTOS POR EL CAMBIO	100	101.	504	PRIMERO SANTA FE	008	008	505	SOMOS FUTURO	005	005	506	PODEMOS	001	001	507	SOBERANIA POPULAR	011	011	508	FRENTE DE TODOS	087	086	VOTOS EN BLANCO (discriminados por categoría)		009	009	VOTOS NULOS (discriminados por categoría)		011	012	VOTOS RECURRIDOS (discriminados por categoría)		000	000	VOTOS DE IDENTIDAD IMPUGNADA (en todas las columnas debe aparecer la misma cantidad)		000	000	TOTAL DE VOTOS POR CATEGORIA		262	262	
Nº	Partido Político /Alianza	CANTIDAD DE VOTOS OBTENIDOS																																																																	
		SENADORES NACIONALES	DIPUTADOS NACIONALES																																																																
87	UNITE POR LA LIBERTAD Y LA DIGNIDAD	006	005																																																																
501	FRENTE AMPLIO PROGRESISTA	027	027																																																																
502	FRENTE DE IZQ. Y DE TRABAJADORES-UNIDAD	003	003																																																																
503	JUNTOS POR EL CAMBIO	100	101.																																																																
504	PRIMERO SANTA FE	008	008																																																																
505	SOMOS FUTURO	005	005																																																																
506	PODEMOS	001	001																																																																
507	SOBERANIA POPULAR	011	011																																																																
508	FRENTE DE TODOS	087	086																																																																
VOTOS EN BLANCO (discriminados por categoría)		009	009																																																																
VOTOS NULOS (discriminados por categoría)		011	012																																																																
VOTOS RECURRIDOS (discriminados por categoría)		000	000																																																																
VOTOS DE IDENTIDAD IMPUGNADA (en todas las columnas debe aparecer la misma cantidad)		000	000																																																																
TOTAL DE VOTOS POR CATEGORIA		262	262																																																																

SR. PRESIDENTE DE MESA: LA SUMA DE LOS totales POR COLUMNA (CATEGORIA) DEBE COINCIDIR CON LA CANTIDAD DE SOBRES UTILIZADOS QUE SE EXTRAJERON DE LA URNA.

INFORMACION INDISPENSABLE PARA EL CORPODE DE LA COMPENSACION-ART.72 DEL CODIGO ELECTORAL NACIONAL

(Por favor: COMPLETAR CON LETRA IMPRESA)

PRESIDENTE DE MESA		1^{er} VOCAL	
LIUTI YANINA		FRAGO LETTI YANNA PAOLA	
Apellido y Nombres		Apellido y Nombres	
Firma	Nº Documento	Firma	Nº Documento
923841120		32454035	
2^{do} VOCAL		FISCALES PARTIDARIOS PRESENTES	
GONZALEZ MARINA		JUAN M. LUSAADI 11326427 JUNTOS POR EL CAMBIO	
Apellido y Nombres		 S. GONZALEZ FIRMA	
Mariangela		FIRMA	
95869553		FIRMA	
		FIRMA	

NO INTRODUCIR EN LA URNA - ENTREGAR AL AGENTE DE CORREOS

A.4. Ejemplo de telegrama con caracteres distintos a números



2108903008X0101

DESTINATARIO: CON COPIA:	JUNTA ELECTORAL NACIONAL - Distrito - SANTA FE DIRECCION NACIONAL ELECTORAL		HOJA 1/1
DISTRITO SANTA FE	SECCION / CIRCUITO 9 - LA CAPITAL	MESA N°: 3008	ELECCIONES GENERALES - 14 DE NOVIEMBRE DE 2021 TELEGRAMA DEL PRESIDENTE DE MESA
CANTIDAD DE SOBRES UTILIZADOS EXTRADIDOS DE LA URNA		2 1 6	
TOTAL DE ELECTORES QUE HAN VOTADO SEGUN EL PADRON		2 1 6	
DIFERENCIA ENTRE LAS ANTERIORES		1 X X	
Nº	Partido Político /Alianza	CANTIDAD DE VOTOS OBTENIDOS	
		SENADORES NACIONALES	DIPUTADOS NACIONALES
		X 2 2	X 2 2
		X 3 3	X 2 9
		XX 5	XX 5
		X 7 9	X 8 1
		X X X	X X X
		X X 6	X X 6
		XX 1	XX 1
		X 5 1	X 5 0
VOTOS EN BLANCO (Descontados por categoria)			
X X 6			
VOTOS NULOS (Descontados por categoria)			
X X 8			
VOTOS RECURRIDOS (Descontados por categoria)			
X X 7			
VOTOS DE IDENTIDAD IMPUGNADA (en todas las columnas debe asentar la misma cantidad)			
X X X			
TOTAL DE VOTOS POR CATEGORIA		2 0 2	2 0 1

SR. PRESIDENTE DE MESA: LA SUMA DE LOS TOTALES POR COLUMNA (CATEGORIA) DEBE COINCIDIR CON LA CANTIDAD DE SOBRES UTILIZADOS QUE SE EXTRAJERON DE LA URNA.

INFORMACION INDISPENSABLE PARA EL COBRO DE LA COMPENSACION-ART. 72 DEL CODIGO ELECTORAL NACIONAL
(Por favor: COMPLETAR CON LETRA IMPRENTA)

PRESIDENTE DE MESA		1º - VOCAL	
ARMANDO LUCILA BELGEM		ANCIO INNIA ISOL	
Apellido y Nombres		Apellido y Nombres	
Firma	3154681160	Firma	4112561227
Nº Documento		Nº Documento	
2º VOCAL		FISCALES PARTIDARIOS PRESENTES	
ARMANDO LUCILA BELGEM			
Apellido y Nombres			
Firma	3810141438	Nº Documento	
NO INTRODUCIR EN LA URNA - ENTREGAR AL AGENTE DE CORREOS			

A.5. Ejemplo de telegrama ceros a la izquierda de la cantidad de votos



DESTINATARIO: CON COPIA:	JUNTA ELECTORAL NACIONAL - Distrito - SANTA FE DIRECCION NACIONAL ELECTORAL		HOJA 1/1
DISTRITO SANTA FE	SECCION / CIRCUITO 13 - ROSARIO 319 - ROSARIO SEC. 2A	MESA N°: 4086	ELECCIONES GENERALES - 14 DE NOVIEMBRE DE 2021 TELEGRAMA DEL PRESIDENTE DE MESA
CANTIDAD DE SOBRES UTILIZADOS EXTRAIDOS DE LA URNA	2 618		
TOTAL DE ELECTORES QUE HAN VOTADO SEGUN EL PADRON	2 618		
DIFERENCIA ENTRE LAS ANTERIORES	0 0 0 0		
Nº	Partido Político /Alianza	CANTIDAD DE VOTOS OBTENIDOS	
		SENADORES NACIONALES	DIPUTADOS NACIONALES
87	UNITE POR LA LIBERTAD Y LA DIGNIDAD	0 0 5	0 0 7
501	FRENTE AMPLIO PROGRESISTA	0 5 3	0 4 6
502	FRENTE DE IZQ. Y DE TRABAJADORES-UNIDAD	0 0 2	0 0 2
503	JUNTOS POR EL CAMBIO	1 4 4	1 4 6
504	PRIMERO SANTA FE	0 0 6	0 0 3
505	SOMOS FUTURO	0 0 0	0 0 0
506	PODEMOS	0 0 1	0 0 2
507	SOBERANIA POPULAR	0 0 7	0 1 9
508	FRENTE DE TODOS	0 3 7	0 3 0
VOTOS EN BLANCO (describirlos por categoria)		0 0 7	0 0 7
VOTOS NULOS (describirlos por categoria)		0 0 6	0 0 6
VOTOS RECURRIDOS (describirlos por categoria)		0 0 0	0 0 0
VOTOS DE IDENTIDAD IMPUGNADA (en todas las columnas debe asentar la misma cantidad)		0 0 0	0 0 0
TOTAL DE VOTOS POR CATEGORIA		2 6 8	2 6 8

SR. PRESIDENTE DE MESA: LA SUMA DE LOS totales por columna (CATEGORIA) DEBE COINCIDIR CON LA CANTIDAD DE SOBRES UTILIZADOS QUE SE EXTRAJERON DE LA URNA.

INFORMACION INDISPENSABLE PARA EL COBRO DE LA COMPENSACION-ART. 72 DEL CODIGO ELECTORAL NACIONAL (Por favor: COMPLETAR CON LETRA IMPRESA)			
PRESIDENTE DE MESA PALAD AGUSTINA		1º - VOCAL	
Apellido y Nombres 37295734 Firma N° Documento		Apellido y Nombres Firma N° Documento	
2º VOCAL		ESCALES PARTIDARIOS PRESENTES	
Apellido y Nombres Paula Alcoba 35570523 Firma N° Documento			
NO INTRODUCIR EN LA URNA - ENTREGAR AL AGENTE DE CORREOS			

Bibliografía

- Akiba, Takuya et al. (2019). «Optuna: A Next-generation Hyperparameter Optimization Framework». En: *Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- Ben-David, Shai et al. (2006). «Analysis of Representations for Domain Adaptation». En: *Advances in Neural Information Processing Systems*. Ed. por B. Schölkopf, J. Platt y T. Hoffman. Vol. 19. MIT Press. URL: <https://proceedings.neurips.cc/paper/2006/file/b1b0432ceafb0ce714426e9114852ac7-Paper.pdf>.
- Bradski, G. (2000). «The OpenCV Library». En: *Dr. Dobb's Journal of Software Tools*.
- Ganin, Yaroslav et al. (2016). «Domain-adversarial training of neural networks». En: *The journal of machine learning research* 17.1, págs. 2096-2030.
- Glorot, Xavier, Antoine Bordes y Yoshua Bengio (2011). «Domain adaptation for large-scale sentiment classification: A deep learning approach». En: *ICML*.
- Gong, Boqing, Kristen Grauman y Fei Sha (2013). «Connecting the dots with landmarks: Discriminatively learning domain-invariant features for unsupervised domain adaptation». En: *International conference on machine learning*. PMLR, págs. 222-230.
- Goodfellow, Ian et al. (2020). «Generative adversarial networks». En: *Communications of the ACM* 63.11, págs. 139-144.
- Gretton, Arthur et al. (2012). «A kernel two-sample test». En: *The Journal of Machine Learning Research* 13.1, págs. 723-773.
- He, Kaiming et al. (2016). «Deep residual learning for image recognition». En: *Proceedings of the IEEE conference on computer vision and pattern recognition*, págs. 770-778.
- Jiang, Junguang et al. (2022). *Transferability in Deep Learning: A Survey*. DOI: [10.48550/ARXIV.2201.05867](https://doi.org/10.48550/ARXIV.2201.05867). URL: <https://arxiv.org/abs/2201.05867>.
- Lamagna, Walter Marcelo (2016). «Lectura artificial de números manuscritos en datos abiertos de elecciones legislativas en la Ciudad de Buenos Aires». Tesis doct. Universidad de Buenos Aires. Facultad de Ciencias Exactas y Naturales.
- LeCun, Yann et al. (1998). «Gradient-based learning applied to document recognition». En: *Proceedings of the IEEE* 86.11, págs. 2278-2324.
- Paszke, Adam et al. (2019). «PyTorch: An Imperative Style, High-Performance Deep Learning Library». En: *Advances in Neural Information Processing Systems* 32. Ed. por H. Wallach et al. Curran Associates, Inc., págs. 8024-8035. URL: <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>.

- Quinonero-Candela, Joaquin et al. (2008). *Dataset shift in machine learning*. Mit Press.
- Sugiyama, Masashi et al. (2007). «Direct importance estimation with model selection and its application to covariate shift adaptation». En: *Advances in neural information processing systems 20*.
- Sutskever, Ilya et al. (2013). «On the importance of initialization and momentum in deep learning». En: *Proceedings of the 30th International Conference on Machine Learning*. Ed. por Sanjoy Dasgupta y David McAllester. Vol. 28. Proceedings of Machine Learning Research 3. Atlanta, Georgia, USA: PMLR, págs. 1139-1147. URL: <https://proceedings.mlr.press/v28/sutskever13.html>.