

## Tesis de Maestría

# Lectura artificial de números manuscritos en datos abiertos de elecciones legislativas en la Ciudad de Buenos Aires

Lamagna, Walter Marcelo

2016-11-24

Este documento forma parte de la colección de tesis doctorales y de maestría de la Biblioteca Central Dr. Luis Federico Leloir, disponible en [digital.bl.fcen.uba.ar](http://digital.bl.fcen.uba.ar). Su utilización debe ser acompañada por la cita bibliográfica con reconocimiento de la fuente.

This document is part of the doctoral theses collection of the Central Library Dr. Luis Federico Leloir, available in [digital.bl.fcen.uba.ar](http://digital.bl.fcen.uba.ar). It should be used accompanied by the corresponding citation acknowledging the source.

Cita tipo APA:

Lamagna, Walter Marcelo. (2016-11-24). Lectura artificial de números manuscritos en datos abiertos de elecciones legislativas en la Ciudad de Buenos Aires. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.

Cita tipo Chicago:

Lamagna, Walter Marcelo. "Lectura artificial de números manuscritos en datos abiertos de elecciones legislativas en la Ciudad de Buenos Aires". Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. 2016-11-24.

**EXACTAS** UBA

Facultad de Ciencias Exactas y Naturales



**UBA**

Universidad de Buenos Aires



UNIVERSIDAD DE BUENOS AIRES  
FACULTAD DE CIENCIAS EXACTAS Y NATURALES  
DEPARTAMENTO DE COMPUTACIÓN

# Lectura artificial de números manuscritos en datos abiertos de elecciones legislativas en la Ciudad de Buenos Aires

Tesis presentada para optar al título de  
Magíster en Explotación de Datos y Descubrimiento de Conocimiento

Ing. Walter Marcelo Lamagna

Directora de Tesis: Dra. Maria Elena Buemi

Buenos Aires, 2016

## RESUMEN

En esta última década, el acceso masivo a las nuevas tecnológicas con medios de almacenamiento reducidos en tamaño y precios accesibles han impulsado la creación y manipulación de grandes volúmenes de datos. En este trabajo se utilizan datos públicos, abiertos, libres de origen gubernamental.

Existen ciertos datos, que como resultado de la gestión de un país y puestos a disposición de la sociedad, significan un enorme beneficio. Los datos abiertos fortalecen la participación democrática motivando a los científicos de datos a utilizar su conocimiento para responder ciertas preguntas ó resolver problemas a través de la tecnología.

Nuestro estudio consistió en tomar los Datos Abiertos Electorales que corresponden a las elecciones de diputados y senadores en Argentina, Buenos Aires, Capital Federal el 27 de Octubre de 2013 y nos propusimos comparar un mismo dato proveniente de dos fuentes diferentes de formato distinto: dígitos tipeados a partir de formularios en papel e imágenes escaneadas. Las planillas manuscritas escaneadas utilizadas en este trabajo albergan datos numéricos, que son datos certeros, ya que son producto del conteo manual supervisado por todos los actores presentes en el evento. Conociendo la metodología utilizada sabemos que este dato debería corresponderse entre ambas fuentes. Se busca determinar si existe una tasa de error entre dos orígenes de datos que deberían coincidir en sus valores.

Se describe el preprocesamiento de las planillas-imágenes para detectar las regiones de interés que localizan cada número que luego será segmentado en sus correspondientes dígitos. La dificultad de decodificar el símbolo numérico que corresponde a cada número manuscrito se encuentra en la necesidad de utilizar técnicas de procesamiento de imágenes y aprendizaje automático.

En esta tesis se evaluó la aplicación de un método de lectura artificial automatizado que clasifica los dígitos manuscritos de los telegramas y luego los compara con

los datos tipeados manualmente buscando inconsistencias.

En diferentes etapas de este trabajo se aplica minería de datos: se utilizan árboles de decisión en la clasificación de las regiones del telegrama y redes neuronales convolucionales en la clasificación de los dígitos manuscritos.

**Palabras claves:** Elecciones, Datamining, Gobierno Abierto, Procesamiento de Imagenes, Redes Neuronales Convolucionales

## AGRADECIMIENTOS

Quiero expresar mi agradecimiento a la Directora de la Maestría, la Dra. Ana Silvia Haedo por permitirme participar y trabajar cómodamente en un ambiente de excelencia humana y profesional.

A mi Directora de Tesis, la Dra. Maria Elena Buemi por brindarme su experiencia y guiarme en cómo comunicar y organizar esta investigación.

A mi esposa Viviana y mis dos hijas: Aileen y Giulia, las genias que me acompañaron embarcadas en este proyecto, a mi lado en todo momento.

# Índice general

<b>1. Introducción</b>	1
1.1. Datos electorales en Argentina	3
<b>2. Antecedentes en el reconocimiento de dígitos</b>	5
2.1. Estado del arte	5
<b>3. Datos abiertos</b>	10
3.1. Base de datos MNIST: Datos abiertos de libre acceso	12
3.2. Datos de Gobierno	13
3.3. Datos abiertos de Gobierno	13
3.4. Datos Electorales abiertos de Gobierno	16
3.5. Obtención de datos abiertos	18
<b>4. Extracción de dígitos a partir de una imagen</b>	21
4.1. Descripción del telegrama	21
4.2. Detección de regiones de interés	22
4.3. Detección de líneas verticales	22
4.4. Detección de líneas horizontales	23
4.5. Clasificación del contenido de los casilleros mediante un árbol de decisión	26
4.6. Segmentación de los dígitos	29
4.7. Base de datos con dígitos manuscritos de telegramas Argentinos	29
4.8. Clasificación de dígitos basado en Redes Neuronales	32
4.8.1. Red Neuronal Convolutacional	33
<b>5. Reconociendo dígitos a partir de telegramas electorales</b>	36
5.1. Preparación de los datos y entrenamiento	36
5.1.1. Formato de los datos	38

---

5.1.2.	SET1: Set de datos sin distorsión . . . . .	39
5.1.3.	SET2: Set de datos modificando alto y ancho . . . . .	39
5.1.4.	SET3: Set de datos con distorsión senoide . . . . .	39
5.2.	Características de la red . . . . .	39
5.2.1.	Inicialización de parámetros . . . . .	41
5.3.	Resultados obtenidos . . . . .	42
5.3.1.	Visualización de la primera capa . . . . .	44
5.3.2.	Errores de transcripción . . . . .	45
<b>6.</b>	<b>Conclusiones y trabajos futuros . . . . .</b>	<b>49</b>
	<b>Apéndice . . . . .</b>	<b>53</b>
<b>A.</b>	<b>Telegramas con inconsistencias . . . . .</b>	<b>54</b>
<b>B.</b>	<b>Código fuente ConvNet . . . . .</b>	<b>69</b>

# 1. INTRODUCCIÓN

En esta última década, el acceso masivo a las nuevas tecnológicas con medios de almacenamiento reducidos en tamaño y precios accesibles han impulsado la creación y manipulación de grandes volúmenes de datos. En la ciencia, en los negocios y en la gestión pública entre otros, se producen datos que pueden representar una ventaja competitiva para aquellos que tengan el conocimiento y las herramientas necesarias. Los datos son la materia prima que junto con las herramientas adecuadas permiten a un científico de datos responder preguntas ó resolver problemas.

En el año 2010, Tim-Berners-Lee (conocido como el padre de la Web) hizo una presentación llamada “El año en que los Datos Abiertos se difundieron” <sup>1</sup> donde muestra ejemplos de impacto social utilizando datos abiertos. Se menciona el caso de la ciudad de Zanesville (Ohio, Estados Unidos) [1] que en el 2008, utilizaron datos abiertos para demostrar una arbitrariedad racial en la distribución del agua corriente en el pueblo. Otra iniciativa de datos abiertos con impacto social que forma parte de la presentación de Tim-Berners-Lee es un proyecto que permite observar en qué se utiliza el dinero de los impuestos en Inglaterra <sup>2</sup>.

Los datos abiertos pueden ser utilizados por la sociedad para ejercer sus derechos frente a situaciones que considere injustas.

La decisión de implementar una estrategia de *Datos Abiertos* es impulsada por los gobiernos que buscan una ideología de transparencia y consideran que la participación de la ciudadanía puede aportar algún valor y/o buscan la eficiencia de gestión a partir de compartir ciertos datos. En el 2010, Davies [2] realiza un estudio sobre quién utiliza datos abiertos de gobierno y con qué finalidad en el Reino Unido. Hunnius et al. 2013 [3] investiga el estado de situación de Datos Abiertos de Gobierno (información del sector público) en Alemania y España. Nahon et al. (2015)

---

<sup>1</sup> [http://www.ted.com/talks/lang/eng/tim\\_berners\\_lee\\_the\\_year\\_open\\_data\\_went\\_worldwide.html](http://www.ted.com/talks/lang/eng/tim_berners_lee_the_year_open_data_went_worldwide.html)

<sup>2</sup> <http://app.wheredoesmymoneygo.org>



[4] publica un trabajo donde analiza la evolución de las políticas de gobierno abierto en 16 ciudades de Estados Unidos. Yang et al. (2015) [5] analiza desde un punto de vista legislativo y político el impacto de la iniciativa de Gobierno Abierto y Datos Abiertos en Taiwan. En el 2015, Zuiderwijk et al. [7] describe una infraestructura de datos abiertos donde se puede compartir datos e información de gestión de un gobierno que aporten a la transparencia y que diferentes grupos de ciudadanos puedan utilizarla anónimamente en su toma de decisiones.

Por su parte, la fundación del conocimiento libre <sup>3</sup> describe características que deben cumplir los Datos Abiertos <sup>4</sup>, entre ellas *Acceso y disponibilidad, Reutilización y redistribución y Participación universal*.

En Software no es lo mismo un Software Libre <sup>5</sup> que un Software Abierto (Open Source) [8], aunque ambos términos se refieren a un mismo concepto: software disponible bajo ciertas licencias que garantiza cierto conjunto específico de libertades. El movimiento del *Software Libre* nace en 1983, encabezado por Richard Stallman. Según la definición establecida por Richard Stallman, un software es “libre” cuando garantiza cuatro libertades: utilizar, estudiar, distribuir y mejorar el programa. La libertad de estudiar y mejorar un programa requieren acceso al código fuente.

Parte de la comunidad de programadores no se identificaba con la filosofía del software libre por que la palabra *libre* en inglés puede significar tanto libre como gratis. Esto impulsó a que en 1988, un grupo de individuos comience a trabajar en la filosofía del *Software Abierto*. En Software Abierto se puede optar por alguna de las licencias disponibles o crear una licencia nueva. Una licencia de Software Abierto debe compatibilizar con los criterios aceptados por la comunidad detallados en el sitio web Open Source Definition [8].

Los conceptos del Software Abierto pueden aplicarse al mundo de los datos online, entendiéndose que también podrían ser abiertos aunque no de libre utilización (por ejemplo un periódico Online ó una red social <sup>6</sup>). Es similar al software libre, donde

---

<sup>3</sup> <https://okfn.org/>

<sup>4</sup> <http://opendefinition.org/od/2.0/en/index.html>

<sup>5</sup> <https://www.fsf.org/es>

<sup>6</sup> [https://www.facebook.com/apps/site\\_scraping\\_tos\\_terms.php?hc\\_location=ufi](https://www.facebook.com/apps/site_scraping_tos_terms.php?hc_location=ufi)

---

libre no significa “gratis” y que la libertad se extiende hasta donde decide el o los autores originales del trabajo.

## 1.1. Datos electorales en Argentina

A partir de la década del 80 comienza un lento proceso de democratización en América latina: en 1970 el gobierno militar ecuatoriano concedió elecciones, el mismo año la revolución encabezada por el sandinismo derrocó al dictador nicaragüense Anastasio Somoza. En 1980, se reestableció el poder civil en Perú, Honduras lo hace en 1981 y Argentina en 1983. Uruguay y Brasil lo hicieron en 1985. En Argentina, en 1983 vuelve a tener vigencia la ley Sáenz Peña, sancionada por el Congreso de la Nación Argentina el 10 de febrero de 1912 (voto universal, secreto y obligatorio para los ciudadanos argentinos varones, nativos o naturalizados, mayores de 18 años de edad habitantes de la Nación inscriptos en el padrón electoral) y la ley 13.010 del 9 de Septiembre de 1947 <sup>7</sup> que reconoce la igualdad en derechos y obligaciones políticas para las mujeres en la República Argentina.

En Argentina, a partir del año 2001 hay datos abiertos de resultados electorales disponibles en el sitio Web de la Dirección Nacional Electoral <sup>8</sup>, lo cuál coincide con la expansión masiva de dispositivos tecnológicos como Redes informáticas, Scanners, Teléfonos inteligentes que producen contenido disponible a pocos segundos de finalizado un acto electoral.

Este estudio consistió en tomar los Datos Abiertos Electorales que corresponden a las elecciones de diputados y senadores en Argentina, Buenos Aires, Capital Federal el 27 de Octubre de 2013. Utilizamos contenido digital consistente en: planillas de cálculo, archivos de imágenes y formatos estandarizados. El contenido de estos archivos, datos numéricos en planillas y formularios, son de vital importancia en el proceso electoral ya que conforman los resultados que obtuvo cada partido político ó candidato de nuestro país en un determinado lugar. Comparamos un mismo dato

---

<sup>7</sup> <http://servicios.infoleg.gob.ar/infolegInternet/anexos/45000-49999/47353/norma.htm>

<sup>8</sup> [http://elecciones.gov.ar/articulo\\_sub\\_sub.php?secc=2&sub\\_secc=9&sub\\_sub\\_secc=9](http://elecciones.gov.ar/articulo_sub_sub.php?secc=2&sub_secc=9&sub_sub_secc=9)

---

proveniente de dos fuentes diferentes de formato distinto: dígitos tipeados a partir de formularios en papel e imágenes escaneadas. Dada la metodología utilizada este dato debería coincidir en ambas fuentes. Para determinar si existen inconsistencias entre dos orígenes de datos diferentes y de formato distinto que constituyen el mismo dato utilizamos técnicas de procesamiento de imágenes y minería de datos. Estos dos orígenes de datos diferentes son datos tipeados a partir de formularios en papel y números manuscritos en telegramas. Debido a que las magnitudes contenidas en estas planillas son muy relevantes, nos hemos centrado en la detección de estos números reflejados. En particular, parte de esta información tiene la característica de ser números manuscritos y deben compararse con otro set que contiene números digitados por computadora. Esta tarea involucra la implementación de un reconocedor de dígitos manuscritos basado en técnicas de aprendizaje automático con entrenamiento supervisado como lo son las **redes neuronales**. Esto nos llevó a involucrarnos en temas de **procesamiento de imágenes**, **deep learning** y automatización de la obtención de datos.

Esta tesis está organizada de la siguiente manera: en el capítulo 2 se describe el estado del arte; en el capítulo 3 se introduce al lector en los conceptos de datos abiertos, datos de Gobierno y software de libre acceso; en el capítulo 4 se expone el procesamiento de los datos utilizados y sus procedimientos. El capítulo 5 expone características de los datos utilizados para entrenar el modelo de red neuronal y qué experimentos se realizaron para clasificar los dígitos manuscritos. Finalmente, en el capítulo 6 se exponen las conclusiones y se proponen trabajos futuros. Finalmente se muestran las referencias utilizadas.

## 2. ANTECEDENTES EN EL RECONOCIMIENTO DE DÍGITOS

El reconocimiento de dígitos manuscritos es un tema que resulta ser de interés en el área de Reconocimiento de Patrones por sus potenciales aplicaciones prácticas. Resuelve la necesidad de automatizar el tratamiento de documentos impresos en el Gobierno, las empresas y una sociedad en expansión permanente. En 1992, Fujisawa et al. [9] aplican análisis de documentos escaneados, segmentación y reconocimiento de dígitos manuscritos. Srihari 1993 [10] aplica reconocimiento de dígitos en clasificación de correo postal. En el 2014, Tom y Davis [11] hacen reconocimiento de números en patentes de vehículos utilizando la base de datos Caltech 1999 <sup>1</sup> con patentes de autos en Estados Unidos.

### 2.1. Estado del arte

La visión por computadora ó visión artificial es de utilidad en el reconocimiento óptico de caracteres (OCR) que se aplica tanto en la lectura de dígitos manuscritos como impresos por máquina. Las técnicas de lectura artificial se utilizan en la transformación de imágenes de caracteres en representaciones simbólicas.

La primera investigación conocida de lectura artificial se remonta a 1929 cuando Gustav Tauscheck [12] obtuvo una patente de su “Máquina que lee”. En 1933, Paul W. Handel [13] patenta una “Máquina Estadista” que realiza sucesivas comparaciones entre un caracter y la imagen de este mismo caracter, lo describe como un aparato foto-eléctrico utilizado para responder a ciertas coincidencias entre un caracter y una imagen.

En 1958, los investigadores Grimsdale, Summer, Tunis y Kilburn [14] crearon uno de los primeros sistemas que reconocen caracteres, consistía en un scanner que

---

<sup>1</sup> <http://vision.caltech.edu/archive.html>

recorría una imagen y la convierte en una matriz bidimensional de puntos, de este patrón se extraen detalles de longitud, curvatura e inclinación. En 1961, IBM produce uno de los primeros dispositivos de reconocimiento de caracteres (OCR) que podían escanear 420 documentos por minuto utilizando un algoritmo personalizado para máquinas de contabilidad, y en 1962 el IBM 1428 que tenía la capacidad de leer todos los caracteres alfanuméricos [15].

En 1966, Casey y Nagy [16] en el centro de investigaciones de IBM Watson presentaron un reconocedor de caracteres chinos. Tou y Gonzalez [17] publican en 1972 un trabajo acerca del reconocimiento de caracteres manuscritos aplicando conceptos de extracción de propiedades. En 1977, Farhat y Pavlidis [20] crean un reconocedor de dígitos manuscritos que convierte los dígitos en representaciones de alto nivel como trazos, arcos y esquinas. Durante la década del 80 continúan desarrollándose las técnicas que pasan desde reconocer un conjunto limitado de caracteres impresos por computadoras, hasta técnicas sofisticadas que reconocen una variedad de caracteres complejos manuscritos de textos completos como se describe en un trabajo de Simon Kahan [21] (1987) acerca de el reconocimiento de caracteres impresos en cualquier tipografía y tamaño. En 1984 Shridhar y Badreldin [22] desarrollan un método para reconocer dígitos manuscritos basado en extracción de características y árboles de decisión. En 1988, Rajavelu [23] hace reconocimiento de caracteres utilizando redes neuronales. Las redes neuronales Perceptrón de Múltiples Capas (MLP) son comúnmente utilizadas en el reconocimiento de caracteres. En una investigación de 1992, Michael Sabourin [24] publicó un sistema de reconocimiento de caracteres utilizando redes neuronales y se menciona que estas redes tienen la ventaja de particionar de manera más eficiente el espacio de propiedades, en comparación con un clasificador de vecinos más cercanos. Los reconocedores de caracteres se utilizan para diferentes idiomas, Kimura et al. 1996 [25] presenta un reconocedor de caracteres Japoneses.

En 1962, Hubel & Wiesel [28] estudian la arquitectura y el funcionamiento de el cortex visual de los felinos, en su publicación describen que en el cortex visual de estos mamíferos (Fig. 2.1) existen dos tipos de células a las que llamaron células

C (Complejas) y S (Simples). En 1980 Fukushima & Miyake [27] presentan una red neuronal inspirada en la arquitectura del cortex visual de los felinos descrita por Hubel & Wiesel [28], la llamaron Neocognitron. En 1983, Fukushima et al. [29] mencionan en su trabajo que las funciones de Convolución/Submuestreo de la Red Neuronal Neocognitron están inspiradas en el funcionamiento de las células C y S descubiertas por Hubel y Wiesel y mencionan las ventajas que tiene este tipo de redes reconociendo patrones visuales. Las redes neuronales Neocognitron tienen mapas de características y pesos compartidos seguidos por una capa de submuestreo ó promedio.

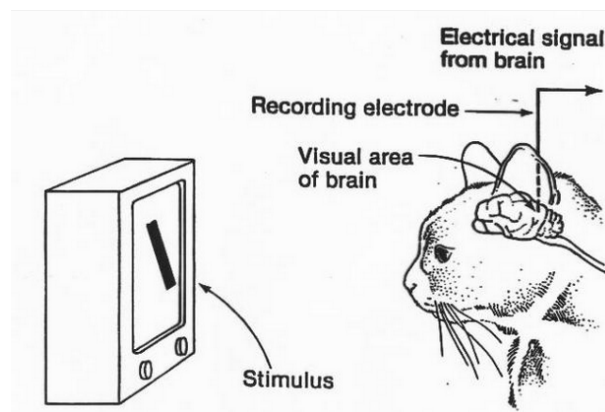


Fig. 2.1: Experimento de Hubel y Wiesel

En 1990 Yan Lecun et al. [30] utiliza la red Neocognitron con el algoritmo Back-propagation de aprendizaje supervisado. Fukushima y Wake (1991) [31] utiliza Neocognitron en reconocimiento de caracteres alfanuméricos. En 1992 Fukushima [32] hace reconocimiento de caracteres manuscritos utilizando la red Neocognitron. El término Red Convolutional es mencionado por primera vez en una publicación de Yan LeCun et al. [33] (1995) donde utiliza Redes Neuronales Convolucionales en reconocimiento dígitos manuscritos de el conjunto de datos MNIST [35] y obtiene un 99,3% de acierto, compara los resultados con otros métodos de aprendizaje:

A	B	C	D	E	F
8.4 %	2.4 %	1.6 %	1.7 %	1.1 %	0.9 %

Tab. 2.1: Error de clasificación obtenido en 1995, Yan LeCun et al. [33]

- A) Clasificador Lineal: 8.4 %.
- B) Clasificador por vecino más cercano ( $k=3$ ): 2.4 %.
- C) Red neuronal multicapa totalmente conectada: 1.6 %.
- D,E,F) Arquitecturas de ConvNet LeNet-1, LeNet-4 , LeNet-5.

La figura 2.2 presenta una línea de tiempo y algunos trabajos publicados acerca de las redes neuronales convolucionales.

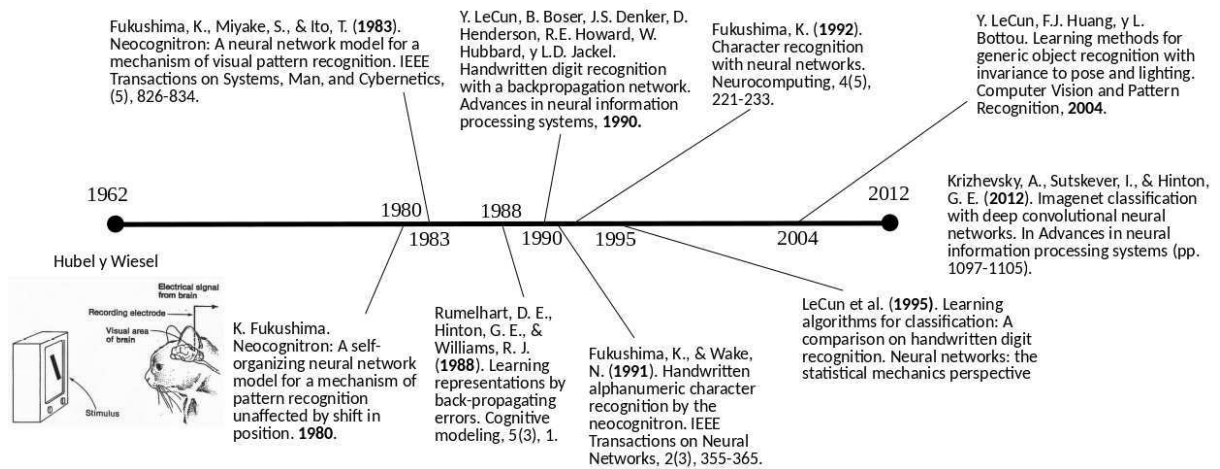


Fig. 2.2: Reseña histórica de las redes neuronales convolucionales

La Red Neuronal Convolutiva (ConvNet) es un método de aprendizaje automático considerado de Aprendizaje Profundo. El concepto de profundidad se refiere al camino más largo desde su nodo de entrada hasta un nodo de salida. En

---

las ConvNet cada capa posee un número diferente de neuronas, estando cada neurona relacionada con una región de neuronas de la capa anterior al compartir pesos sinápticos. Los pesos a su vez son compartidos en diferentes lugares de la misma capa. Estas redes están especialmente diseñadas para reconocer figuras con un alto grado de invariancia de translación, cambio de escala, sesgado de la imagen y otras formas de distorsión. Durante el entrenamiento, la Red Neuronal Convolutiva aprende de ejemplos clasificados lo cual hace al aprendizaje supervisado.

En el 2004, LeCun et al. [36] utilizan una Red Neuronal Convolutiva para reconocer un mismo objeto en diferentes posiciones y variada iluminación. Ahn et al. (2010) [37] utilizan una ConvNet entrenada con SGD y obtiene un 99,17% de acierto sobre un conjunto de datos con imágenes de caracteres Farsi/Árabicos.

Algunas investigaciones acerca de ConvNet utilizan un conjunto de datos llamado Imagenet, esta fue creada por el laboratorio de visión de la Universidad de Stanford<sup>2</sup>. Es una base de datos compuesta por más de 100.000 categorías y su tamaño es de aproximadamente 1.2TB. Imagenet se utiliza en la competencia LSVRC (Large Scale Visual Recognition Challenge) [38] de minería de datos y reconocimiento de imágenes, gana quién logre clasificar correctamente la mayor cantidad de imágenes en categorías dadas. En 2012, Krizhevsky et al. [39] obtuvieron los mejores resultados en la competencia ImageNet LSVRC utilizando redes neuronales convolucionales. Las ConvNets también son eficientes en el abordaje de problemas tales como detección de personas, en una publicación del año 2013, Sermanet et al. [40] obtuvieron muy buenos resultados utilizando ConvNets en detección de peatones sobre los conjuntos de datos Caltech [41] e Inria [42].

---

<sup>2</sup> <http://www.image-net.org/>



### 3. DATOS ABIERTOS

Pese a que ciertos datos sean de fácil acceso y exista la posibilidad técnica de almacenarlos para luego hacer con ellos lo que uno quiera no significa que sea legalmente viable. La legislación ha comenzado a incluir el término Dato Abierto ó Dato Público de manera que los ciudadanos puedan utilizar estos datos y tengan la posibilidad de incluir un valor agregado.

Naser et al. (2012) [43] describe Datos Abiertos como un proceso en el que se pone a disposición de la sociedad, de manera libre, una gran cantidad de datos procedentes de diferentes organizaciones, fundamentalmente del ámbito de la Administración Pública o de aquellos proyectos que han sido financiados con dinero público o creados por una institución pública.

En Software no es lo mismo un Software Libre <sup>1</sup> que un Software Abierto (Open Source) <sup>2</sup>, aunque ambos términos se refieren a un mismo concepto: software disponible bajo ciertas licencias que garantiza cierto conjunto específico de libertades. El movimiento del *Software Libre* nace en 1983, encabezado por Richard Stallman. El Software Libre se refiere a qué libertad decide el autor que tenga su programa. Según la definición establecida por Richard Stallman, un software es “libre” cuando garantiza cuatro libertades: utilizar, estudiar, distribuir y mejorar el programa. La libertad de estudiar y mejorar un programa requieren acceso al código fuente. Parte de la comunidad de programadores no estaba representada por la filosofía del software libre por que la palabra *libre* tiene un significado ambiguo y puede significar tanto libre como gratis. Esto impulsó a que en 1988, un grupo de individuos comience a trabajar en la filosofía del *Software Abierto* (Software de código abierto). El Software Abierto también tiene licencias que debe permitir la libre utilización del software <sup>3</sup>, su modificación y distribución.

---

<sup>1</sup> <https://www.fsf.org/es>

<sup>2</sup> <https://opensource.org/>

<sup>3</sup> <https://opensource.org/osd>

Los conceptos del Software Abierto pueden aplicarse al mundo de los datos online, entendiéndose que también podrían ser abiertos aunque no de libre utilización (por ejemplo un periódico Online ó una red social <sup>4</sup>). Es similar al software libre, donde libre no significa “gratis” y que la libertad se extiende hasta donde decide el o los autores originales del trabajo.

Por su parte, la fundación del conocimiento libre <sup>5</sup> describen características que deben cumplir los Datos Abiertos <sup>6</sup>, entre ellas:

1. Acceso y disponibilidad: Los datos deben estar disponibles por completo, si existiese un costo asociado, este debe ser razonable, preferentemente accesible sin costo por Internet. Deben estar disponibles en un formato conveniente y modificable.
2. Reutilización y redistribución: Los datos deben permitir su re-distribución y reutilización, incluyendo su comercialización ya sea individualmente o con entrecruzamiento y/o entremezclado con otras bases de datos.
3. Participación universal: La licencia debe permitir que todos tengan la posibilidad de utilizar, reutilizar y redistribuir, sin ningún tipo de discriminación a personas o grupos.

Como puede observarse en el tercer punto se habla de licencias. Toda base de datos abierta debería estar acompañada de su correspondiente licencia, la misma puede solicitar una mención a quienes generaron los datos ó reservarse el derecho de que cualquier trabajo derivado mantenga la misma licencia, etc. Es similar al software libre, donde libre no significa “gratis” y que la libertad se extiende hasta donde decide el o los autores originales del trabajo. En un sitio web acerca de Datos Abiertos Electorales <sup>7</sup> se menciona una característica complementaria que deben

---

<sup>4</sup> [https://www.facebook.com/apps/site\\_scraping\\_tos\\_terms.php?hc\\_location=ufi](https://www.facebook.com/apps/site_scraping_tos_terms.php?hc_location=ufi)

<sup>5</sup> <https://okfn.org/>

<sup>6</sup> <http://opendefinition.org/od/2.0/en/index.html>

<sup>7</sup> <http://openelectiondata.net/es/guide/principles/non-discriminatory/>

---

cumplir los datos para ser Abiertos, se trata de que el acceso a los mismos sea anónimo, esto elimina la posibilidad de que se tome algún tipo de acción que perjudique la libertad de expresión y acceso a la información. Diferentes definiciones acerca de Datos Abiertos coinciden en conceptos como la disponibilidad, acceso, participación y reutilización, esto manifiesta que en materia de datos existe un responsable que puede decidir qué hacer.

En la ciencia, en los negocios y en la gestión pública entre otros, se producen datos que pueden representar una ventaja competitiva para aquellos que tengan el conocimiento y las herramientas necesarias. En el año 2010, Tim-Berners-Lee (conocido como el padre de la Web) hizo una presentación llamada “El año en que los Datos Abiertos se difundieron”<sup>8</sup> donde muestra ejemplos de impacto social utilizando datos abiertos. En esta presentación se cita el ejemplo de la ciudad de Zanesville (Ohio, Estados Unidos) [1] que en el 2008, utilizaron datos abiertos para demostrar una arbitrariedad en la distribución del agua corriente en el pueblo donde se privilegiaba a las familias de raza blanca. Los datos abiertos pueden ser utilizados por la sociedad para ejercer sus derechos frente a situaciones que considere injustas. Otra iniciativa de datos abiertos con impacto social que forma parte de la presentación de Tim-Berners-Lee es un proyecto que permite observar en qué se utiliza el dinero de los impuestos en Inglaterra<sup>9</sup>.

Todavía no están claras las consecuencias de que ciudadanos empoderados participen e interactúen en las decisiones de un país, pero indudablemente es una tendencia global en expansión.

### 3.1. Base de datos MNIST: Datos abiertos de libre acceso

La base de datos MNIST [35] constituye un estándar a la hora de testear y comparar métodos de reconocimiento de dígitos manuscritos. La misma contiene 70000 imágenes de números escritos a mano que incluyen diferentes estilos de escritura,

---

<sup>8</sup> [http://www.ted.com/talks/lang/eng/tim\\_berners\\_lee\\_the\\_year\\_open\\_data\\_went\\_worldwide.html](http://www.ted.com/talks/lang/eng/tim_berners_lee_the_year_open_data_went_worldwide.html)

<sup>9</sup> <http://app.wheredoesmymoneygo.org>

---

extraídos de un conjunto mucho mayor de patrones manuscritos provisto por NIST, *National Institute of Standards and Technology* [Seijas et al, 2003 [6]].

MNIST tiene derechos de autor de Yann LeCun <http://yann.lecun.com/> y Corinna Cortes, bajo la licencia **Creative Commons Attribution-Share Alike 3.0**, estos son datos abiertos si se menciona a sus autores en trabajos derivados.

### 3.2. Datos de Gobierno

Datos de Gobierno son aquellos producidos en la gestión de un país, se utilizan en la toma de decisiones estratégicas y están relacionados con políticas educativas, sanitarias, judiciales, datos obtenidos por organismos científicos entre otros. Durante la gestión gubernamental se hizo necesario disponer de infraestructura adecuada al volumen, intercambio y preparación de la disponibilidad, así como el requisito que sean especialistas en datos quienes tengan la responsabilidad de darles acceso libre. El acceso a los datos puede significar un enorme beneficio para la sociedad sin obstaculizar la gobernabilidad de un país.

### 3.3. Datos abiertos de Gobierno

Los Datos Abiertos de Gobierno son de utilidad para la ciudadanía aunque también para el Gobierno que puede rendir cuentas durante su gestión fomentando la transparencia <sup>10</sup>.

La decisión de implementar una estrategia de *Datos Abiertos* es impulsada por los gobiernos que buscan una ideología de transparencia y consideran que la participación de la ciudadanía puede aportar algún valor y/o buscan la eficiencia de gestión a partir de compartir ciertos datos. En el 2010, Davies [2] realiza un estudio en el Reino Unido acerca de quién utiliza Datos Abiertos de Gobierno y con qué finalidad. Hunnius et al. 2013 [3] investiga el estado de situación de Datos Abiertos de Gobierno (información del sector público) en Alemania y España. Nahon et al.

---

<sup>10</sup> <http://www.datos.gob.ar/datos-abiertos>

(2015) [4] publica un trabajo donde analiza la evolución de políticas de Gobierno Abierto en 16 ciudades de Estados Unidos. Yang et al. (2015) [5] analiza desde un punto de vista legislativo y político el impacto de la iniciativa de Gobierno Abierto y Datos Abiertos en Taiwan. En el 2015, Zuiderwijk et al. [7] describe una infraestructura de datos abiertos donde se comparten datos e información de gestión de un gobierno que aporten a la transparencia y que diferentes grupos de ciudadanos puedan utilizarla anónimamente en su toma de decisiones.

La Comisión de la Unión Europea <sup>11</sup> utiliza el término Información del Sector público en referencia a los Datos Abiertos de Gobierno y los define como “aquella información producida, recolectada o por la que se paga en los organismos de gestión pública”. En 2011, Geiger y Lucke [44] describe los datos abiertos como “todos los datos almacenados correspondientes al sector público que puedan ser puestos a disposición por parte del gobierno por ser de interés público, sin ningún tipo de restricción para su uso y distribución”, es una definición que contempla la necesidad de reservar ciertos datos por ser confidenciales, privados o incluso posibles secretos industriales estratégicos para un país.

Anneke Zuiderwijk en el 2015 [7] realiza un estudio en Holanda acerca de Datos Abiertos de Gobierno e identifica tres tipos de usuarios:

1. Proveedores de Datos Abiertos de Gobierno.
2. Usuarios de estos Datos Abiertos.
3. Legisladores de Datos Abiertos de Gobierno (Figura 3.1).

Un usuario de estos datos es un Científico de Datos que tiene las herramientas para utilizar los datos públicos y responder preguntas ó resolver problemas realizando estudios ó un periodista de datos redactando una nota periodística <sup>12</sup>; o incluso un ciudadano que desee obtener información acerca de su vecindario por ejemplo.

<sup>11</sup> <https://ec.europa.eu/digital-single-market/en/open-data>

<sup>12</sup> <http://data.lanacion.com.ar/dashboards/4610/indicadores-generales/>

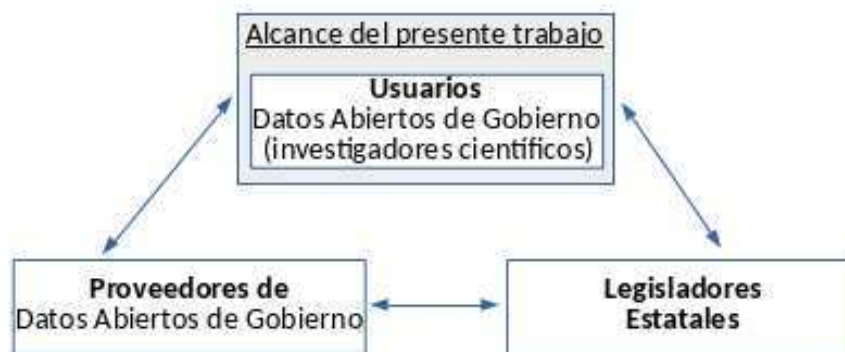


Fig. 3.1: Usuarios de Datos Abiertos de Gobierno

En un trabajo de Tim Davies (2010) [2] acerca de Datos Abiertos y Democracia en el sector público del Reino Unido <sup>13</sup> se encontraron cinco diferentes usos que se le da a los Datos Abiertos de Gobierno:

1. **Del dato al hecho:** extracción de hechos específicos a partir de los datos.
2. **Del dato a la información:** Generar una representación de los datos que permita realizar interpretaciones y reportes.
3. **Del dato al dato:** Extender el conjunto original de datos, combinarlo con otros datos, cambiar su formato ó manipularlo y compartir el conjunto extendido de datos.
4. **Del dato a la interfaz:** Proveer una interfaz que permita acceder y explorar los datos interactivamente.
5. **Del dato al servicio:** Integrar conjuntos de datos y producir productos o servicios.

El Gobierno de la Ciudad de Buenos Aires tiene un sitio web con datos públicos, <sup>14</sup> en el que define los Datos Abiertos de la siguiente manera: “Poner información del Estado en un catálogo al alcance de todos, en formatos digitales, estándar y abiertos,

<sup>13</sup> <http://data.gov.uk>

<sup>14</sup> <http://data.buenosaires.gob.ar/about>

siguiendo una estructura clara que permita su fácil comprensión y reutilización por parte de la ciudadanía.”. Otras regiones de la Argentina como Bahía Blanca <sup>15</sup> y Misiones <sup>16</sup>, tienen su portal de Datos Públicos siendo una herramienta poderosa que inspira el deseo de implementar mejoras que han sido efectivas en otras ciudades ó países.

En Argentina, los Datos abiertos de Gobierno tienen un marco legal que los regula. En 1998, la Legislatura de la Ciudad de Buenos Aires crea la Ley de Acceso a la Información Pública Nro. 104, <sup>17</sup> que regula el derecho de los ciudadanos a solicitar información pública. En el año 2003, el ex-presidente Néstor Kirchner firmó el decreto 1172 <sup>18</sup>, que reglamenta el Acceso a la Información Pública para el Poder Ejecutivo Nacional. En el año 2016, se firmó el Decreto 117/2016 <sup>19</sup> que reglamenta plazos tanto para la publicación de los primeros datasets en el Portal Nacional de Datos Públicos <sup>20</sup> como también para detallar los activos de datos de los distintos organismos de la Nación y su cronograma de publicación.

### 3.4. Datos Electorales abiertos de Gobierno

En Argentina, la gestión de la digitalización de los telegramas electorales y la contabilización de los votos que ofrece un resultado provisional, lo realiza la empresa Indra <sup>21</sup>. El Correo Argentino traslada los telegramas originales en papel hasta uno de los Centros de Gestión de Datos de Indra donde se valida la integridad del documento y se lo digitaliza. Cada telegrama es leído y tipeado por dos personas diferentes, esto reduce la posibilidad de errores en la lectura y transcripción de los números escritos en los telegramas. En aquellos casos en los que no coincide la transcripción del número por parte de ambas personas, el telegrama es enviado

---

<sup>15</sup> <http://bahiablanca.opendata.junar.com/home/>

<sup>16</sup> <http://www.datos.misiones.gov.ar/>

<sup>17</sup> <http://www.buenosaires.gob.ar/asuntospublicos/ley-104-acceso-la-informacion-publica>

<sup>18</sup> <http://www.infoleg.gov.ar/infolegInternet/anexos/90000-94999/90763/norma.htm>

<sup>19</sup> <http://servicios.infoleg.gov.ar/infolegInternet/anexos/255000-259999/257755/norma.htm>

<sup>20</sup> <http://www.datos.gob.ar/>

<sup>21</sup> <http://www.indracompany.com/pais/argentina/df>

a un área llamada Resolución de incidentes, donde se decidirá qué hacer con ese dato. De corresponderse el número transcripto por ambos sujetos, se almacena y es contabilizado. Las actividades mencionadas y el flujo de información se presenta en la Fig. 3.2.

El resultado provisional se anuncia a las pocas horas de concluidas las elecciones; siendo este una aproximación al resultado definitivo que se obtiene varias semanas después. Una vez concluidas las elecciones, los partidos políticos resuelven diferencias en el transcurso de las siguientes semanas, estas no suelen alterar el resultado final de las elecciones.

### Funciones del sistema

- Captura de datos, doble digitación
- Resolución de incidencias
- Escaneo y codificación de telegramas en paralelo a la grabación
- Seguimiento del proceso
- Fiscalización mesa a mesa
- Transmisión de datos
- Monitorización del sistema de carga



Fig. 3.2: Gestión tradicional de los telegramas (Fuente: Dirección Nacional Electoral, Ministerio del Interior y Transporte)

Los datos utilizados en este trabajo son Datos Electorales Abiertos de Gobierno. Tanto los votos transcriptos como los telegramas escaneados son publicados en la



Página Web Oficial del Gobierno Argentino <sup>22</sup>, constituyéndose, a partir de ese momento en datos abiertos de gobierno. El marco legal es la disposición 408/2013, en el Anexo I del Decreto Nr. 682 del 14 de mayo de 2010, que aprobó la directiva de datos públicos abiertos para la administración electoral <sup>23</sup>. Esta directiva establece que las siguientes bases de datos deberán adecuarse progresivamente permitiendo la interacción con los ciudadanos:

- Resultados electorales
- Asignación de espacios para Campañas Electorales en Servicios de Comunicación Audiovisual
- Aportes partidarios y de campañas electorales
- Cartografía electoral digital

### 3.5. Obtención de datos abiertos

Los Datos Abiertos de un Gobierno suelen encontrarse en diferentes sitios de Internet y el investigador debe obtener los datos para combinarlos localmente. Los formatos de los datos abiertos, su acceso y disponibilidad es una de las barreras que actualmente el científico de datos debe sortear si desea hacer una investigación. Cuando se trata de Datos Abiertos de Gobierno cuyo dimensionamiento supera el Gigabyte, su obtención se dificulta y puede llevar desde días a meses de trabajo. El Manual de Periodismo de Datos <sup>24</sup> describe algunas herramientas útiles que automatizan la obtención de datos abiertos <sup>25</sup>. En el 2001, Mike Thelwal [45] propone el diseño de un sistema distribuido que ayude a los científicos de datos en el preprocesamiento de información obtenida de páginas web.

Los telegramas obtenidos para el presente trabajo son datos que no cumplen con la disponibilidad ni accesibilidad: están disponibles por pocos días concluidas

<sup>22</sup> <http://www.resultados.gob.ar/inicio.htm>

<sup>23</sup> <http://www.infoleg.gob.ar/infolegInternet/anexos/220000-224999/221756/norma.htm>

<sup>24</sup> [http://datajournalismhandbook.org/1.0/en/getting\\_data\\_3.html](http://datajournalismhandbook.org/1.0/en/getting_data_3.html)

<sup>25</sup> <http://scrapy.org/>

las elecciones y dispersos en miles de archivos a los que se debe acceder de a uno por vez. Se automatizó la tarea con un programa desarrollado en lenguaje de programación Bash <sup>26</sup>. La obtención de todos los telegramas llevó 18 horas, de haber interrupciones en la transferencia de datos el programa debe ser personalizado para que obtenga solamente los archivos restantes.

Siendo el volumen disponible de datos decenas de miles de telegramas completados manualmente y teniendo a disposición el resultado de un tipeo manual de los mismos, resulta interesante combinar técnicas de procesamiento de imágenes y minería de datos para una lectura automatizada de los telegramas y lograr comparar ambos resultados (Fig. 3.3).

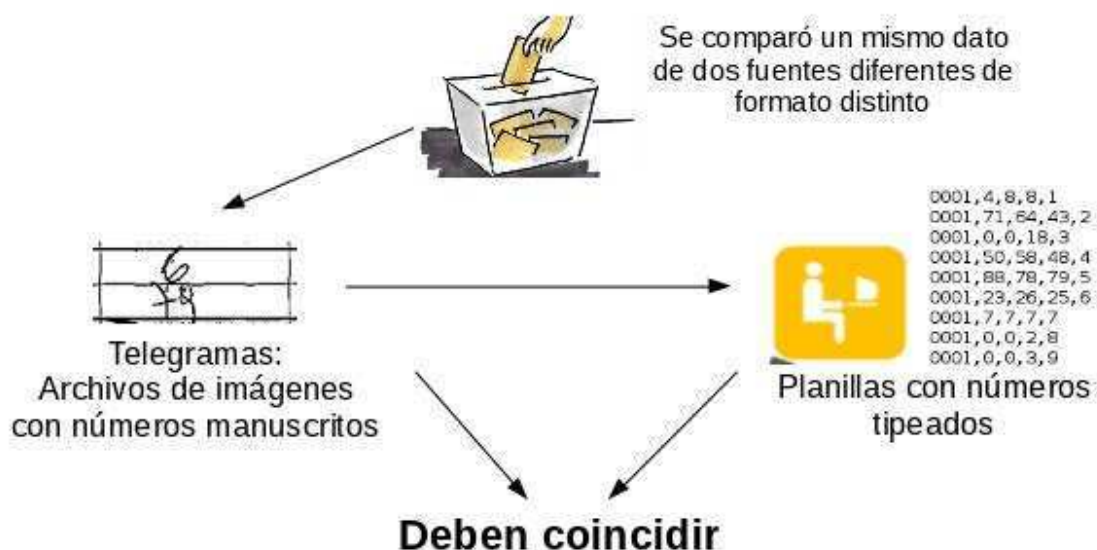


Fig. 3.3: Comparamos dos orígenes de datos diferentes y de formato distinto que constituyen el mismo dato

En esta tesis se desarrolló un método de lectura artificial automatizado que compara los números manuscritos con los datos transcritos considerados válidos y certeros. En este trabajo se evaluó la aplicación de lectura artificial sobre los dígitos manuscritos de los telegramas.

<sup>26</sup> <https://www.gnu.org/software/bash/>

Se utilizaron tres tipos de datos:

- Dígitos manuscritos que provienen de una imagen.
- El número en texto plano, resultado del clasificador sobre la imagen.
- El número en texto plano, tipeado por una persona que lee el telegrama.

La minería de datos es aplicada en diferentes etapas de este trabajo: se utilizó un árbol de decisión en la clasificación de las regiones del telegrama y una red neuronal convolucional en la clasificación de los dígitos manuscritos.

## 4. EXTRACCIÓN DE DÍGITOS A PARTIR DE UNA IMAGEN

Aquí se describe el tratamiento de los datos provenientes de las planillas manuscritas escaneadas que son llamados telegramas. Estas planillas-imágenes albergan datos numéricos, que son datos certeros, ya que son producto del conteo manual supervisado por todos los actores presentes en el evento. Este capítulo trata de cómo se detectan las regiones de interés para localizar cada número. También se describe cómo se segmenta cada número en sus correspondientes dígitos para su posterior clasificación.

### 4.1. Descripción del telegrama

Los telegramas de escrutinio en formato planilla-imagen tienen un diseño impreso que no varía en todo el conjunto de datos (posición, dimensiones, encabezado, tablas, columnas). Los telegramas presentan tres grandes regiones: (Fig. 4.1) la superior contiene el logo de Correo Argentino, los títulos y datos de la mesa. La región media (de nuestro interés) es una tabla de 30 filas. Cada una de las primeras 25 filas corresponde a un partido político. Estas primeras 25 filas están divididas en 5 columnas: en las primeras dos están impresos los nombres de los candidatos y las tres columnas restantes que contienen los resultados asentados de la suma de votos obtenidos para cada candidato de un partido político. En este caso presenta tres columnas porque en las elecciones estudiadas en este trabajo se elegían tres tipos de cargos políticos. Las últimas 5 filas están organizadas en 4 columnas. La primer columna tiene un título en cada una de las cinco filas: votos nulos, votos recurridos, impugnados, en blanco y total por columna. Las tres columnas restantes contienen la cantidad de votos que corresponden a cada uno de los tres puestos elegibles en estas elecciones. La tercer región es el pie de página y tiene dos rectángulos donde

firma el presidente y el suplente de mesa.

## 4.2. Detección de regiones de interés

La homogeneidad en el diseño de los telegramas nos permitió obtener los números manuscritos aplicando técnicas de procesamiento de imágenes. Estos telegramas escaneados ó planillas-imagen poseen localizadas en la misma región una gran cantidad de casilleros.

Se utilizó un operador de gradiente (Sobel) y una función de detección de bordes (edge) sobre cada telegrama en formato imagen, se presentan las funciones de Matlab utilizadas en el algoritmo 1. Estas funciones se utilizaron al momento de leer la imagen escaneada en un paso previo a detectar las regiones de interés que se detalla en las secciones siguientes. Posteriormente se utilizó un algoritmo de segmentación de objetos que permitió obtener los dígitos que puedan ser clasificados utilizando una red neuronal convolucional <sup>1</sup>.

---

**Algoritmo 1** operador de gradiente y detección de bordes

---

```
1: procedure PRE_PROCESAMIENTO
2:   [ , threshold] = edge(telegrama_imagen, 'sobel');
3:   fudgeFactor = .5;
4:   BWs = edge(telegrama_imagen,'sobel', threshold * fudgeFactor);
5: end procedure
```

---

## 4.3. Detección de líneas verticales

Los telegramas de escrutinio poseen tres columnas donde se asientan manuscritos los resultados de la suma de los votos obtenidos para cada candidato de un partido político. Las tres columnas están graficadas por cuatro líneas verticales equidistantes que forman el borde izquierdo y derecho de cada columna como se observa en la fig. 4.1.

---

<sup>1</sup> <http://deeplearning.net/tutorial/lenet.html>

A partir del telegrama digitalizado se obtuvo una sub-imagen binaria de una región rectangular que no incluye al encabezado ni pie de página (fig. 4.2,B). La sub-imagen obtenida tiene píxeles con valores 0 ó 1 y representan color negro o blanco.

Se creó una proyección de la sub-imagen con el porcentaje de color blanco en cada hilera vertical de píxeles. En la proyección obtenida a partir de la sub-imagen (fig. 4.2, C), se observa mayor cantidad de píxeles negros en aquellos lugares donde hay una línea recta vertical. El criterio utilizado en la detección de las líneas verticales de las tres columnas B1, B2 y B3 (Fig. 4.1) es encontrar las posiciones en la proyección donde la cantidad de color negro supera a 2.5 desvíos estandar de la media.


#### 4.4. Detección de líneas horizontales

Las líneas horizontales dentro de la región de interés dividen los casilleros del formulario. Estas líneas delimitan las filas y forman el borde superior e inferior de los casilleros donde se encuentran los números manuscritos con la cantidad de votos. El formulario impreso en el telegrama digitalizado tiene tres columnas que llamamos B1, B2 y B3 (fig. 4.1). Cada una de las tres columnas tiene líneas rectas horizontales de trazo continuo color negro.

A partir de la sub-imagen B1 se creó una proyección con el porcentaje de color blanco en cada hilera horizontal de píxeles. En la región superior de esta proyección se observa una zona de aproximadamente 200 píxeles de alto (fig. 4.3) que tiene solamente píxeles blancos. Este diseño en los telegramas permite determinar el límite superior donde comienzan las columnas llamadas B1, B2 y B3.

Se generaron, a partir de las sub-imágenes B2 y B3, dos proyecciones adicionales con el porcentaje de color negro que tienen en cada hilera horizontal de píxeles. En las proyección se observa un porcentaje de color negro superior al 80% en aquellos lugares donde hay impresa una línea horizontal negra, esto nos permitió saber donde se encuentran las líneas horizontales y luego segmentar los casilleros de las columnas.

NO INTRODUCIR EN LA URNA - ENTREGAR AL EMPLEADO DEL CORREO - ART. 105, Ley N° 19945



01 10001X 1 1

### TELEGRAMA DE ESCRUTINIO DE MESA

Elecciones Generales del 27 de octubre de 2013  
Art. 105 - Código Electoral Nacional - LEY N° 19945  
**DISTRITO CAPITAL FEDERAL**

COMPLETAR EL FORMULARIO INDISPENSABLE PARA LA COMPENSACION - ART. 14 Y 72 DEL CEN

Destinatario: JUNTA ELECTORAL NACIONAL  
Con copia a: DIRECCION NACIONAL ELECTORAL

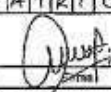

MESA  
**1**

Cantidad de electores que han votado:	
Cantidad de sobres en la urna:	

Sección	Circuito
<b>1</b>	<b>1</b>

Lista N°	PARTIDOS Y ALIANZAS	SENADORES NACIONALES	DIPUTADOS NACIONALES	DIPUTADOS/AS DE LA LEGISLATURA
187	Partido Autodeterminación y Libertad	4	8	8
601	Alianza Frente para la Victoria	71	64	43
605	Alianza Alternativa Popular	/	/	18
602	Alianza UNEN	50	38	48
603	Alianza Unión Pro	88	78	79
606	Alianza Pte. de Izq. y de los Trabajadores	23	26	25
608	Alianza Camino Popular	7	7	7
5	Partido Demócrata Cristiano	/	/	2
63	Partido Es Posible	/	/	3
68	Partido Nueva Izquierda	/	/	4
179	Partido Acción Ciudadana	/	/	/
262	Partido El Movimiento	/	/	/
264	Partido Mov. Indep. Justicia y Dignidad	/	/	/
312	Partido Federal	/	/	/
317	Partido Nacionalista Constitucional-UNIR	/	/	/
323	Partido Movimiento al Socialismo	/	/	3
800	Partido Liberal Libertario	/	/	/
901	Partido de la Red	/	/	2
902	Partido Bandera Vecinal	/	/	/
903	Partido Confianza Pública	/	/	7
904	Partido Sumar Izquierda Democrática	/	/	/
905	Partido Convergencia Popular Porteña	/	/	/
906	Partido Aluvión Ciudadano	/	/	2
907	Partido Popular de la Reconstrucción	/	/	/
Votos NULOS				1
Votos RECURRIDOS		<b>B1</b>	<b>B2</b>	<b>B3</b>
Votos IMPUGNADOS				
Votos EN BLANCO				13
TOTAL POR COLUMNA				265

Información indispensable para el cobro de la compensación de Autoridad de Mesa

A C I O S T I A C I L I A U D I A H F A T R I C I A APELLIDO Y NOMBRE	A U H E I O N I E R O M O N I C A APELLIDO Y NOMBRE
 20191750151 Documento	 1134615346 Documento
<b>PRESIDENTE</b>	<b>SUPLENTE</b>

Firma, Aclaración, Número de Documento y Número de Lista de los Fiscales

Fig. 4.1: El telegrama escaneado y las tres columnas que denominamos B1, B2 y B3. Estas columnas contienen los números manuscritos y constituyen la región de interés. B1 son votos para el puesto de Senador, B2 Diputado y B3 Diputados de la legislatura.

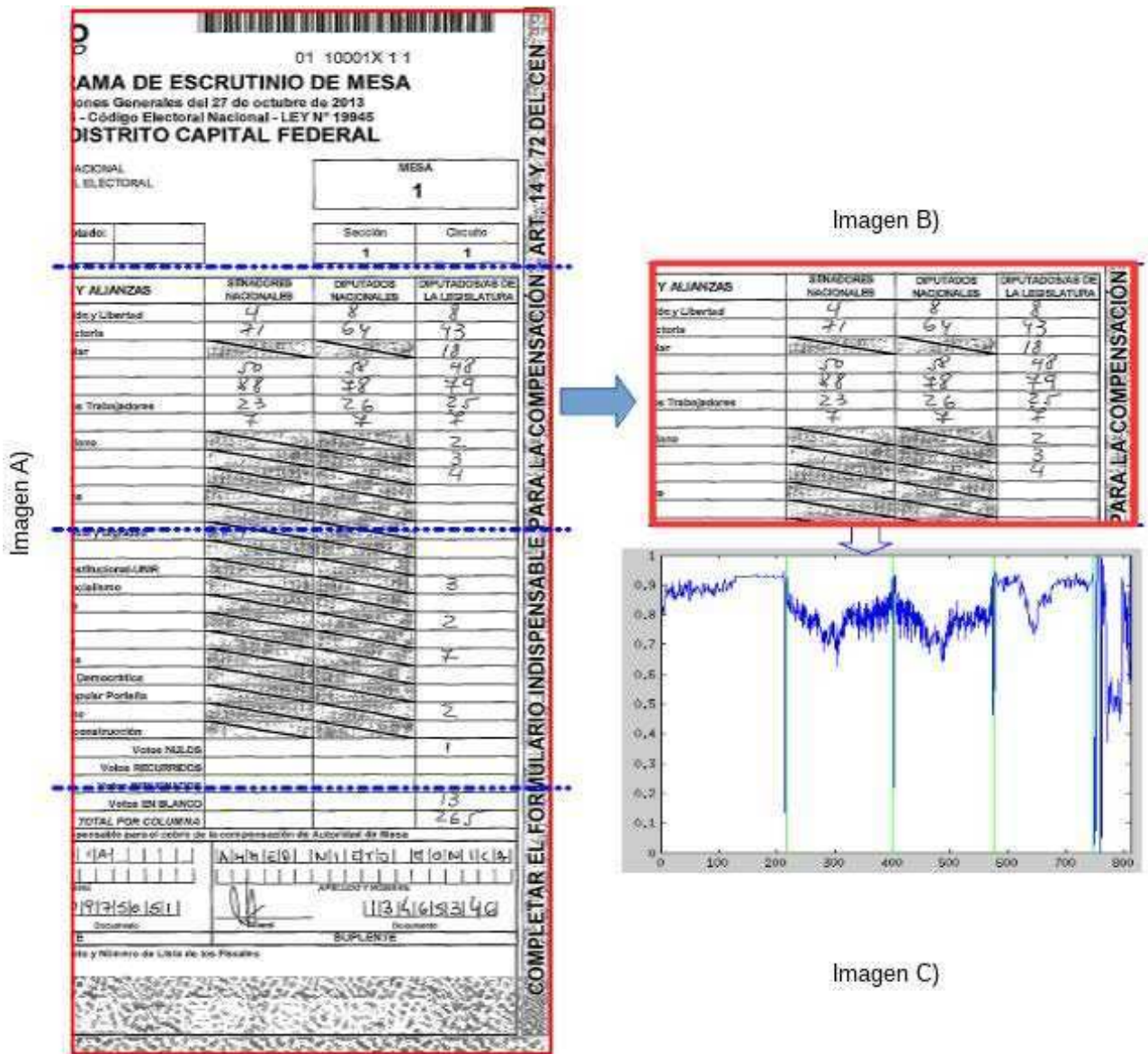


Fig. 4.2: A) El telegrama y la región de interés utilizada para extraer la Sub-Imagen. B) La Sub-Imagen de la cuál se obtiene una proyección. C) La proyección de la sub-región, aquí se observan cuatro posiciones verticales con menos cantidad de color blanco. Esto es un indicio de que hay una línea vertical.



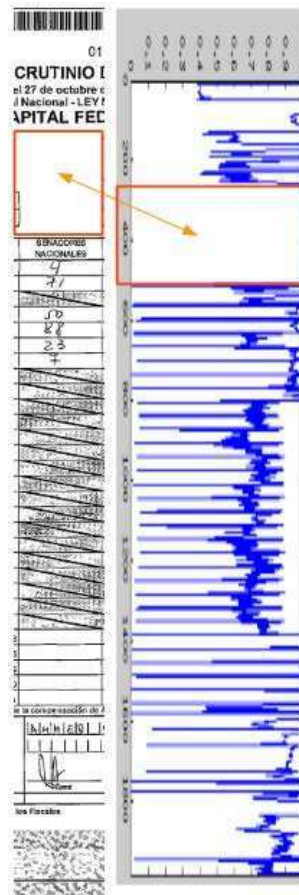


Fig. 4.3: Obteniendo los casilleros, sobre la primer columna se genera una proyección vertical. Utilizando esta proyección se determina donde se encuentra el primer casillero de la columna.

#### 4.5. Clasificación del contenido de los casilleros mediante un árbol de decisión

En la sección 4.4 se obtuvo la ubicación de los casilleros en el telegrama. Los casilleros contienen asentado en números manuscritos la suma de votos obtenidos por partido político, también hay casilleros que se encuentran vacíos (donde el fiscal de mesa no completó ningún dato) y casilleros anulados. Los casilleros anulados son algunos que por diseño del telegrama están pintados de color negro indicando que no debe completarse, sucede cuando no hay candidatos disponibles para un puesto determinado. Utilizamos un árbol de decisión para automatizar la clasificación de

cada casillero y obtener solamente casilleros con números.

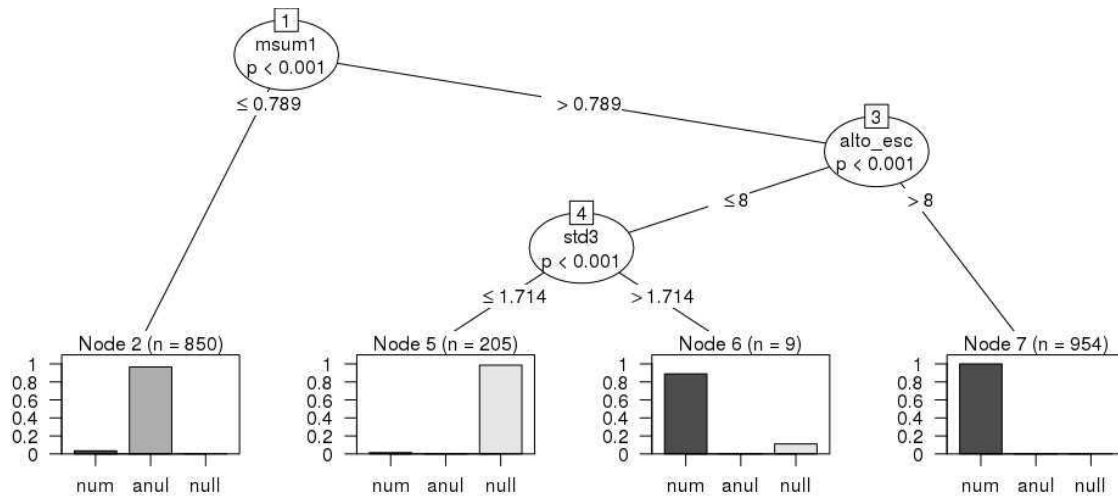
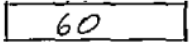

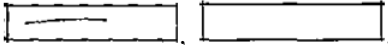


Fig. 4.4: Árbol para clasificar casilleros entre una de tres clases: casilleros con número, casillero anulado ó sin ningún contenido. El modelo utiliza tres variables.

Un modelo de clasificación utilizado en el ámbito de la inteligencia artificial es el árbol de decisión. Torsten Hothorn et al. 2006 [18] propone un algoritmo de árbol que reduce el sobreajuste. En este trabajo utilizamos la implementación de este algoritmo que se encuentra en el módulo *ctree* [19] en lenguaje R dentro del paquete *partykit*. Entrenamos el árbol de clasificación *ctree* y luego se lo utilizó para clasificar automáticamente los casilleros de los telegramas digitalizados en una de tres clases:

- Con números de uno ó varios dígitos: .
- Anulado: .
- Sin ningún tipo de contenido (en blanco): .

El conjunto de datos utilizado para entrenar el árbol de decisión tiene 2018 registros de 12 variables cada uno. Los 2018 registros se obtuvieron de 29 telegramas, cada registro tiene variables obtenidas de los casilleros. Las variables obtenidas de cada casillero son estadísticas puntuales con alguna información acerca de la imagen

y el color. Observando cada casillero se le asignó la clase correspondiente. Se utilizó el 80% de los registros obtenidos al azar para entrenamiento y el 20% restante para testing. Las variables estadísticas obtenidas de los casilleros se describen en la tabla 4.1. El modelo final puede observarse en la Fig. 4.4, este utiliza solamente tres variables del conjunto de datos.

El árbol clasificó correctamente en testing el 98% de los casilleros en una de las tres clases. Este modelo se utiliza para determinar automáticamente si un casillero tiene un número ó no.

<b>Variable</b>	<b>Descripción</b>
alto_esc	Altura del número en la celda
ancho_esc	Ancho del número en la celda
msum1	Promedio proyección vertical de la casilla
msum2	Promedio proyección vertical del número
msum3	Promedio proyección horizontal de la casilla
msum4	Promedio proyección vertical del número
std1	Desvio estandar de proyección vertical de la casilla
std2	Desvio estandar de proyección horizontal de la casilla
std3	Desvio estandar de proyección horizontal del número
std4	Desvio estandar de proyección vertical del número
std5	Desvio estandar de proyección vertical de la mitad superior del número
std6	Desvio estandar de proyección vertical de la mitad inferior del número

Tab. 4.1: Variables estadísticas puntuales obtenidas de las imágenes de los casilleros del telegrama. Estas variables conforman el conjunto de datos utilizado para entrenar un árbol de decisión y clasificar automáticamente los casilleros.

## 4.6. Segmentación de los dígitos

Un mismo número escrito por distintas personas puede tener diferencias considerables, esto se acentúa si se trata de números con múltiples dígitos. El ruido causado por diferente forma de escribir disuelve las regularidades necesarias para abordar el problema desde las técnicas de aprendizaje automático. Los números manuscritos de los telegramas están compuestos de uno y hasta tres dígitos lo cuál representa un desafío adicional dado que los métodos utilizados para clasificarlos utilizan de a un dígito en el rango de 0 a 9.

En 1995, Congedo et al. [46] presenta un procedimiento para segmentar dígitos manuscritos denominado *Deslizamiento de Gota*. En 1998, Salman Kahn [47] utiliza entre otros métodos el de “deslizamiento de gota” para segmentar y reconocer montos manuscritos en cheques de banco. La técnica está inspirada en una gota de agua hipotética que cae sobre el número y se desliza hasta llegar a la base, el recorrido que realiza esta gota deja un trazo por donde se segmentan los dígitos. En la Fig. 4.5, se observa el resultado de este método.



Fig. 4.5: Ejemplo de separación de dígitos mediante el algoritmo *Deslizamiento de gota*. En amarillo la gota cayendo y en azul la gota tras hacer contacto con el número. Este cambio de color es con fines de analizar el correcto funcionamiento del algoritmo.

## 4.7. Base de datos con dígitos manuscritos de telegramas Argentinos

La segmentación de los números manuscritos en sus correspondientes dígitos (Sección 4.6) nos permitió crear una base de imágenes de dígitos manuscritos. La

asignación de clase a cada dígito se automatizó utilizando los números en texto plano, tipeados por una persona que lee el telegrama (Sección 3.5). Cada dígito está centrado en un casillero de 24x24 píxeles, fondo negro y dígito color blanco (Fig. 4.6). Este es un conjunto de datos con imágenes de dígitos manuscritos que provienen de los telegramas de elecciones legislativas en la República Argentina, lo llamamos **telSET1**. El conjunto de dígitos manuscritos está compuesto por 96.927 dígitos, escritos por 2151 personas diferentes.

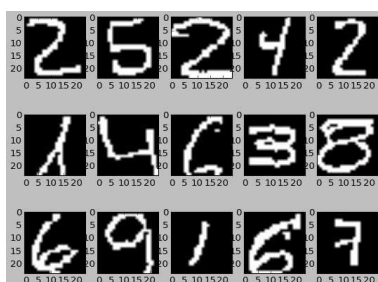


Fig. 4.6: Algunos dígitos del conjunto telSET1.

Dígito MNIST									
0	1	2	3	4	5	6	7	8	9
5923	6742	5958	6131	5842	5421	5918	6265	5851	5949

Tab. 4.2: Cantidad de dígitos por categoría del conjunto MNIST utilizado para entrenar.

(Fig. 4.7)

Dígito Telegrama									
0	1	2	3	4	5	6	7	8	9
14414	25278	9477	6857	7371	7082	7793	7052	6426	5177

Tab. 4.3: Cantidad de dígitos por categoría de los telegramas utilizados para entrenar.

(Fig. 4.8)

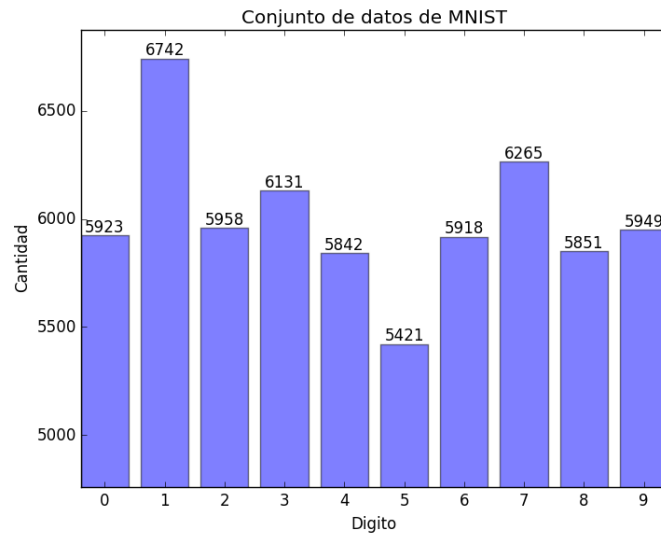


Fig. 4.7: Cantidad de dígitos por categoría del conjunto MNIST utilizado para entrenar. (Tabla 4.2)

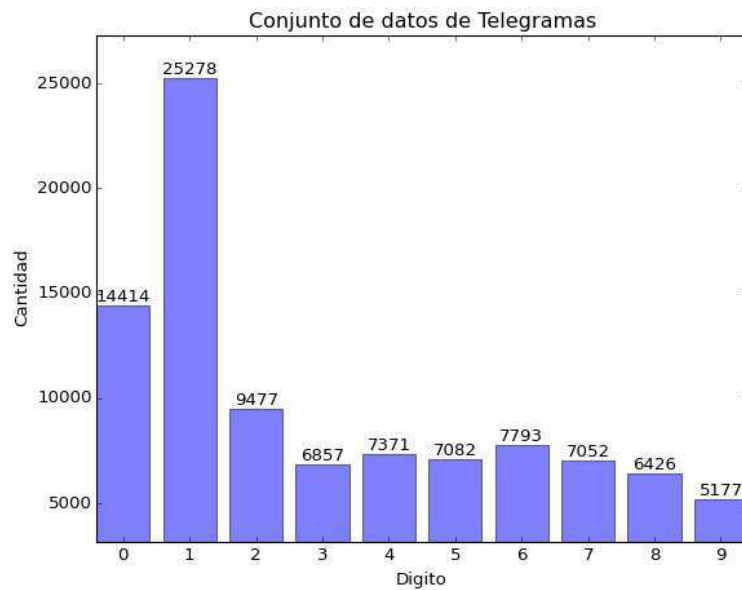


Fig. 4.8: Cantidad de dígitos por categoría en telegramas utilizados para testear. Una de las características de los resultados en estas elecciones es una mayor cantidad de dígitos uno, seguidos por el dígito cero. (Tabla 4.3)

A diferencia del conjunto de datos de dígitos manuscritos MNIST compuesto por 60.000 dígitos escritos por 250 personas distintas <sup>2</sup>. y que tiene una distribución homogénea de clase de dígitos (Fig. 4.7), el nuevo conjunto de datos **telSET1**, basado en telegramas electorales Argentinos tiene una distribución de dígitos desbalanceada (Fig. 4.8). El desbalance se debe a factores como: son datos de elecciones, la cantidad de candidatos que se presentan, cantidad de puestos a elegir, cantidad de votantes que participaron.

En este trabajo se clasifican los dígitos de el conjunto de datos **telSET1** utilizando una red neuronal que aprende las características de dígitos manuscritos de diferente procedencia como lo son los dígitos MNIST.

#### 4.8. Clasificación de dígitos basado en Redes Neuronales

Una de las ramas más destacadas del campo científico de la Inteligencia Artificial es la que corresponde a Redes Neuronales Artificiales (RNA), son redes en las que existen elementos procesadores de información de cuyas interacciones locales depende el comportamiento del conjunto del sistema (Hilera y Martínez, 1995, p. 16) [48]. Sietsma y Dow (1991) [49] analizan diferentes arquitecturas de redes neuronales y mencionan que una de las fortalezas de este tipo de redes es su habilidad de reconocer ó clasificar correctamente patrones que no se hayan presentado anteriormente a la red. En el 2009, Ordoñez et al. [50] aplican redes neuronales en reconocimiento de caracteres ascii, mencionan que una de las fortalezas de las RNA es la posibilidad de aprender mediante la experiencia.

Al crear la red neuronal se comienza entrenándola, se le “enseña” la respuesta esperada a una entrada específica. Este proceso de aprendizaje está controlado por un agente externo (supervisor ó maestro) que establece cuál es la respuesta que debe generar la red a partir de una muestra dada. El supervisor observa la salida de la red y puede proceder a modificar los pesos de las conexiones si la red ha clasificado incorrectamente, con el fin de que la salida de la red se aproxime al

---

<sup>2</sup> <http://yann.lecun.com/exdb/mnist/>

resultado esperado.

Cada presentación de todo el conjunto de datos de entrenamiento a la red se denomina una época (epoch en inglés). La cantidad de épocas (cantidad de veces que se le muestra a la red el conjunto de datos) afecta la performance de la red. En 2001, Rafiq et al. [58] publica un trabajo acerca de redes neuronales artificiales aplicadas a ingeniería, aquí describe los siguientes factores que condicionan el número de épocas a utilizar:

- Cantidad de muestras en el conjunto de datos
- Cantidad de capas ocultas
- Cantidad de neuronas en las capas ocultas
- Cantidad de parámetros de salida

#### 4.8.1. Red Neuronal Convolutiva

Las redes neuronales consisten en la organización y disposición de las neuronas en red formando capas o agrupaciones de neuronas. El aprendizaje es el proceso por el cual una red neuronal modifica sus pesos en respuesta a una información de entrada. En este trabajo se utilizó un tipo de red neuronal con entrenamiento supervisado llamada Red Neuronal Convolutiva. Esta clase de redes ofrecen cierto grado de tolerancia a distorsiones y desplazamiento, ventajas que se obtienen por la implementación de pesos compartidos, campos receptivos locales y sub-muestreo. Los pesos son compartidos por neuronas en diferentes posiciones de la misma capa y cada capa de convolución aprende filtros de manera automática a partir de ciertos patrones que encuentra reiteradas veces en determinadas posiciones de las imágenes (Dean et al. 2012 [51]).

En 1998, Yan LeCun et al. [34] utilizó redes neuronales convolucionales en reconocimiento de documentos sobre el conjunto de dígitos manuscritos MNIST [35] y clasificó correctamente el 99,3% de los 10.000 dígitos de test. En este trabajo se utiliza una ConvNet de cinco capas: dos de convolución, dos de sub-muestreo y



una de clasificación. Para su implementación se utilizó lenguaje Python utilizando la librería de aprendizaje profundo Lasagne (Dieleman et al., 2015 [52]) y Theano (Bastien et al., 2012 [53]).

Se validó la ConvNet utilizada en el presente trabajo sobre el conjunto de dígitos manuscritos MNIST sin distorsiones. Se le presentan los datos de entrenamiento 30 veces (30 épocas) y para el test se utilizaron los 10000 dígitos del conjunto de datos destinados a este propósito. En la Fig. 4.9 se presentan las 30 épocas y la clasificación de los dígitos de testing del conjunto de datos MNIST. El acierto obtenido es de 99.02 % que es similar a los resultados obtenidos en publicaciones como Yan LeCun et al. [34] donde que se utilizan redes convolucionales junto con los datos de MNIST. Los parámetros utilizados en la red se describen a continuación.

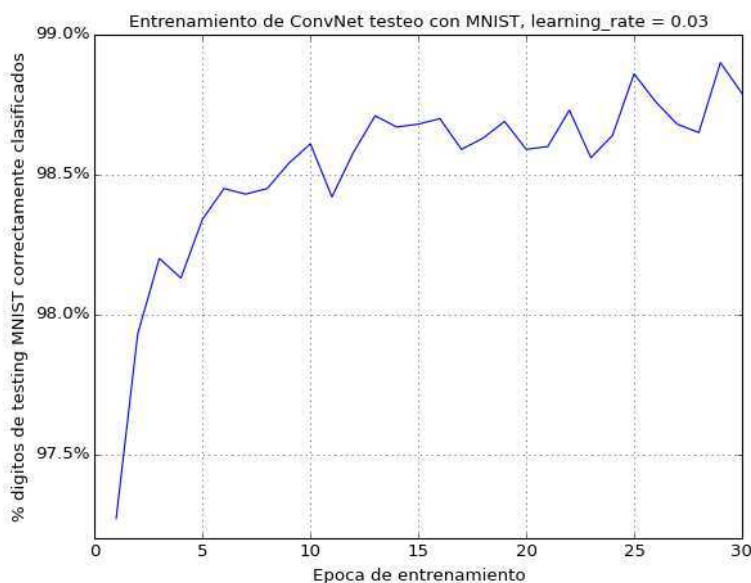


Fig. 4.9: Entrenamiento de la ConvNet y test sobre MNIST

Los componentes de la ConvNet están implementados en la librería de aprendizaje profundo Lasagne (Dieleman et al., 2015 [52]), se utilizó para la capa de entrada a la ConvNet la función `lasagne.layers.InputLayer`<sup>3</sup> a la que se le especifica una entrada de 24x24.

<sup>3</sup> <http://lasagne.readthedocs.io/en/latest/user/layers.html>

Las **capas convolutivas** se definieron utilizando la función `lasagne.layers.Conv2DLayer`<sup>4</sup> especificando la cantidad de filtros para cada capa, el tamaño del padding y la cantidad de píxeles que debe moverse el filtro en cada operación de convolución (*stride*). La primera capa convolutiva tiene 6 filtros y la segunda capa tiene 16 filtros (*kernels*). El tamaño de padding para ambas capas convolutivas es de 2 píxeles, el parámetro *stride* es de 1.

Para las **capas de sub-muestreo** se llamó a la función `lasagne.layers.MaxPool2DLayer`<sup>5</sup> donde se le indica el parámetro *stride* con valor 1 y un tamaño de la ventana de 2x2.

La última capa de la red neuronal es **totalmente conectada** compuesta por 10 neuronas que se conectan a todas las neuronas de la capa previa. Estas 10 neuronas de salida representan los 10 dígitos posibles, se utilizó la función `lasagne.layers.DenseLayer`<sup>6</sup>.

Se adjunta al apéndice B el código fuente de la implementación utilizada de ConvNet en Python. La Fig. 4.10 es un esquema de la arquitectura de la red utilizada en este trabajo.

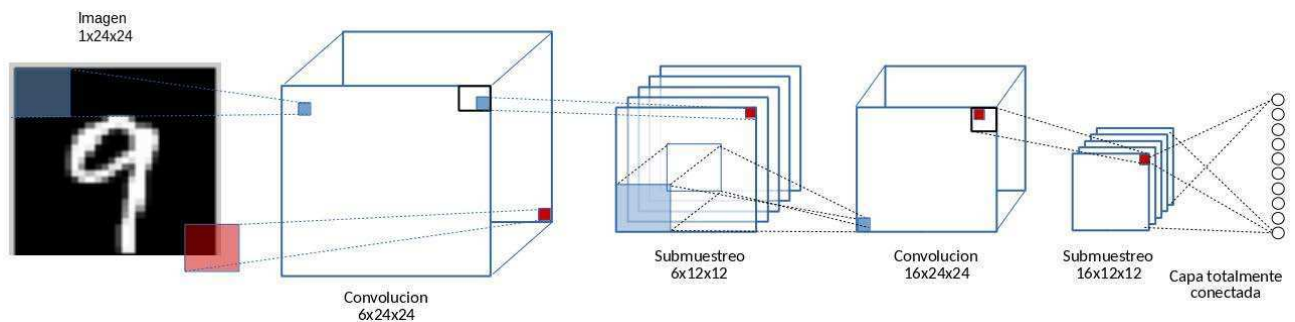


Fig. 4.10: Arquitectura de la ConvNet utilizada en el presente trabajo

<sup>4</sup> <http://lasagne.readthedocs.io/en/latest/modules/layers/conv.html>

<sup>5</sup> <http://lasagne.readthedocs.io/en/latest/modules/layers/pool.html>

<sup>6</sup> <http://lasagne.readthedocs.io/en/latest/modules/layers/dense.html>

## 5. RECONOCIENDO DÍGITOS A PARTIR DE TELEGRAMAS ELECTORALES

En esta sección se describen los experimentos realizados y los resultados obtenidos. Al tratarse de datos que consisten en dígitos manuscritos tomamos un set de datos de imágenes de números escritos a mano muy difundido: MNIST. La base de datos MNIST [35] constituye un estándar a la hora de testear y comparar métodos de reconocimiento de dígitos manuscritos. La misma contiene 70000 imágenes de números escritos a mano que incluyen diferentes estilos de escritura (Seijas et al, 2003 [6]).

En esta tesis se crearon conjuntos de datos derivados del set de dígitos manuscritos MNIST, con ellos se entrenó una ConvNet y se obtuvieron tres modelos. Cada uno de los tres modelos de ConvNet aprendidos se utilizaron para clasificar los dígitos manuscritos extraídos de telegramas. Los dígitos de los telegramas no fueron utilizados durante el aprendizaje de la red. Se desea comparar los modelos entrenados con muestras obtenidas del conjunto MNIST y testeados con dígitos manuscritos de un origen diferente como son los dígitos manuscritos de los telegramas electorales de escrutinio.

### 5.1. Preparación de los datos y entrenamiento

Se prepararon tres conjuntos de datos a partir de los dígitos manuscritos MNIST y se los utilizó para entrenar la ConvNet. Se llamó a los conjuntos de datos SET1, SET2 y SET3. De cada conjunto de datos se obtuvo un clasificador. El SET1 está compuesto por los 60000 dígitos de entrenamiento MNIST. Al SET2 y SET3 lo conforman los mismos dígitos del SET1 sumando los mismos dígitos habiéndoles

aplicado pequeñas distorsiones de tamaño y apariencia <sup>1</sup>. Esto da como resultado un nuevo conjunto de datos con mayor cantidad de muestras del mismo tamaño. Las distorsiones de aspecto aplicadas reducen el tamaño del dígito manteniéndolo dentro de los márgenes originales de la imagen de 24x24 píxeles.

Ciresan et al., 2010 [57] aplica distorsiones sobre los dígitos MNIST para entrenar una ConvNet. En el presente trabajo se utilizaron dos tipos de distorsiones diferentes sobre los dígitos originales. Cada distorsión se utilizó por separado en cada nuevo conjunto de datos llamados SET2 y SET3. Al aplicar estas distorsiones sobre la base de datos original, se incorpora variabilidad en los datos. Se desea validar si mejora la capacidad de generalización del modelo.

También se preparó un set de datos de dígitos manuscritos provenientes de telegramas electorales. A este conjunto de datos se lo llamó TelSET1, compuesto por 96927 dígitos, escritos por 2151 personas diferentes. Estos dígitos manuscritos de telegramas son utilizados para testear los tres modelos aprendidos.

El tamaño final de cada uno de los set de datos es:

- SET1: Sin distorsión, base de datos original (60.000 dígitos).
- SET2: Distorsión alto y ancho (540.000 dígitos).
- SET3: Distorsión senoide (540.000 dígitos).
- TelSET1: 96927 dígitos manuscritos de los telegramas.

En 1998, Yan LeCun et al. [34] utiliza redes neuronales convolucionales en reconocimiento de documentos sobre el conjunto de dígitos manuscritos MNIST y sugiere que los resultados obtenidos de ConvNets podrían mejorar a medida que aumenta el tamaño del conjunto de datos utilizado para entrenamiento. Utilizar un set de datos de mayor tamaño podría sesgar el resultado a favor de un modelo simplemente por haber entrenado con mas muestras y no por la posible información adicional que puedan aportar distorsiones aplicadas sobre los datos. Para resolverlo

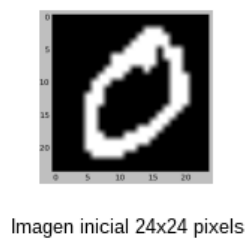
---

<sup>1</sup> <http://www.mathworks.com/help/images/examples/creating-a-gallery-of-transformed-images.html>

se parametriza que durante cada época de entrenamiento, la red neuronal aprenderá de 60.000 dígitos tomados al azar (es la cantidad de dígitos que componen el set de datos de entrenamiento MNIST). Para el SET2 y SET3 se toman al azar 60.000 dígitos en cada época. Como se menciona en la sección 4.8, cada presentación de todo el conjunto de datos de entrenamiento a la red se denomina una época (epoch en inglés).

### 5.1.1. Formato de los datos

Enachescu y Miron, 2009 [59] utilizan redes neuronales convolucionales para clasificar dígitos manuscritos, respecto al formato de las imágenes (grayscale, binario) menciona que influye en el proceso de aprendizaje de la red neuronal y puede significar menos tiempo necesario por época de entrenamiento. Las imágenes provenientes del conjunto MNIST y telSET1 (dígitos manuscritos de los telegramas) fueron formateadas de la siguiente manera: los píxeles blancos del dígito manuscrito se teados en 1 y los píxeles del fondo en 0. Los píxeles de los dígitos tienen valores entre 0 y 1, se los binariza utilizando un umbral de 0.4: Todo píxel mayor a 0.4 es seteado en 1, menor ó igual a 0.4 en 0. Un ejemplo de la codificación se presenta en la Fig. 5.1



```

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 0 1 1 1 0 0 0 0
0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 0 1 1 1 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 0 0 1 1 0 0 0 0
0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 0 0 1 1 0 0 0 0
0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 0 0 1 0 0 0 0 1 1 0 0 0
0 0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0
0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0
0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0
0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0
0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0
0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0
0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0
0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0
0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0
0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

```

Matriz del dígito para entrenamiento

Fig. 5.1: Ejemplo de formato de un dígito MNIST

A continuación se describe la composición de los tres conjuntos de datos utilizados para crear los modelos.

### 5.1.2. SET1: Set de datos sin distorsión

En un conjunto de datos se utiliza la base de datos MNIST sin aplicar ninguna distorsión a los datos. Se entrena con los 60.000 dígitos. En la Fig. 5.2 (a) se observan los primeros dígitos del set MNIST sin aplicar distorsiones.

### 5.1.3. SET2: Set de datos modificando alto y ancho

Se generan nuevos dígitos modificando el tamaño de los datos originales. Las imágenes del conjunto de datos original son modificadas en tamaño, se redujeron hasta cuatro y dos píxeles de alto y ancho cada dígito, el tamaño original de los casilleros que contienen los dígitos se mantuvo en 24x24 píxeles. El conjunto de datos pasa de 60.000 a  $60.000 * 3 * 3 = 540.000$  registros. En la Fig. 5.2 (b) se observan algunos dígitos de este nuevo conjunto de datos.

### 5.1.4. SET3: Set de datos con distorsión senoide

Se aplica una transformación senoide utilizando la función `tformarray` de Matlab <sup>2</sup>. Se observó que aplicada una distorsión senoide utilizando valores entre -2 y 2 se obtienen dígitos similares a los originales aunque con pequeñas diferencias como se observa en la Fig. 5.2 (c). El conjunto de datos se incrementa 9 veces, de 60.000 a 540.000 dígitos incluidos en el SET3.

## 5.2. Características de la red

Las redes neuronales utilizan el algoritmo descenso de gradiente para aprender los pesos. Este algoritmo puede ser implementado utilizando un método por lotes ó

---

<sup>2</sup> <http://www.mathworks.com/help/images/examples/creating-a-gallery-of-transformed-images.html>

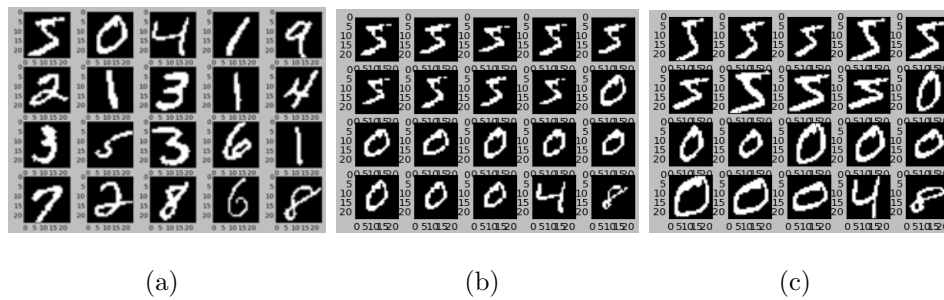


Fig. 5.2: (a) Dígitos MNIST sin modificaciones. (b) Dígitos a los que se les modifica la escala. (c) Dígitos con distorsión sinusoidal.

uno en línea (Mu Li et al., 2014 [60]). En el método por lotes, los cambios en los pesos se aplican tras recorrer todo el conjunto de datos de entrenamiento (una época). Por otro lado, el entrenamiento en-línea actualiza los pesos tras la presentación de cada registro del conjunto de datos de entrenamiento (ó instancia). Una tercer alternativa, intermedia entre estos dos métodos es el entrenamiento por mini-lotes<sup>3</sup>, en el que los cambios en los pesos se acumulan tras una cantidad de instancias  $\mu$ , antes de modificar los pesos. Si la frecuencia (ó tamaño de mini-lote) con la que se modifican los pesos es  $\mu = 1$  el método es en línea, mientras que si  $\mu = N$  es entrenamiento por lotes, donde  $N$  es la cantidad de instancias en todo el conjunto de datos de entrenamiento (Wilson y Martinez, 2003 [61]).

Experimentalmente se buscan los valores para los parámetros de esta red neuronal. Los parámetros a determinar son: tamaño de mini-lote y velocidad de aprendizaje  $\alpha$ . Para esta exploración de parámetros se utilizó el conjunto de datos SET1 y los resultados se evalúan con 9000 dígitos manuscritos tomados al azar procedentes de los telegramas.

Estos experimentos se realizaron en una computadora con procesador Intel(R) Core(TM) i7 CPU 950 @ 3.07GHz, sistema operativo Debian Linux 8, Kernel 3.16.0-4-amd64 y 8 Gigas de memoria.

Se probaron diferentes valores de mini-lote (entre 5 y 1300), tras 30 épocas la red estabiliza su aprendizaje (Fig. 5.4). Puede observarse en la Fig. 5.3 que un

<sup>3</sup> <http://www.faqs.org/faqs/ai-faq/neural-nets/part2/>

mini-lote de tamaño entre 10 y 300 es recomendable para esta red y conjunto de datos. Este tamaño de mini-lote coincide con resultados obtenidos un trabajo de Glorot et al., 2011 [62] en el que evalúa diferentes parámetros para redes neuronales convolucionales.

El parámetro  $\alpha$  (velocidad de aprendizaje) utilizado en las redes neuronales tiene un valor entre 0.001 y 0.5, el valor óptimo depende en parte del conjunto de datos y de las características de la red. En este trabajo, se probaron valores entre 0.001 y 0.1 (0.001; 0.01; 0.05) utilizando el conjunto de datos SET1 y se testeó con el set de dígitos manuscritos obtenidos de los telegramas TelSET1. En la Fig. 5.5 se observan los resultados de tres experimentos: las 30 épocas con la tasa de acierto sobre 9000 dígitos provenientes de los telegramas, seleccionados al azar. La tasa de acierto sobre todo el conjunto de dígitos de telegramas se evalúa tras la última época, en la Fig. 5.6 se comparan los diferentes valores de  $\alpha$  utilizados y el resultado obtenido. El valor de  $\alpha$  con el que se obtuvieron mejores resultados es 0.03 para esta red y este conjunto de datos.

### 5.2.1. Inicialización de parámetros

Los pesos en cada capa fueron inicializados utilizando una distribución Gaussiana con media cero y desvío estandar 0.01. Los valores de sesgo (bias) en las capas convolutivas y la capa totalmente conectadas fueron inicializadas con el valor constante 1. El valor de Momentum utilizado fue 0.9. Otros estudios muestran que esta inicialización de parámetros acelera las etapas iniciales de aprendizaje cuando se utiliza la función de activación ReLU (unidades lineales rectificadas). ReLU calcula la función  $f(x) = \max(0, x)$ , es simplemente un umbral en cero (Fig. 5.7). Otros autores como Dahl et al., 2013 [54], Chen et al. y 2014 [55] utilizan esta función de activación en implementaciones de redes convolucionales, también Krizhevsky et al., 2012 [39] utiliza esta función de activación en la ConvNet ImageNet con la que obtuvo el primer lugar en la competencia ImageNet LSVRC.



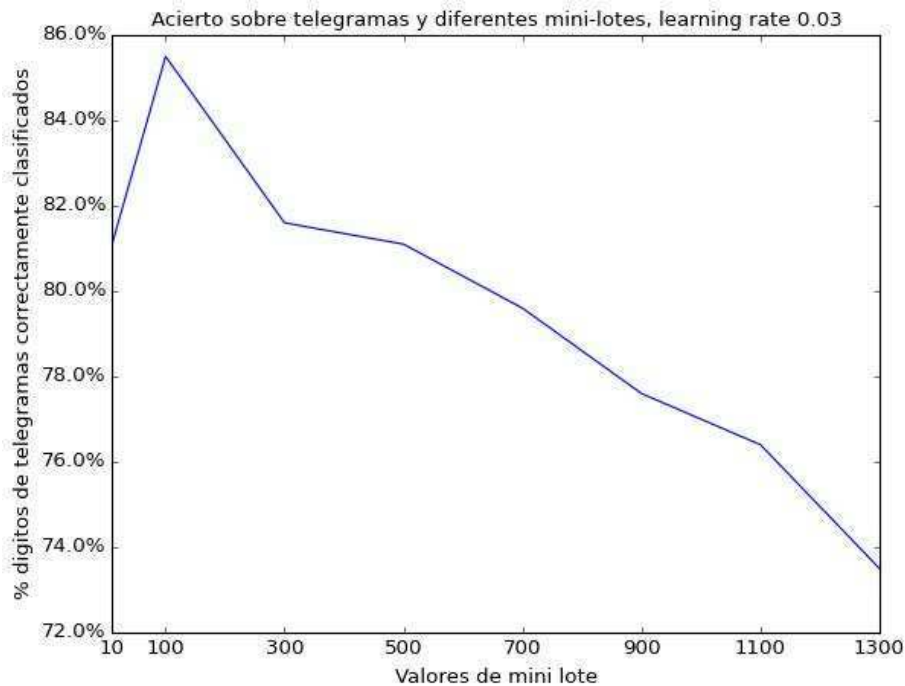


Fig. 5.3: Se exploran valores de mini-lote. Un mini-lote de 100 muestras posibilita obtener un mayor porcentaje de dígitos correctamente clasificados.

### 5.3. Resultados obtenidos

En cada experimento se utilizó un set de datos diferente con el que se entrenó la red neuronal. Cada modelo obtenido se utilizó para reconocer los dígitos manuscritos provenientes de los telegramas.

- Experimento 1 : Se utilizó el SET1, se obtuvo Modelo 1
- Experimento 2 : Se utilizó el SET2, se obtuvo Modelo 2
- Experimento 3 : Se utilizó el SET3, se obtuvo Modelo 3

Se buscó determinar si se obtienen mejoras en la clasificación de dígitos al extender el conjunto de datos utilizado para entrenar, aplicando ciertas distorsiones. En cada uno de los experimentos, la ConvNet entrenó durante 30 épocas, al final de cada época de aprendizaje se testeó con 9000 dígitos tomados al azar del conjunto TelSET1. En la Fig. 5.8 se observan las 30 épocas y el porcentaje de dígitos

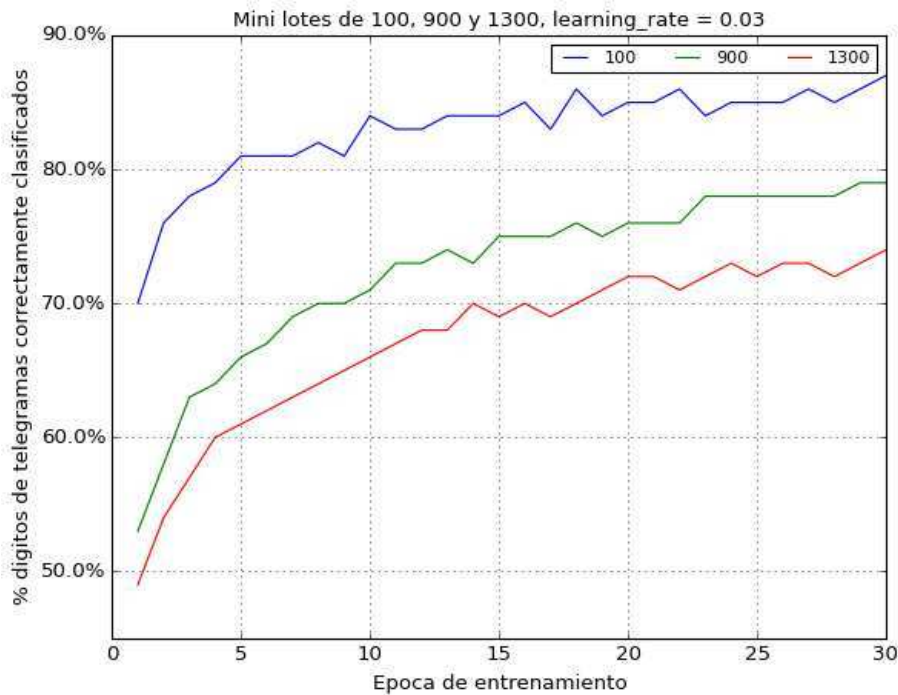


Fig. 5.4: Tres tamaños de mini-lote (100, 900 y 1300) por época de entrenamiento.

manuscritos de telegramas leídos correctamente por cada uno de los modelos. Con cada uno de los tres modelos creados se testeó la clasificación de dígitos manuscritos de telegramas. Puede observarse que el Modelo 3 (obtenido con el SET3) permite obtener una lectura con mayor cantidad de aciertos que los otros modelos Fig. 5.9.

En la tabla Tab. 5.1 se presenta la cantidad de aciertos y errores detallado por dígito. En la Tab. 5.2 se muestran los errores en valores relativos al total, los casilleros color naranja son aquellos dígitos que tuvieron mayor cantidad de errores. Analizando la tabla de errores relativos se observa que el modelo utilizado ha cometido mayor cantidad de errores para los dígitos 1, 7 y 8. En particular el dígito 1 es confundido como 7 y 4. El dígito 7 es confundido por el modelo por el número 2 y el dígito 8 ha sido incorrectamente clasificado como 3 en ciertos casos. Algunos de estos dígitos incorrectamente clasificados se muestran en la Fig. 5.10.

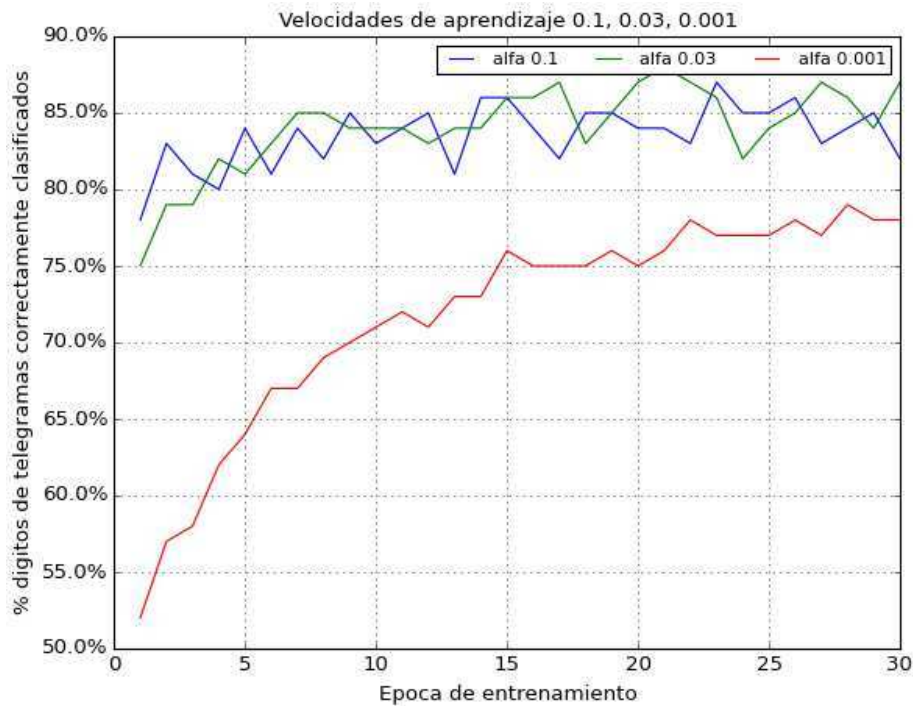


Fig. 5.5: Tres velocidades de aprendizaje distintas (0.1, 0.03 y 0.001).

### 5.3.1. Visualización de la primera capa

En redes neuronales convolucionales, cuando se trabaja con imágenes, puede resultar de utilidad visualizar los filtros aprendidos. En este caso, se trabaja con dígitos manuscritos y pueden distinguirse ciertos patrones, este es un indicio de que el entrenamiento ha sido correcto. En la Fig. 5.11 se observan algunas propiedades aprendidas en la primera y en la segunda capa convolutiva. Cada imagen es un mapa de propiedades diferente (también conocidos como filtros ó kernels). Cada mapa tiene un tamaño de 5 x 5 correspondiente a los pesos de 5 x 5 en el campo receptivo. Bloques blancos son valores de pesos pequeños (o más negativos), por lo tanto el mapa de propiedades tendrá menor respuesta a los píxel que reciba. Bloques oscuros significan pesos de mayor valor, aquí el mapa de propiedades tendrá mayor respuesta a píxeles de entrada. El hecho de que los mapas de propiedades tengan claramente sub regiones definidas con color claro y otras con color oscuro, permite entender que la red está aprendiendo estructuras de formas.

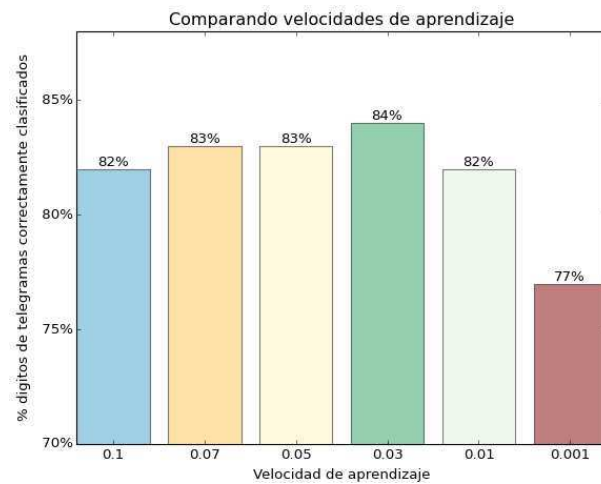


Fig. 5.6: Se exploran velocidades de aprendizaje (0.1 a 0.001), un valor de 0.03 da mejores resultados en la clasificación.

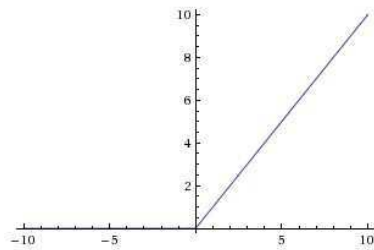


Fig. 5.7: Función de activación Rectified Linear Unit (ReLU), devuelve cero cuando  $x < 0$ .

### 5.3.2. Errores de transcripción

De los primeros 2000 telegramas seleccionados la ConvNet detecta correctamente en 21 telegramas algún dígito inconsistente con lo transcripto. En el apéndice A se muestran algunos de estos telegramas.

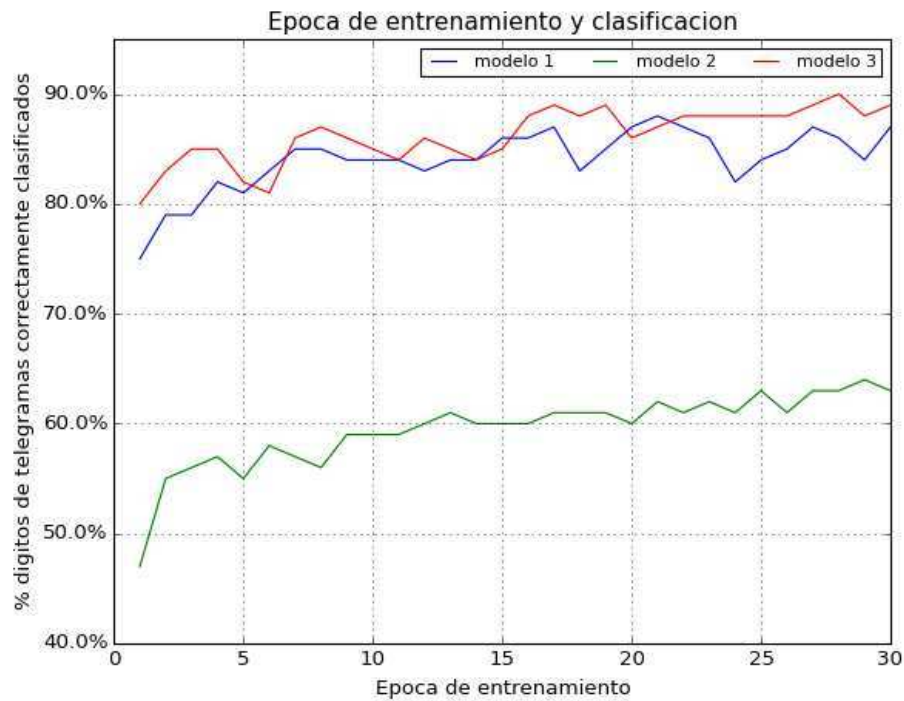


Fig. 5.8: Para cada uno de los tres modelos, el porcentaje de dígitos correctamente clasificados por época de entrenamiento.

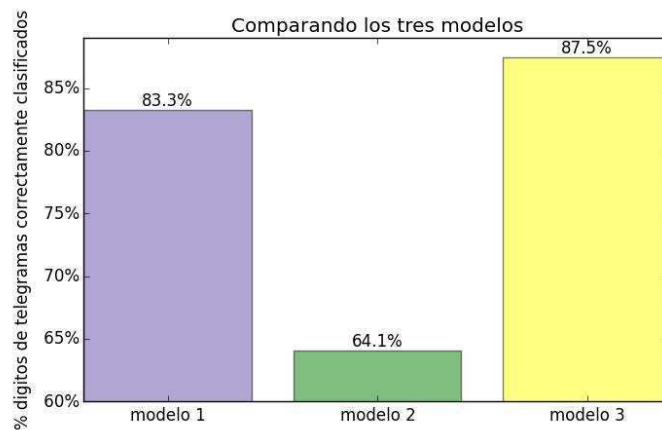


Fig. 5.9: Modelo 1: entrenado con MNIST sin modificaciones. Modelo 2: entrenado con conjunto MNIST de dígitos reescalados. Modelo 3: entrenado con dígitos MNIST habiendo aplicado una distorsión sinusoidal. Con cada modelo se clasificaron los dígitos provenientes de telegramas y cada uno tuvo un porcentaje de dígitos correctamente clasificados. El modelo 3 dió mejores resultados.

		Dígito correcto									
		0	1	2	3	4	5	6	7	8	9
Dígito estimado	0	14043	120	13	0	27	2	143	0	33	9
	1	1	20389	4	0	62	0	4	66	0	0
	2	158	429	8870	58	51	6	64	558	79	68
	3	18	151	85	6741	2	67	81	107	342	114
	4	4	1375	9	0	6755	0	0	299	7	96
	5	26	134	12	12	13	6943	1172	9	42	21
	6	9	272	3	0	179	6	6050	1	7	4
	7	60	1953	49	39	33	26	4	5960	25	137
	8	61	299	235	2	22	19	268	40	5788	523
	9	34	156	197	5	227	13	7	12	103	4205

Tab. 5.1: Matriz de confusión en valores absolutos. La diagonal muestra cuántos dígitos fueron correctamente clasificados. Hay mayor cantidad de dígitos 1 (uno) incorrectamente debido a que este dígito es más frecuente en estos datos. El acierto de este modelo es de 88.46%. En la Tab. 5.2 se calculó la cantidad relativa de errores

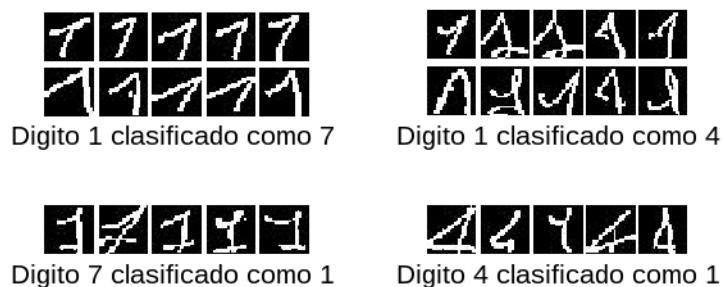


Fig. 5.10: Aquí pueden observarse algunos de los dígitos que fueron incorrectamente clasificados por la ConvNet.

		Dígito correcto									
		0	1	2	3	4	5	6	7	8	9
Dígito estimado	0	0.97	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.01	0.00
	1	0.00	0.81	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00
	2	0.01	0.02	0.94	0.01	0.01	0.00	0.01	0.08	0.01	0.01
	3	0.00	0.01	0.01	0.98	0.00	0.01	0.01	0.02	0.05	0.02
	4	0.00	0.05	0.00	0.00	0.92	0.00	0.00	0.04	0.00	0.02
	5	0.00	0.01	0.00	0.00	0.00	0.98	0.15	0.00	0.01	0.00
	6	0.00	0.01	0.00	0.00	0.02	0.00	0.78	0.00	0.00	0.00
	7	0.00	0.08	0.01	0.01	0.00	0.00	0.00	0.85	0.00	0.03
	8	0.00	0.01	0.02	0.00	0.00	0.00	0.03	0.01	0.90	0.10
	9	0.00	0.01	0.02	0.00	0.03	0.00	0.00	0.00	0.02	0.81

Tab. 5.2: Matriz de confusión en valores relativos. Los recuadros naranja señalan aquellos dígitos para los que hubo mayor cantidad de errores al clasificarlos. En la fig. 5.10 se muestran algunos ejemplos de los errores de detección automática.

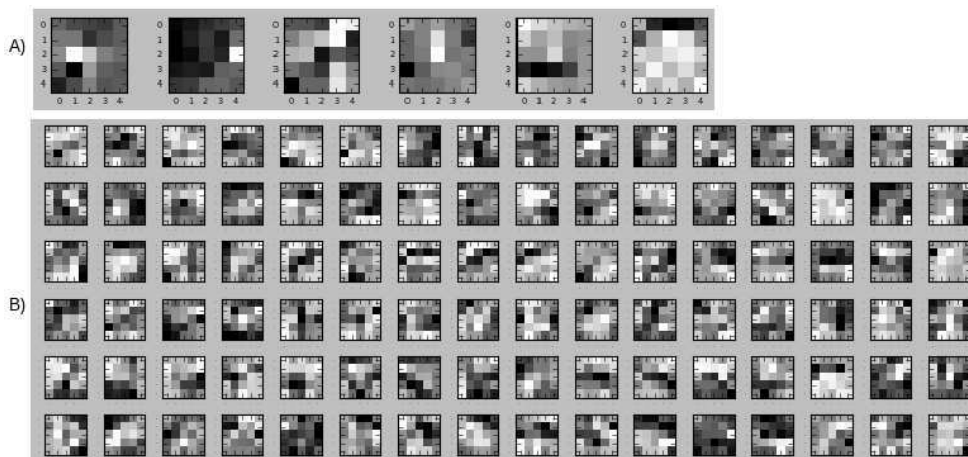


Fig. 5.11: a) 6 propiedades de la primera capa convolutiva de tamaño 5x5. b) 16 propiedades de la segunda capa convolutiva de tamaño 5x5x6.

## 6. CONCLUSIONES Y TRABAJOS FUTUROS

El presente trabajo es un caso de estudio en el que se han utilizado Datos Abiertos de Gobierno producidos en Buenos Aires, Capital Federal durante la jornada electoral del 27 de Octubre de 2013. Se ha utilizado contenido digital consistente en: planillas de cálculo, archivos de imágenes y formatos estandarizados.

En el acceso a Datos Abiertos de Gobierno identificamos ventajas y dificultades	
Ventajas	Dificultades
Los dígitos manuscritos son consistentes con los datos tipeados	Breve disponibilidad de tiempo por el que se encuentran disponibles los telegramas manuscritos digitalizados, una vez concluidas las elecciones.
Los archivos de texto con resultados tipeados se obtienen del portal de Datos Abiertos de la Ciudad de Buenos Aires [69]	Los archivos de texto con resultados tipeados disponibles no están acompañados por la imagen del telegrama de donde proceden dichos datos, dificultando la validación de lo manuscrito con lo transcripto.
Posterior a las elecciones, por unos días, están disponibles los telegramas manuscritos escaneados.	Para obtener los telegramas escaneados es necesario utilizar algún programa específico que automatice la extracción.
El acceso a los telegramas y datos tipeados es anónimo sin intermediarios ni demoras	



Algunas de las inconsistencias identificadas corresponden al partido político número 905 denominado Convergencia Popular Porteña. En la Fig. 6.1 se observan los resultados finales publicados en el diario La Nación. El resultado de cero votos para un partido político llamó la atención de la prensa que publicó la anécdota en varios medios como el diario La Nación [65], La Gaceta [66] y La Voz [67] entre otros. El motivo por el que este partido político obtuvo pocos votos se debe a que en un comunicado de último momento solicitaron que sus electores votasen a otro partido denominado “Aluvión Ciudadano”.

Agrupaciones políticas	Votos		Estimación de cargos
UNION PRO	618.496	33,59%	12
UNEN	454.354	24,68%	8
FRENTE PARA LA VICTORIA	315.418	17,13%	6
CONFIANZA PUBLICA	108.231	5,88%	2
FRENTE DE IZQUIERDA Y DE LOS TRABAJADORES	91.313	4,96%	1
ALTERNATIVA POPULAR	77.880	4,23%	1
AUTODETERMINACION Y LIBERTAD	51.059	2,77%	
CAMINO POPULAR	35.247	1,91%	
PARTIDO DE LA RED	21.368	1,16%	
ES POSIBLE	13.082	0,71%	
MOVIMIENTO AL SOCIALISMO	8.344	0,45%	
NUEVA IZQUIERDA	6.772	0,37%	
LIBERAL LIBERTARIO	6.761	0,37%	
POPULAR DE LA RECONSTRUCCION	5.042	0,28%	
ALUVION CIUDADANO	4.953	0,27%	
SUMAR IZQUIERDA DEMOCRATICA	4.930	0,27%	
BANDERA VECINAL	3.845	0,21%	
DEMOCRATA CRISTIANO	2.978	0,16%	
PARTIDO FEDERAL	2.813	0,15%	
EL MOVIMIENTO	2.633	0,14%	
PARTIDO NACIONALISTA CONSTITUCIONAL UNIR	2.094	0,11%	
ACCION CIUDADANA	2.057	0,11%	
MOVIMIENTO INDEPENDIENTE JUSTICIA Y DIGNIDAD	1.566	0,09%	
CONVERGENCIA POPULAR PORTEÑA	0	0%	

Fig. 6.1: Resultados finales oficiales. Fuente: [65]

Telegrama	Código	Partido	Dígito 1	Dígito 2	Dígito 3
1147	905	Partido Convergencia Popular Porteña	1	0	0
1407	905	Partido Convergencia Popular Porteña	3	3	0
1427	905	Partido Convergencia Popular Porteña	1	0	0
1774	905	Partido Convergencia Popular Porteña	8	2	0
1947	905	Partido Convergencia Popular Porteña	2	2	0
2017	905	Partido Convergencia Popular Porteña	2	2	0
2088	905	Partido Convergencia Popular Porteña	1	1	0
2187	905	Partido Convergencia Popular Porteña	1	1	0
2462	905	Partido Convergencia Popular Porteña	2	2	0
2685	905	Partido Convergencia Popular Porteña	5	5	0

Fig. 6.2: El partido Convergencia Popular Porteña obtuvo pocos votos pero no cero. En amarillo el dígito tipeado, en azul el dígito reconocido por OCR

Aquí se presentan las conclusiones alcanzadas durante el presente trabajo:

- Se comparó un mismo dato proveniente de dos fuentes diferentes de formato distinto: dígitos tipeados a partir de formularios en papel e imágenes escaneadas.
- Se comparó el acierto clasificando dígitos manuscritos de telegramas habiendo entrenado una ConvNet con datos más prolijos como es el set de datos abiertos MNIST.
- Se creó un conjunto de datos de dígitos manuscritos obtenidos en condiciones ambientales reales
- De 2151 telegramas se identificó que menos del 1% tenía diferencias entre lo manuscrito y lo transcripto.

El reconocimiento automático de los dígitos manuscritos utilizando una ConvNet nos permitió encontrar los votos del partido Convergencia Popular Porteña 6.2. Si se analizan los Datos Abiertos de Gobierno disponibles en el sitio web del Gobierno de la Ciudad [68] [69], sorprende que el partido político ni siquiera aparece en el

conjunto de datos.

Esto es un ejemplo de cómo se encuentran historias y generar investigaciones a partir de los datos abiertos de gobierno.

El objetivo de este trabajo es encontrar inconsistencias entre imágenes escaneadas y dígitos tipeados a partir de formularios en papel, dada la metodología utilizada este dato debería coincidir en ambas fuentes.

Como trabajo futuro es viable utilizar los datos abiertos de números manuscritos para evaluar métodos de segmentación de números y luego determinar si se obtienen mejores resultados en la clasificación de los dígitos utilizando métodos de minería de datos. Otra propuesta es evaluar las diferencias entre el número clasificado por la ConvNet y el número manuscrito en los telegramas escaneados.


## Apéndice

## A. TELEGRAMAS CON INCONSISTENCIAS

Se presenta a continuación una visualización de los telegramas manuscritos en los que se han encontrado inconsistencias entre lo tipeado por una persona y lo clasificado por la ConvNet. En cada casillero del telegrama se ha incluido con un color distintivo el número que ha tipeado la persona (en color amarillo) y el número determinado por la red neuronal convolucional (en color azul). En aquellos casos en los que coincide lo tipeado con lo clasificado se utiliza color verde.

Detalle de las inconsistencias	
Descripción	Imagen
Error en partido 905 y votos nulos	Fig. A.1
Error en partido 905 y votos nulos	Fig. A.2
Error en partido 905 y votos nulos	Fig. A.3
Error en partido 905 y votos nulos	Fig. A.4
Error en partido 905 y votos nulos	Fig. A.5
Error en partido 905 y votos nulos	Fig. A.6
Error en partido 905 y votos nulos	Fig. A.7
Error en votos nulos	Fig. A.8
Error en partido 905 y votos nulos	Fig. A.9
Error en partido 902, 905 y votos nulos	Fig. A.10
Error en partido 317, 323, 900, 901, 902, 903	Fig. A.11
Partidos 68 al 323 y del 904, 905, 906 y 907 se traza una línea vertical indicando cero votos que confunde a la ConvNet.	Fig. A.12
Error en partido 905 y votos nulos	Fig. A.13
Error en votos nulos	Fig. A.14

NO INTRODUCIR EN LA URNA - ENTREGAR AL EMPLEADO DEL CORREO - ART. 105 Ley N° 19945



**CORREO ARGENTINO**  
CORREO OFICIAL

01 31147X 1 1

**TELEGRAMA DE ESCRUTINIO DE MESA**  
Elecciones Generales del 27 de octubre de 2013  
Art. 105 - Código Electoral Nacional - LEY N° 19945  
**DISTRITO CAPITAL FEDERAL**

COMPLETAR EL FORMULARIO INDISPENSABLE PARA LA COMPENSACION ART. 14 Y 72 DEL CEN

Destinatario: JUNTA ELECTORAL NACIONAL  
Con copia a: DIRECCION NACIONAL ELECTORAL

MESA  
**1147**

Cantidad de electores que han votado:	249
Cantidad de sobres en la urna:	249

Sección	3
Circuito	34


Lista N°	PARTIDOS Y ALIANZAS	SENADORES NACIONALES	DIPUTADOS NACIONALES	DIPUTADOS/AS DE LA LEGISLATURA
167	Partido Autodeterminación y Libertad	7	13	8
501	Alianza Frente para la Victoria	63	59	40
605	Alianza Alternativa Popular			0
602	Alianza UNEN	55	57	55
603	Alianza Unión Pro	78	85	84
505	Alianza Pto. de Izq. y de los Trabajadores	10	11	12
606	Alianza Camino Popular	8	3	2
5	Partido Demócrata Cristiano			0
63	Partido Es Posible			0
68	Partido Nueva Izquierda			1
179	Partido Acción Ciudadana			1
262	Partido El Movimiento			0
204	Partido Mov. indep. Justicia y Dignidad			0
312	Partido Federal			0
317	Partido Nacionalista Constitucional-UNIR			0
323	Partido Movimiento al Socialismo			1
900	Partido Liberal Libertario			0
901	Partido de la Red			1
902	Partido Bandera Vecinal			1
903	Partido Confianza Pública			19
904	Partido Sumar Izquierda Democrática			0
905	Partido Convergencia Popular Porteña			1
906	Partido Aluvión Ciudadano			0
907	Partido Popular de la Reconstrucción			0
	Votos NULOS	0	0	2
	Votos RECURRIDOS	0	0	0
	Votos IMPUGNADOS			
	Votos EN BLANCO	10	14	11
	<b>TOTAL POR COLUMNA</b>	<b>249</b>	<b>249</b>	<b>249</b>

Información indispensable para el cobro de la compensación de Autoridad de Mesa


APELLIDO Y NOMBRE: [Firma]	APELLIDO Y NOMBRE: [Firma]
Documento: 11811521316	Documento: 1179124318
<b>PRESIDENTE</b>	<b>SUPLENTE</b>

Firma, Aclaración, Número de Documento y Número de Lista de los Fiscales

Fig. A.1: Error en partido 905 (escrito: 1, tipeo: 0, convnet: 1) y votos nulos (escrito: 1, tipeo: 0, convnet: 1)



**CORREO ARGENTINO**  
CORREO OFICIAL



01 31407X 1 1

**TELEGRAMA DE ESCRUTINIO DE MESA**  
Elecciones Generales del 27 de octubre de 2013  
Art. 105 - Código Electoral Nacional - LEY N° 19945  
**DISTRITO CAPITAL FEDERAL**

Destinatario: JUNTA ELECTORAL NACIONAL  
Con copia a: DIRECCION NACIONAL ELECTORAL

MESA  
**1407**

Cantidad de electores que han votado:	277
Cantidad de sobres en la urna:	271

Sección	Círculo
3	42

Lista N°	PARTIDOS Y ALIANZAS	SENADORES NACIONALES	DIPUTADOS NACIONALES	DIPUTADOS/AS DE LA LEGISLATURA
187	Partido Autodeterminación y Libertad	8 6 6 10	10	8 8
501	Alianza Frente para la Victoria	8 85 85 79	79	67 67
605	Alianza Alternativa Popular			13 13
502	Alianza UNEN	52 53 53 57	57	50 50
503	Alianza Unión Pro	90 90 75 75	75	78 78
506	Alianza Pto. de Izq. y de los Trabajadores	17 17 21 21	21	21 21
506	Alianza Camino Popular	4 4 2 2	2	2 2
5	Partido Demócrata Cristiano			1 1
63	Partido Es Posible			3 3
68	Partido Nueva Izquierda			1 1
179	Partido Acción Ciudadana			1 1
262	Partido El Movimiento			0 0
264	Partido Mov. Indep. Justicia y Dignidad			1 1
312	Partido Federal			0 0
317	Partido Nacionalista Constitucional-UNIR			0 0
323	Partido Movimiento al Socialismo			1 1
900	Partido Liberal Libertario			2 2
901	Partido de la Red			1 1
902	Partido Bandera Vecinal			0 0
903	Partido Confianza Pública			11 11
904	Partido Sumar Izquierda Democrática			0 0
905	Partido Convergencia Popular Porteña			3 3
906	Partido Aluvión Ciudadano			0 0
907	Partido Popular de la Reconstrucción			0 0
	Votos NULOS	1 1 4 4	4	1 4
	Votos RECURRIDOS	7 0 0 0	0	0 0
	Votos IMPUGNADOS			
	Votos EN BLANCO	20 20 29 29	29	14 14
	<b>TOTAL POR COLUMNA</b>			

Información indispensable para el cobro de la compensación de Autoridad de Mesa

APELLIDO Y NOMBRE [Firma] [Documento]	APELLIDO Y NOMBRE [Firma] [Documento]
PRESIDENTE	SUPLENTE

Firma, Aclaración, Número de Documento y Número de Lista de los Fiscales


503 206444791  
503 37619856

NO INTRODUCIR EN LA URNA - ENTREGAR AL EMPLEADO DEL CORREO - ART. 105 Ley N° 19945

COMPLETAR EL FORMULARIO INDISPENSABLE PARA LA COMPENSACION ART. 14 Y 72 DEL CEN

Fig. A.2: Error en partido 905 (escrito: 3, tipeo: 0, convnet: 3) y votos nulos (escrito: 1, tipeo: 4, convnet: 1)

NO INTRODUCIR EN LA URNA ENTREGAR AL EMPLEADO DEL CORREO ART. 105 LEY N° 19945



**CORREO ARGENTINO**  
CORREO OFICIAL

01 31427X 1 1

**TELEGRAMA DE ESCRUTINIO DE MESA**  
Elecciones Generales del 27 de octubre de 2013  
Art. 105 - Código Electoral Nacional - LEY N° 19945  
**DISTRITO CAPITAL FEDERAL**

COMPLETAR EL FORMULARIO INDISPENSABLE PARA LA COMPENSACION ART. 14 Y 72 DEL CEN

Destinatario: JUNTA ELECTORAL NACIONAL  
Con copia a: DIRECCION NACIONAL ELECTORAL

MESA  
**1427**

Cantidad de electores que han votado: <u>247</u>		Sección	Circuito
Cantidad de sobres en la urna: <u>247</u>		3	42

Lista N°	PARTIDOS Y ALIANZAS	SENADORES NACIONALES	DIPUTADOS NACIONALES	DIPUTADOS/AS DE LA LEGISLATURA
187	Partido Autodeterminación y Libertad	7 7	8 8	8 8
501	Alianza Frente para la Victoria	60 60	56 56	45 45
605	Alianza Alternativa Popular			0
602	Alianza UNEN	55 55	69 69	51 51
503	Alianza Unión Pro	88 88	76 76	75 75
506	Alianza Pte. de Izq. y de los Trabajadores	13 13	18 18	16 16
508	Alianza Camino Popular	6 6	5 5	4 4
5	Partido Demócrata Cristiano			0
63	Partido Es Posible			3 3
58	Partido Nueva Izquierda			0
179	Partido Acción Ciudadana			0
262	Partido El Movimiento			1 1
264	Partido Mov. Indep. Justicia y Dignidad			0
312	Partido Federal			0
317	Partido Nacionalista Constitucional-UNIR			2 2
323	Partido Movimiento al Socialismo			1 1
900	Partido Liberal Libertario			0
901	Partido de la Red			0
902	Partido Bandera Vecinal			23 23
903	Partido Confianza Pública			0
904	Partido Sumar Izquierda Democrática			0
905	Partido Convergencia Popular Porteña			1 1
908	Partido Aluvión Ciudadano			0
907	Partido Popular de la Reconstrucción			0
Votos NULOS		4 4	4 4	4 4 5
Votos RECURRIDOS		0	0	0
Votos IMPUGNADOS				
Votos EN BLANCO		14 14	11 11	9 9 2
<b>TOTAL POR COLUMNA</b>		<b>247</b>	<b>247</b>	<b>247</b>

Información indispensable para el cobro de la compensación de Autoridad de Mesa

<p>APPELLIDO Y NOMBRE</p> <p><u>[Firma]</u></p> <p>Documento: <u>131911141631</u></p> <p><b>PRESIDENTE</b></p>	<p>APPELLIDO Y NOMBRE</p> <p><u>[Firma]</u></p> <p>Documento: <u>114188139109</u></p> <p><b>SULENTE</b></p>
--	---

Firma, Aclaración, Número de Documento y Número de Lista de los Fiscales

[Firma]

36.949.168.1030


[Firma]

34340835 - LISTA 2013

Fig. A.3: Error en partido 905 (escrito: 1, tipeo: 0, convnet: 1) y votos nulos (escrito: 4, tipeo: 5, convnet: 4)



NO INTRODUCIR EN LA URNA - ENTREGAR A EMPLEADO DEL CORREO - ART. 105 Ley N° 19945



01 41774X 1 1

### TELEGRAMA DE ESCRUTINIO DE MESA

Elecciones Generales del 27 de octubre de 2013  
Art. 105 - Código Electoral Nacional - LEY N° 19945  
**DISTRITO CAPITAL FEDERAL**

COMPLETAR EL FORMULARIO INDISPENSABLE PARA LA COMPENSACION - ART. 14 Y 72 DEL CEN

Destinatario: JUNTA ELECTORAL NACIONAL  
Con copia a: DIRECCION NACIONAL ELECTORAL

MESA  
**1774**

Cantidad de electores que han votado:	
Cantidad de sobres en la urna:	

Sección	Círculo
4	53

Lista N°	PARTIDOS Y ALIANZAS	SENADORES NACIONALES	DIPUTADOS NACIONALES	DIPUTADOS/AS DE LA LEGISLATURA
187	Partido Autodeterminación y Libertad	6 6	6 6	6 6
501	Alianza Frente para la Victoria	94 94	40 40	80 72
805	Alianza Alternativa Popular			1 16 16
602	Alianza UNEN	35 35	38 39	39 35 35
603	Alianza Unión Pro	78 78	74 74	74 74
505	Alianza Pte. de Izq. y de los Trabajadores	16 16	16 16	45 15 15
506	Alianza Camino Popular	7 7	7 7	7 7
5	Partido Demócrata Cristiano			2 2
63	Partido Es Posible			0 -
68	Partido Nueva Izquierda			1 1
179	Partido Acción Ciudadana			0 -
262	Partido El Movimiento			0 -
264	Partido Mov. Indep. Justicia y Dignidad			1 1
312	Partido Federal			0 -
317	Partido Nacionalista Constitucional-UNIR			0 -
323	Partido Movimiento al Socialismo			1 1
900	Partido Liberal Libertario			0 -
901	Partido de la Red			1 1 1
902	Partido Bandera Vecinal			0 -
903	Partido Confianza Pública			1 6 6
904	Partido Sumar Izquierda Democrática			0 -
905	Partido Convergencia Popular Porteña			0 2 0
906	Partido Aluvión Ciudadano			0 -
907	Partido Popular de la Reconstrucción			0 -
	Votos NULOS	5 5	5 5	5 5 7
	Votos RECURRIDOS	0 -	0 -	0 -
	Votos IMPUGNADOS	-	-	-
	Votos EN BLANCO	18 18	22 22	12 15 15
	<b>TOTAL POR COLUMNA</b>	<b>259</b>	<b>259</b>	<b>259</b>


Información indispensable para el cobro de la compensación de Autoridad de Mesa

<p>ALLOGNO</p> <p>EL OTRO</p> <p>Apellido y Nombre</p> <p>Firma</p> <p>Documento</p> <p style="text-align: center;"><b>PRESIDENTE</b></p>	<p>Apellido y Nombre</p> <p>Firma</p> <p>Documento</p> <p style="text-align: center;"><b>SUPLENTE</b></p>
---	---

Firma, Aclaración, Número de Documento y Número de Lista de los Fiscales

Fig. A.4: Error en partido 905 (escrito: 2, tipeo: 0, convnet: 2) y votos nulos (escrito: 5, tipeo: 7, convnet: 5)

NO INTRODUCIR EN LA URNA. ENTREGAR AL EMPLEADO DEL CORREO ART. 105 Ley N° 19945



**CORREO ARGENTINO**  
CORREO OFICIAL

01 41947X 1 1

**TELEGRAMA DE ESCRUTINIO DE MESA**  
Elecciones Generales del 27 de octubre de 2013  
Art. 105 - Código Electoral Nacional - LEY N° 19945  
**DISTRITO CAPITAL FEDERAL**

COMPLETAR EL FORMULARIO INDISPENSABLE PARA LA COMPENSACION - ART. 14 Y 72 DEL GEN

Destinatario: JUNTA ELECTORAL NACIONAL  
Con copia a: DIRECCION NACIONAL ELECTORAL

MESA  
**1947**


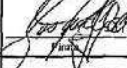
Cantidad de electores que han votado: **277**

Cantidad de sobres en la urna: **277**

Sección	Círculo
4	57

Lista N°	PARTIDOS Y ALIANZAS	SENADORES NACIONALES	DIPUTADOS NACIONALES	DIPUTADOS/AS DE LA LEGISLATURA
187	Partido Autodeterminación y Libertad	362 107	8	0
501	Alianza Frente para la Victoria	84	821 84	0
605	Alianza Alternativa Popular			86 86
502	Alianza UNEN	59 8	59 63	62
503	Alianza Unión Pro	78 78	70	69 59 54
505	Alianza Pte. de Izq. y de los Trabajadores	13	113 13	74 80 14
506	Alianza Camino Popular	5	5	4 4
5	Partido Demócrata Cristiano			2 2
63	Partido Es Posible			1 1
68	Partido Nueva Izquierda			2 2
179	Partido Acción Ciudadana			0
262	Partido El Movimiento			0
264	Partido Mov. Indep. Justicia y Dignidad			1 1
312	Partido Federal			0
317	Partido Nacionalista Constitucional-UNR			1 1
323	Partido Movimiento al Socialismo			2 2
900	Partido Liberal Libertario			2 2
901	Partido de la Red			1 0
902	Partido Bandera Vecinal			1 0
903	Partido Confianza Pública			8 8
904	Partido Sumar Izquierda Democrática			0
905	Partido Convergencia Popular Porteña			2 0
906	Partido Alusión Ciudadano			0
907	Partido Popular de la Reconstrucción			2 0 1
	Votos NULOS	5 5	5 5	5 5 7
	Votos RECURRIDOS	0	0	0
	Votos IMPUGNADOS			
	Votos EN BLANCO	15 75 15	13 47 17 8	8
	<b>TOTAL POR COLUMNA</b>	<b>277</b>	<b>277</b>	<b>277</b>

Información indispensable para el cobro de la compensación de Autoridad de Mesa


NESTOR A. MARADETI A JERRO APELLIDO Y NOMBRE	GUSTAVO BUCCIANI APELLIDO Y NOMBRE
 PRESIDENTE	 SUPLENTE

Firma, Aclaración, Número de Documento y Número de Lista de los Fiscales

Fig. A.5: Error en partido 905 (escrito: 2, tipeo: 0, convnet: 2) y votos nulos (escrito: 5, tipeo: 7, convnet: 5)

NO INTRODUCIR EN LA URNA - ENTREGAR AL EMPLEADO DEL CORREO - ART. 105 Ley N° 19945

COMPLETAR EL FORMULARIO INDISPENSABLE PARA LA COMPENSACION - ART. 14 Y 72 DEL CEN



01 52017X 1 1

### TELEGRAMA DE ESCRUTINIO DE MESA

Elecciones Generales del 27 de octubre de 2013  
Art. 105 - Código Electoral Nacional - LEY N° 19945  
**DISTRITO CAPITAL FEDERAL**

Destinatario: JUNTA ELECTORAL NACIONAL  
Con copia a: DIRECCION NACIONAL ELECTORAL

Cantidad de electores que han votado:	264
Cantidad de sobres en la urna:	264

MESA	
2017	

Sección	Circuito
5	59

Lista N°	PARTIDOS Y ALIANZAS	SENADORES NACIONALES	DIPUTADOS NACIONALES	DIPUTADOS/AS DE LA LEGISLATURA
187	Partido Autodeterminación y Libertad	6	9	9
601	Alianza Frente para la Victoria	85	81	67
605	Alianza Alternativa Popular	13	13	13
602	Alianza UNEN	55	63	48
503	Alianza Unión Pro	79	65	64
505	Alianza Fte. de Izq. y de los Trabajadores	9	14	11
606	Alianza Camino Popular	8	7	5
5	Partido Demócrata Cristiano	0	0	0
63	Partido Es Posible	0	0	0
68	Partido Nueva Izquierda	1	1	1
179	Partido Acción Ciudadana	1	1	1
262	Partido El Movimiento	0	0	0
264	Partido Mov. indep. Justicia y Dignidad	1	1	1
312	Partido Federal	2	2	2
317	Partido Nacionalista Constitucional-UNIR	0	0	0
323	Partido Movimiento al Socialismo	0	0	0
900	Partido Liberal Libertario	0	0	0
901	Partido de la Red	35	5	5
902	Partido Bandera Vecinal	2	2	2
903	Partido Confianza Pública	18	18	18
904	Partido Sumar Izquierda Democrática	0	0	0
905	Partido Convergencia Popular Porteña	2	2	0
906	Partido Aluvión Ciudadano	1	1	1
907	Partido Popular de la Reconstrucción	0	0	0
Votos NULOS		5	5	7
Votos RECURRIDOS		0	0	0
Votos IMPUGNADOS		-	-	-
Votos EN BLANCO		16	20	9
<b>TOTAL POR COLUMNA</b>		<b>264</b>	<b>264</b>	<b>264</b>

Información indispensable para el cobro de la compensación de Autoridad de Mesa


EVANGELISTA PAULIA APELLIDO Y NOMBRE [Firma] [2631116814] Documento	CIRIACIANTTANEO APELLIDO Y NOMBRE [Firma] [313892858] Documento
<b>PRESIDENTE</b>	<b>SULENTE</b>

Firma, Aclaración, Número de Documento y Número de Lista de los Fiscales

NOMBRE Y APELLIDO: [Handwritten] / NOMBRE Y APELLIDO: [Handwritten]  
 DOCUMENTO: [Handwritten] / DOCUMENTO: [Handwritten]

Fig. A.6: Error en partido 905 (escrito: 2, tipeo: 0, convnet: 2) y votos nulos (escrito: 5, tipeo: 7, convnet: 5)

NO INTRODUCIR EN LA URNA - ENTREGAR AL EMPLEADO DEL CORREO - ART. 105 Ley N° 19945



**CORREO ARGENTINO**  
CORREO OFICIAL

01 52088X 1 1

**TELEGRAMA DE ESCRUTINIO DE MESA**  
Elecciones Generales del 27 de octubre de 2013  
Art. 105 - Código Electoral Nacional - LEY N° 19945  
**DISTRITO CAPITAL FEDERAL**

COMPLETAR EL FORMULARIO INDISPENSABLE PARA LA COMPENSACION - ART. 14 Y 72 DEL CEN

Destinatario: JUNTA ELECTORAL NACIONAL  
Con copia a: DIRECCION NACIONAL ELECTORAL

MESA  
**2088**

Cantidad de electores que han votado:	284	Sección	5	Circuito	61
Cantidad de sobres en la urna:	284				

Lista N°	PARTIDOS Y ALIANZAS	SENADORES NACIONALES	DIPUTADOS NACIONALES	DIPUTADOS/AS DE LA LEGISLATURA
187	Partido Autodeterminación y Libertad	6	17	127
501	Alianza Frente para la Victoria	81	56	42
605	Alianza Alternativa Popular	61	56	42
502	Alianza UNEN	4	92	75
503	Alianza Unión Pro	97	30	80
505	Alianza Pte. de Izq. y de los Trabajadores	15	18	15
506	Alianza Camino Popular	8	4	9
5	Partido Demócrata Cristiano			0
63	Partido Es Posible			0
68	Partido Nueva Izquierda			0
179	Partido Acción Ciudadana			0
262	Partido El Movimiento			0
264	Partido Mov. Indep. Justicia y Dignidad			0
312	Partido Federal			0
317	Partido Nacionalista Constitucional-UNIR			1
323	Partido Movimiento al Socialismo			1
900	Partido Liberal Libertario			0
901	Partido de la Red			7
902	Partido Bandera Vecinal			2
903	Partido Confianza Pública			14
904	Partido Sumar Izquierda Democrática			0
905	Partido Convergencia Popular Porteña			1
906	Partido Aluvión Ciudadano			0
907	Partido Popular de la Reconstrucción			0
	Votos NULOS	0	0	0
	Votos RECURRIDOS	0	0	0
	Votos IMPUGNADOS			
	Votos EN BLANCO	9	17	13
	<b>TOTAL POR COLUMNA</b>	<b>284</b>	<b>284</b>	<b>284</b>


Información indispensable para el cobro de la compensación de Autoridad de Mesa

<p>11132734131</p> <p>Apellido y Nombre</p> <p>Firma</p> <p><b>PRÉSIDENTE</b></p>	<p>310456426</p> <p>Apellido y Nombre</p> <p>Firma</p> <p><b>SULENTE</b></p>
---	--

Firma, Aclaración, Número de Documento y Número de Lista de los Fiscales

Fig. A.7: Error en partido 905 (escrito: 1, tipeo: 0, convnet: 1) y votos nulos (escrito: 0, tipeo: 1, convnet: 0)

NO INTRODUCIR EN LA URNA - ENTREGAR AL EMPLEADO DEL CORREO - ART. 105 Ley N° 19945



01 52170X 1 1

### TELEGRAMA DE ESCRUTINIO DE MESA

Elecciones Generales del 27 de octubre de 2013  
Art. 105 - Código Electoral Nacional - LEY N° 19945  
DISTRITO CAPITAL FEDERAL

COMPLETAR EL FORMULARIO INDISPENSABLE PARA LA COMPENSACION - ART. 14 Y 72 DEL GEN.

Destinatario: JUNTA ELECTORAL NACIONAL  
Con copia a: DIRECCION NACIONAL ELECTORAL


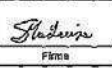
MESA  
**2170**

Cantidad de electores que han votado:	281
Cantidad de sobres en la urna:	281

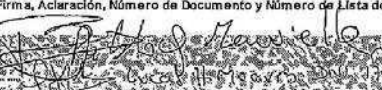

Sección	5
Círculo	63

Lista N°	PARTIDOS Y ALIANZAS	SENADORES NACIONALES	DIPUTADOS NACIONALES	DIPUTADOS/AS DE LA LEGISLATURA
187	Partido Autodeterminación y Libertad	4 4	8 8	4 4
501	Alianza Frente para la Victoria	70 70	63 63	53 53
605	Alianza Alternativa Popular			13 13
502	Alianza UNEN	69 69	71 71	63 53 63
503	Alianza Unión Pro	1013 103	103 92	88 88
505	Alianza Fte. de Izq. y de los Trabajadores	13 13	18 18	13 83 13
506	Alianza Camino Popular	5 5	8 8	4 6 4
5	Partido Demócrata Cristiano			1 2 1
63	Partido Es Posible			3 3
68	Partido Nueva Izquierda			2 3 2
179	Partido Acción Ciudadana			1 1
262	Partido El Movimiento			0 0
264	Partido Mov. Indep. Justicia y Dignidad			0 0
312	Partido Federal			0 0
317	Partido Nacionalista Constitucional-UNIR			0 0
323	Partido Movimiento al Socialismo			2 2
900	Partido Liberal Libertario			2 2 9 2
901	Partido de la Red			2 2
902	Partido Bandera Vecinal			1 1
903	Partido Confianza Pública			16 85 16
904	Partido Sumar Izquierda Democrática			3 3
905	Partido Convergencia Popular Porteña			0 0
906	Partido Aluvión Ciudadano			0 0
907	Partido Popular de la Reconstrucción			2 2
Votos NULOS		0 0	0 0	1 1 0
Votos RECURRIDOS		0 0	0 0	0 0
Votos IMPUGNADOS		0 0	0 0	0 0
Votos EN BLANCO		14 17	19 289 19	4 74 7
<b>TOTAL POR COLUMNA</b>		<b>281</b>	<b>281</b>	<b>281</b>

Información indispensable para el cobro de la compensación de Autoridad de Mesa


CARRERENO CELIA EUSA APELLIDO Y NOMBRE	CASTILLO MARIA APELLIDO Y NOMBRE
 PRESIDENTE	 SUPLENTE

Firma, Aclaración, Número de Documento y Número de Lista de los Fiscales

(\*) salvado al don so =

Fig. A.8: Error en votos nulos (escrito: 1, tipeo: 0, convnet: 1)


01 52187X 1 1  
**TELEGRAMA DE ESCRUTINIO DE MESA**  
 Elecciones Generales del 27 de octubre de 2013  
 Art. 105 - Código Electoral Nacional - LEY N° 19945  
**DISTRITO CAPITAL FEDERAL**



Destinatario: JUNTA ELECTORAL NACIONAL  
 Con copia a: DIRECCION NACIONAL ELECTORAL

MESA  
**2187**

Cantidad de electores que han votado: 249	Sección: 5
Cantidad de sobres en la urna: 249	Circuito: 63

Lista N°	PARTIDOS Y ALIANZAS	SENADORES NACIONALES	DIPUTADOS NACIONALES	DIPUTADOS/AS DE LA LEGISLATURA
187	Partido Autodeterminación y Libertad	5	12	5
501	Allianza Frente para la Victoria	61	55	48
606	Allianza Alternativa Popular			8
602	Allianza UNEN	60	69	59
603	Allianza Unión Pro	79	69	70
605	Allianza Pte. de Izq. y de los Trabajadores	15	18	16
506	Allianza Camino Popular	5	4	4
5	Partido Demócrata Cristiano			0
63	Partido Es Posible			3
68	Partido Nueva Izquierda			1
179	Partido Acción Ciudadana			0
262	Partido El Movimiento			1
264	Partido Mov. Indep. Justicia y Dignidad			2
312	Partido Federal			0
317	Partido Nacionalista Constitucional-UNIR			0
323	Partido Movimiento al Socialismo			1
900	Partido Liberal Libertario			1
901	Partido de la Red			1
902	Partido Bandera Vecinal			0
903	Partido Confianza Pública			15
904	Partido Sumar Izquierda Democrática			0
905	Partido Convergencia Popular Porteña			1
906	Partido Aluvión Ciudadano			1
907	Partido Popular de la Reconstrucción			2
	Votos NULOS	0	1	4
	Votos RECURRIDOS	0	0	0
	Votos IMPUGNADOS			
	Votos EN BLANCO	24	21	6
	<b>TOTAL POR COLUMNA</b>	<b>249</b>	<b>249</b>	<b>249</b>


Información indispensable para el cobro de la compensación de Autoridad de Mesa

EBRRIAND EABRIANAI SHUDANAI APELLIDO Y NOMBRE  Firma 261181011917 Documento	JALIFEE PLOMINA JEISICIA APELLIDO Y NOMBRE  Firma 2181810112146 Documento
<b>PRESIDENTE</b>	<b>SUPLENTE</b>

Firma, Aclaración, Número de Documento y Número de Lista de los Fiscales

Fig. A.9: Error en partido 905 (escrito: 1, tipeo: 0, convnet: 1) y votos nulos (escrito: 4, tipeo: 5, convnet: 4)

NO INTRODUCIREN LA URNA - ENTREGAR AL EMPLEADO DEL CORREO - ART. 105 Ley N° 19945



01 62462X 1 1

### TELEGRAMA DE ESCRUTINIO DE MESA

Elecciones Generales del 27 de octubre de 2013  
Art. 105 - Código Electoral Nacional - LEY N° 19945  
**DISTRITO CAPITAL FEDERAL**

COMPLETAR EL FORMULARIO INDISPENSABLE PARA LA COMPENSACION - ART. 14 Y 12 DEL GEN

Destinatario: JUNTA ELECTORAL NACIONAL  
Con copia a: DIRECCION NACIONAL ELECTORAL

MESA  
**2462**

Cantidad de electores que han votado: <b>273</b>		Sección <b>6</b>	Circuito <b>68</b>
Cantidad de sobres en la urna: <b>273</b>			


Lista N°	PARTIDOS Y ALIANZAS	SENADORES NACIONALES	DIPUTADOS NACIONALES	DIPUTADOS/AS DE LA LEGISLATURA
187	Partido Autodeterminación y Libertad	40 80 10	11 11 11	10
801	Allianza Frente para la Victoria	57 57	56 56	42 512 42
805	Allianza Alternativa Popular			12 72 12
802	Allianza UNEN	83 83 98	98	74 34 74
503	Allianza Unión Pro	79 9 79	63	68
505	Allianza Fle. de Izq. y de los Trabajadores	19 49 19	20 20	20 2 20
506	Allianza Camino Popular	8 8	8 3 8	7 2 7
5	Partido Demócrata Cristiano			0
83	Partido Es Posible			0
68	Partido Nueva Izquierda			3 33 3
179	Partido Acción Ciudadana			0
262	Partido El Movimiento			0
264	Partido Mov. Indep. Justicia y Dignidad			0
312	Partido Federal			0
317	Partido Nacionalista Constitucional-UNIR			1 34 1
323	Partido Movimiento al Socialismo			2 2
900	Partido Liberal Libertario			0
901	Partido de la Red			3 3
902	Partido Bandera Vecinal			917 917 17
903	Partido Confianza Pública			1 1
904	Partido Sumar Izquierda Democrática			0
905	Partido Convergencia Popular Porteña			2 2 0
906	Partido Aluvión Ciudadano			0
907	Partido Popular de la Reconstrucción			1 11 1
	Votos NULOS	3 03	3 03	3 5
	Votos RECURRIDOS	1 1	1 1	1 4 1
	Votos IMPUGNADOS			
	Votos EN BLANCO	13 13	13 13	7 7
	<b>TOTAL POR COLUMNA</b>	<b>273</b>	<b>273</b>	<b>273</b>

Información indispensable para el cobro de la compensación de Autoridad de Mesa

PRESIDENTE FIRMAS: <i>[Firma]</i> / <i>[Firma]</i> APELLIDO Y NOMBRE: <b>FERRERUCIA</b> / <b>MAURICIO</b> Documento: <b>23164164179</b> / <b>316217188919</b>	SUPLENTE FIRMAS: <i>[Firma]</i> / <i>[Firma]</i> APELLIDO Y NOMBRE: <b>MAURICIO</b> / <b>MAURICIO</b> Documento: <b>316217188919</b>
--	---

Firma, Aclaración, Número de Documento y Número de Lista de los Fiscales

Fig. A.10: Error en partido 902 (escrito: 917, tipeo: 17, convnet: 917), partido 905 (escrito: 2, tipeo: 0, convnet: 2) y votos nulos (escrito: 3, tipeo: 5, convnet: 3)



01 62528X 1 1

### TELEGRAMA DE ESCRUTINIO DE MESA

Elecciones Generales del 27 de octubre de 2013  
Art. 105 - Código Electoral Nacional - LEY N° 19945  
DISTRITO CAPITAL FEDERAL

Destinatario: JUNTA ELECTORAL NACIONAL  
Con copia a: DIRECCION NACIONAL ELECTORAL

MESA  
**2528**

Cantidad de electores que han votado:	230
Cantidad de sobres en la urna:	230

Sección	6
Circuito	69

Lista N°	PARTIDOS Y ALIANZAS	SENADORES NACIONALES	DIPUTADOS NACIONALES	DIPUTADOS/AS DE LA LEGISLATURA
187	Partido Autodeterminación y Libertad	3	16	12
501	Alianza Frente para la Victoria	44	41	30
605	Alianza Alternativa Popular			12
602	Alianza UNEN	95	103	71
503	Alianza Unión Pro	85	75	41
506	Alianza Fle. de Izq. y de los Trabajadores	12	10	7
506	Alianza Camino Popular	5	6	4
5	Partido Demócrata Cristiano			0
63	Partido Es Posible			1
68	Partido Nueva Izquierda			1
179	Partido Acción Ciudadana			0
262	Partido El Movimiento			2
264	Partido Mov. Indep. Justicia y Dignidad			0
312	Partido Federal			0
317	Partido Nacionalista Constitucional-UNIR			0
323	Partido Movimiento al Socialismo			1
900	Partido Liberal Libertario			0
901	Partido de la Red			3
902	Partido Bandera Vecinal			0
903	Partido Confianza Pública			27
904	Partido Sumar Izquierda Democrática			0
905	Partido Convergencia Popular Porteña			0
908	Partido Aluvión Ciudadano			0
907	Partido Popular de la Reconstrucción			0
Votos NULOS		4	4	4
Votos RECURRIDOS		0	0	0
Votos IMPUGNADOS				
Votos EN BLANCO		16	15	6
TOTAL POR COLUMNA		230	230	230

Información indispensable para el cobro de la compensación de Autoridad de Mesa

APELLIDO Y NOMBRE: <u>Suñer</u> Documento: <u>21374113</u>	APELLIDO Y NOMBRE: <u>R. 19819593</u> Documento: <u>19819593</u>
PRESIDENTE	SUPLENTE


Firma, Aclaración, Número de Documento y Número de Lista de los Fiscales

NO INTRODUCIR EN LA URNA ENTREGAR AL EMPLEADO DEL CORREO ART. 105 Ley N° 19945

COMPLETAR EL FORMULARIO INDISPENSABLE PARA LA COMPENSACION ART. 14 Y 72 DEL CEN

Fig. A.11: Error en partido 317 (escrito: 0, tipeo: 1, convnet: 0), partido 323 (escrito: 1, tipeo: 0, convnet: 1), partido 900 (escrito: 0, tipeo: 3, convnet: 3), partido 901 (escrito: 3, tipeo: 0, convnet: 3), partido 902 (escrito: 0, tipeo: 27, convnet: 0), partido 903 (escrito: 27, tipeo: 0, convnet: 27)





01 62634X 1 1

### TELEGRAMA DE ESCRUTINIO DE MESA

Elecciones Generales del 27 de octubre de 2013  
Art. 105 - Código Electoral Nacional - LEY N° 19945  
DISTRITO CAPITAL FEDERAL

Destinatario: JUNTA ELECTORAL NACIONAL  
Con copia a: DIRECCION NACIONAL ELECTORAL

MESA  
**2634**

Cantidad de electores que han votado:	282
Cantidad de sobres en la urna:	282

Sección	Circuito
6	71


Lista N°	PARTIDOS Y ALIANZAS	SENADORES NACIONALES	DIPUTADOS NACIONALES	DIPUTADOS/AS DE LA LEGISLATURA
187	Partido Autodeterminación y Libertad	9	13	11
501	Alianza Frente para la Victoria	97	47	35
605	Alianza Alternativa Popular			9
502	Alianza UNEN	78	98	67
503	Alianza Unión Pro	106	89	91
505	Alianza Fle. de Izq. y de los Trabajadores	16	17	17
506	Alianza Camino Popular	11	11	09
5	Partido Demócrata Cristiano			1
63	Partido Es Posible			3
68	Partido Nueva Izquierda			0
179	Partido Acción Ciudadana			0
262	Partido El Movimiento			0
264	Partido Mov. Indep. Justicia y Dignidad			0
312	Partido Federal			0
317	Partido Nacionalista Constitucional-UNIR			0
323	Partido Movimiento al Socialismo			0
900	Partido Liberal Libertario			3
901	Partido de la Red			1
902	Partido Bandera Vocinal			0
903	Partido Confianza Pública			20
904	Partido Sumar Izquierda Democrática			0
905	Partido Convergencia Popular Porteña			0
906	Partido Aluvión Ciudadano			0
907	Partido Popular de la Reconstrucción			0
Votos NULOS		1	1	1
Votos RECURRIDOS		1	1	1
Votos IMPUGNADOS				0
Votos EN BLANCO		12	13	13
<b>TOTAL POR COLUMNA</b>		<b>282</b>	<b>282</b>	<b>980</b>

Información indispensable para el cobro de la compensación de Autoridad de Mesa

APELLIDO Y NOMBRE: [Firma]	APELLIDO Y NOMBRE: [Firma]
Firma: [Firma] Documento: [Número]	Firma: [Firma] Documento: [Número]
<b>PRESIDENTE</b>	<b>SUPLENTE</b>

Firma, Aclaración, Número de Documento y Número de Lista de los Fiscales

Fig. A.12: Partidos 68 al 323 y del 904, 905, 906 y 907 se traza una línea vertical indicando cero votos que confunde a la ConvNet.



**CORREO ARGENTINO**  
CORREO OFICIAL

01 62685X 1 1

**TELEGRAMA DE ESCRUTINIO DE MESA**  
Elecciones Generales del 27 de octubre de 2013  
Art. 105 - Código Electoral Nacional - LEY N° 19945  
**DISTRITO CAPITAL FEDERAL**

Destinatario: JUNTA ELECTORAL NACIONAL  
Con copia a: DIRECCION NACIONAL ELECTORAL

MESA  
**2685**

Sección	Circuito
6	71

Cantidad de electores que han votado:	
Cantidad de sobres en la urna:	

Lista N°	PARTIDOS Y ALIANZAS	SENADORES NACIONALES	DIPUTADOS NACIONALES	DIPUTADOS/AS DE LA LEGISLATURA
187	Partido Autodeterminación y Libertad	9	13	11
501	Allianza Frente para la Victoria	51	43	38
605	Allianza Alternativa Popular			5
502	Allianza UNEN	91	101	80
503	Allianza Unión Pro	94	84	86
506	Allianza Fte. de Izq. y de los Trabajadores	7	11	9
506	Allianza Camino Popular	9	7	6
5	Partido Demócrata Cristiano			0
63	Partido Es Posible			2
68	Partido Nueva Izquierda			1
179	Partido Acción Ciudadana			0
262	Partido El Movimiento			1
264	Partido Mov. Indep. Justicia y Dignidad			0
312	Partido Federal			0
317	Partido Nacionalista Constitucional-UNIR			0
323	Partido Movimiento al Socialismo			0
900	Partido Liberal Libertario			0
901	Partido de la Red			0
902	Partido Bandera Vecinal			1
903	Partido Confianza Pública			18
904	Partido Sumar Izquierda Democrática			0
905	Partido Convergencia Popular Porteña			5
906	Partido Aluvión Ciudadano			0
907	Partido Popular de la Reconstrucción			0
Votos NULOS		3	2	3
Votos RECURRIDOS		0	0	0
Votos IMPUGNADOS		-	-	-
Votos EN BLANCO		7	2	3
<b>TOTAL POR COLUMNA</b>				

Información indispensable para el cobro de la compensación de Autoridad de Mesa

<p>MARTINEZ KRUIS</p> <p>AMTOMILO</p> <p style="text-align: center;">APELLIDO Y NOMBRE</p> <p style="text-align: center;">Firma:  Documento: 22744033</p> <p style="text-align: center;"><b>PRESIDENTE</b></p>	<p style="text-align: center;">APELLIDO Y NOMBRE</p> <p style="text-align: center;">Firma: _____ Documento: _____</p> <p style="text-align: center;"><b>SULENTE</b></p>
--	---


Firma, Aclaración, Número de Documento y Número de Lista de los Fiscales

Verónica DOCCALO Vecinal 1503 39.184.935

NO INTRODUCIR EN LA URNA - ENTREGAR AL EMPLEADO DEL CORREO - ART. 105 Ley N° 19945

COMPLETAR EL FORMULARIO INDISPENSABLE PARA LA COMPENSACION - ART. 14 Y 72 DEL CEN

Fig. A.13: Error en partido 905 (escrito: 5, tipeo: 0, convnet: 5) y votos nulos (escrito: 3, tipeo: 8, convnet: 3)



01 62688X 1 1

### TELEGRAMA DE ESCRUTINIO DE MESA

Elecciones Generales del 27 de octubre de 2013  
Art. 105 - Código Electoral Nacional - LEY N° 19945  
**DISTRITO CAPITAL FEDERAL**

Destinatario: JUNTA ELECTORAL NACIONAL  
Con copia a: DIRECCION NACIONAL ELECTORAL

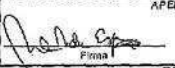

MESA  
**2688**

Cantidad de electores que han votado:	269
Cantidad de sobres en la urna:	269

Sección	Circuito
6	71

Lista N°	PARTIDOS Y ALIANZAS	SENADORES NACIONALES	DIPUTADOS NACIONALES	DIPUTADOS/AS DE LA LEGISLATURA
187	Partido Autodeterminación y Libertad	3	3	6
501	Alianza Frente para la Victoria	42	42	31
606	Alianza Alternativa Popular			11
502	Alianza UNEN	86	86	78
503	Alianza Unión Pro	112	112	89
506	Alianza Pte. de Izq. y de los Trabajadores	11	11	12
506	Alianza Camino Popular	4	4	3
5	Partido Demócrata Cristiano			1
63	Partido Es Posible			2
68	Partido Nueva Izquierda			1
179	Partido Acción Ciudadana			0
262	Partido El Movimiento			0
264	Partido Mov. Indep. Justicia y Dignidad			0
312	Partido Federal			0
317	Partido Nacionalista Constitucional-UNIR			0
323	Partido Movimiento al Socialismo			0
900	Partido Liberal Libertario			1
901	Partido de la Red			1
902	Partido Bandera Vecinal			0
903	Partido Confianza Pública			17
904	Partido Sumar Izquierda Democrática			0
905	Partido Convergencia Popular Porteña			0
906	Partido Aluvión Ciudadano			1
907	Partido Popular de la Reconstrucción			1
	Votos NULOS	0	0	0
	Votos RECURRIDOS	0	0	0
	Votos IMPUGNADOS			
	Votos EN BLANCO	12	12	12
	<b>TOTAL POR COLUMNA</b>	<b>269</b>	<b>269</b>	<b>269</b>

Información indispensable para el cobro de la compensación de Autoridad de Mesa

MARTHA CRISTINA MARTINEZ APELLIDO Y NOMBRE  Firma Documento: 616191415119191	MENENDEZ RAMIREZ APELLIDO Y NOMBRE  Firma Documento: 2311462319
<b>PRESIDENTE</b>	<b>SUPLENTE</b>

Firma, Aclaración, Número de Documento y Número de Lista de los Fiscales

NO INTRODUCIR EN LA URNA ENTREGAR AL EMPLEADO DEL CORREO ARGENTINO

COMPLETAR EL FORMULARIO INDISPENSABLE PARA LA COMPENSACION - ART. 14 Y 72 DEL CEN

Fig. A.14: Error en votos nulos (escrito: 1, tipeo: 0, convnet: 2)

## B. CODIGO FUENTE CONVNET

Este código fuente y los archivos necesarios se encuentran disponibles en Github <sup>1</sup>.

*Listing B.1: ConvNet en Python*

```
#!/usr/bin/env python
"""
Red neuronal convolucional para clasificar conjunto de datos
_MNIST
"""

from __future__ import print_function

import sys
import os
import time
import random
import numpy as np
import theano
import theano.tensor as T

import lasagne
import matplotlib.cm as cm
import matplotlib.pyplot as plt

# Cargar el conjunto de datos MNIST
def load_dataset(training_set, training_labels):
```

---

<sup>1</sup> <https://github.com/walt3rminer/datamining>

---

```
# We then define functions for loading MNIST images and  
labels.  
# For convenience, they also download the requested  
files if needed.
```

```
import gzip
```

```
def load_experimental_images(filename):
```

```
    f = open(filename, "rb")
```

```
    data = np.frombuffer(f.read(), np.uint8)
```

```
    data = data.reshape(-1, 1, 24, 24)
```

```
    #~ return np.uint8(data)
```

```
    return data / np.float32(255)
```

```
def load_experimental_labels(filename):
```

```
    with open(filename) as f:
```

```
        data = f.read().splitlines()
```

```
    data=np.array(data, dtype=np.uint8)
```

```
    return data
```

```
X_train = load_experimental_images(training_set)
```

```
y_train = load_experimental_labels(training_labels)
```

```
X_test = load_experimental_images('tes1_images.csv')
```

```
y_test = load_experimental_labels('tes1_labels.csv')
```

```
return X_train, y_train, X_test, y_test
```

```
# Build the neural network model
```

---

```
# This script supports three types of models. For each one,  
we define a  
# function that takes a Theano variable representing the  
input and returns  
# the output layer of a neural network model built in  
Lasagne.
```

```
def build_cnn(input_var=None):  
    # Capa de entrada  
    network = lasagne.layers.InputLayer(shape=(None, 1, 24,  
        24), input_var=input_var)  
    # Convolution Layer 6 5x5 kernels  
    network = lasagne.layers.Conv2DLayer(  
        network, num_filters=6, pad=2, stride=1,  
        filter_size=(5, 5),  
        nonlinearity=lasagne.nonlinearities.relu,  
        W=lasagne.init.GlorotUniform(gain='relu'))  
  
    # Max-pooling layer  
    network = lasagne.layers.MaxPool2DLayer(network, stride  
        =2, pool_size=(2, 2))  
    # Convolution Layer con 16 5x5 kernels  
    network = lasagne.layers.Conv2DLayer(  
        network, num_filters=16, pad=2, stride=1,  
        filter_size=(5, 5),  
        nonlinearity=lasagne.nonlinearities.relu)  
    # Max-pooling layer  
    network = lasagne.layers.MaxPool2DLayer(network, stride  
        =2, pool_size=(2, 2))
```

---

```
# Ultima capa
network = lasagne.layers.DenseLayer(
    lasagne.layers.dropout(network, p=.5),
    num_units=10,
    nonlinearity=lasagne.nonlinearities.softmax)

return network

# Batch iterator
# This is just a simple helper function iterating over
training data in
# mini-batches of a particular size, optionally in random
order. It assumes
# data is available as numpy arrays. For big datasets, you
could load numpy
# arrays as memory-mapped files (np.load(..., mmap_mode='r')
) , or write your
# own custom data iteration function. For small datasets,
you can also copy
# them to GPU at once for slightly improved performance.
This would involve
# several changes in the main program, though, and is not
demonstrated here.

def iterate_minibatches(inputs, targets, batchsize, shuffle=
    False):
    assert len(inputs) == len(targets)
    if shuffle:
```

---

```
indices = np.arange(len(inputs))
np.random.shuffle(indices)
for start_idx in range(0, len(inputs) - batchsize + 1,
batchsize):
    if shuffle:
        excerpt = indices[start_idx:start_idx +
batchsize]
    else:
        excerpt = slice(start_idx, start_idx + batchsize
)
    yield inputs[excerpt], targets[excerpt]

# Main program
# Everything else will be handled in our main program now.
    We could pull out
# more functions to better separate the code, but it wouldn't
    make it any
# easier to read.
def main(batch_size=100,num_epochs=20):
    # Load the dataset
    X_train, y_train, X_test, y_test = load_dataset('
exp1_images.csv', 'exp1_labels.csv')
    X_train[ X_train > 0.4 ] = 1
    X_train[ X_train <= 0.4 ] = 0
    X_test[ X_test > 0.4 ] = 1
    X_test[ X_test <= 0.4 ] = 0
    X_train = (np.uint8(X_train))
    X_test = (np.uint8(X_test))
```



---

```
training_samples = X_train.shape[0]
test_samples = X_test.shape[0]
# Prepare Theano variables for inputs and targets
input_var = T.tensor4('inputs')
target_var = T.ivector('targets')
train_size = 60000
# Create neural network model (depending on first
  command line parameter)
print("Building_model_and_compiling_functions...")
network = build_cnn(input_var)

# Create a loss expression for training, i.e., a scalar
  objective we want
# to minimize (for our multi-class problem, it is the
  cross-entropy loss):
prediction = lasagne.layers.get_output(network)
loss = lasagne.objectives.categorical_crossentropy(
  prediction, target_var)
loss = loss.mean()
# We could add some weight decay as well here, see
  lasagne.regularization.
# Create update expressions for training, i.e., how to
  modify the
# parameters at each training step. Here, we'll use
  Stochastic Gradient
# Descent (SGD) with Nesterov momentum, but Lasagne
  offers plenty more.
params = lasagne.layers.get_all_params(network,
  trainable=True)
```

---

```
updates = lasagne.updates.nesterov_momentum(loss , params
    , learning_rate=0.03, momentum=0.9)

# Create a loss expression for validation/testing. The
  crucial difference
# here is that we do a deterministic forward pass
  through the network,
# disabling dropout layers.
test_prediction = lasagne.layers.get_output(network ,
    deterministic=True)
test_loss = lasagne.objectives.categorical_crossentropy(
    test_prediction , target_var)
test_loss = test_loss.mean()
# As a bonus, also create an expression for the
  classification accuracy:
train_acc = T.mean(T.eq(T.argmax(prediction , axis=1),
    target_var) , dtype=theano.config.floatX)
test_acc = T.mean(T.eq(T.argmax(test_prediction , axis=1)
    , target_var) , dtype=theano.config.floatX)

# Compile a function performing a training step on a
  mini-batch (by giving
# the updates dictionary) and returning the
  corresponding training loss:
train_fn = theano.function([input_var , target_var] , [
    loss , train_acc ] , updates=updates)

# Compile a second function computing the validation
  loss and accuracy:
```

---

```
val_fn = theano.function([input_var, target_var], [
    test_loss, test_acc])
pred_fn = theano.function([input_var], test_prediction)

print('train_epoch, minibatch, loss, test, elapsed');
test_err = 0
test_acc = 0
# We iterate over epochs:
for epoch in range(num_epochs):
    # In each epoch, we do a full pass over the training
    data:
    train_err = 0
    train_acc = 0
    train_batches = 0
    start_time = time.time()
    start_sample = 0
    if (training_samples > 60000):
        start_sample = random.randrange(0, (
            training_samples - 60000))
    start = time.time()
    for batch in iterate_minibatches(X_train[
        start_sample:(start_sample+60000)], y_train[
        start_sample:(start_sample+60000)], batch_size,
        shuffle=True):
        inputs, targets = batch
        err, acc = train_fn(inputs, targets)
        train_err += err
        train_acc += acc
        train_batches += 1
```

---

```

end = time.time()
batch_test = iterate_minibatches(X_test, y_test,
    10000, shuffle=True)
inputs, targets = batch_test.next()
test_err, test_acc = val_fn(inputs, targets)
print ('%02d,%02d/%i,%0.4f,%06.4f,%0.4f' %
    (epoch, train_batches*batch_size, train_size,
    train_err/train_batches, test_acc, (end - start))
    )

np.savez('model_mnist.npz', *lasagne.layers.
    get_all_param_values(network))
batch_valid = iterate_minibatches(X_test, y_test,
    test_samples, shuffle=False)
inputs, targets = batch_valid.next()
err, acc = val_fn(inputs, targets)
print ('samples:%d,loss_%,acc:_%06.4f' % (test_samples,
    err, acc))

if __name__ == '__main__':
    if ('--help' in sys.argv) or ('-h' in sys.argv):
        print ("Trains a neural network on MNIST using
            Lasagne.")
        print ("Usage: %s [batch_size]" % sys.argv[0])
        print ()
        print ("EPOCHS: number of training epochs to perform
            (default: 500)")
    else:

```

```
kwargs = {}  
if len(sys.argv) > 1:  
    kwargs['batch_size'] = int(sys.argv[1])  
main(**kwargs)
```

## Bibliografía

- [1] Colfax, R. N. (2009). Kennedy v. City of Zanesville-Making the Case for Water. *Hum. Rts.*, 36, 18.
- [2] Davies, T. (2010). Open data, democracy and public sector reform. A look at open government data use from data.gov.uk.
- [3] Hunnius, S., Krieger, B., & Schuppan, T. (2014, September). Providing, guarding, shielding: Open government data in Spain and Germany. In 2014 EGPA Annual Conference, 10-12 September 2014 in Speyer, Germany.
- [4] Nahon, K., Peled, A., & Shkabatur, J. Open Government Data Heartbeat: Cities Commitment to OGD.
- [5] Yang, T. M., Lo, J., & Shiang, J. (2015). To open or not to open? Determinants of open government data. *Journal of Information Science*, 41(5), 596-612.
- [6] Seijas, L. M., & Segura, D. D. T. D. E. (2003). Reconocimiento de dígitos manuscritos mediante redes neuronales: una técnica Híbrida. Departamento de Computación, Facultad de Ciencias exactas y Naturales. Universidad de Buenos Aires. Buenos Aires, 1.
- [7] Zuiderwijk, A. M. G. (2015). Open data infrastructures: The design of an infrastructure to enhance the coordination of open data use (Doctoral dissertation, TU Delft, Delft University of Technology).
- [8] Open Source Initiative. (2016, 5 Sep). <https://opensource.org/osd>
- [9] Fujisawa, H., Nakano, Y., & Kurino, K. (1992). Segmentation methods for character recognition: from segmentation to document structure analysis. *Proceedings of the IEEE*, 80(7), 1079-1092.

- 
- [10] Srihari, S. N. (1993). Recognition of handwritten and machine-printed text for postal address interpretation. *Pattern recognition letters*, 14(4), 291-302.
- [11] Tom, R., & Devis, J. License Plate Character Recognition.
- [12] Gustav Tauschek (1935). Reading machine. US Patent 2.026.330.
- [13] P. W. Handel (1933), Statistical machine. US. Patent 1.915.993.
- [14] Grimsdale, R. L., Sumner, F. H., Tunis, C. J., & Kilburn, T. (1959). A system for the automatic recognition of patterns. *Proceedings of the IEE-Part B: Radio and Electronic Engineering*, 106(26), 210-221.
- [15] The IBM 1401 manufacturing and test production line. (2016) URL: [https://www.researchgate.net/figure/224111701\\_fig4\\_FIGURE-7-The-IBM-1401-manufacturing-and-test-production-line-in-Endicott-New-York-in](https://www.researchgate.net/figure/224111701_fig4_FIGURE-7-The-IBM-1401-manufacturing-and-test-production-line-in-Endicott-New-York-in)
- [16] Casey, R., & Nagy, G. (1966). Recognition of printed Chinese characters. *IEEE Transactions on Electronic Computers*, (1), 91-101.
- [17] Tou, J. T., and R. C. Gonzalez. Recognition of handwritten characters by topological feature extraction and multilevel categorization. *IEEE Transactions on Computers* 100.7 (1972): 776-785.
- [18] Hothorn, T., & Zeileis, A. (2015). partykit: A modular toolkit for recursive partytioning in R. *Journal of Machine Learning Research*, 16, 3905-3909.
- [19] Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased Recursive Partitioning: A Conditional Inference Framework.
- [20] Ali, F., & Pavlidis, T. (1977). Syntactic recognition of handwritten numerals. *IEEE Transactions on Systems, Man, and Cybernetics*, 7(7), 537-541.
- [21] Kahan, S., Pavlidis, T., & Baird, H. S. (1987). On the recognition of printed characters of any font and size. *Pattern Analysis and Machine Intelligence*, *IEEE Transactions on*, (2), 274-288.

- 
- [22] Shridhar, M., & Badreldin, A. (1984). Handwritten numeral recognition by tree classification methods. *Image and Vision Computing*, 2(3), 143-149.
- [23] Rajavelu, A., Musavi, M. T., & Shirvaikar, M. V. (1989). A neural network approach to character recognition. *Neural Networks*, 2(5), 387-393.
- [24] Sabourin, M., & Mitiche, A. (1992). Optical character recognition by a neural network. *Neural networks*, 5(5), 843-852.
- [25] Kimura, F., Wakabayashi, T., Tsuruoka, S., & Miyake, Y. (1997). Improvement of handwritten Japanese character recognition using weighted direction code histogram. *Pattern recognition*, 30(8), 1329-1337.
- [26] Chu, J. L., & Krzyżak, A. (2014, May). Analysis of feature maps selection in supervised learning using convolutional neural networks. In *Canadian Conference on Artificial Intelligence* (pp. 59-70). Springer International Publishing.
- [27] K. Fukushima. *Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position*. 1980.
- [28] Hubel, David H and Wiesel, Torsten N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160(1):106, 1962.
- [29] Fukushima, K., Miyake, S., & Ito, T. (1983). Neocognitron: A neural network model for a mechanism of visual pattern recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, (5), 826-834.
- [30] Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, y L.D. Jackel. Handwritten digit recognition with a backpropagation network. *Advances in neural information processing systems*, 1990.
- [31] Fukushima, K., & Wake, N. (1991). Handwritten alphanumeric character recognition by the neocognitron. *IEEE Transactions on Neural Networks*, 2(3), 355-365.



- 
- [32] Fukushima, K. (1992). Character recognition with neural networks. *Neurocomputing*, 4(5), 221-233.
- [33] LeCun, Y., Jackel, L. D., Bottou, L., Cortes, C., Denker, J. S., Drucker, H. & Vapnik, V. (1995). Learning algorithms for classification: A comparison on handwritten digit recognition. *Neural networks: the statistical mechanics perspective*, 261, 276.
- [34] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324.
- [35] Y. LeCun, "The MNIST database of handwritten digits," <http://yann.lecun.com/exdb/mnist>.
- [36] Y. LeCun, F.J. Huang, y L. Bottou. Learning methods for generic object recognition with invariance to pose and lighting. *Computer Vision and Pattern Recognition*, 2004.
- [37] Ahranjany, S. S., Razzazi, F., & Ghassemian, M. H. (2010, September). A very high accuracy handwritten character recognition system for Farsi/Arabic digits using Convolutional Neural Networks. In *Bio-Inspired Computing: Theories and Applications (BIC-TA)*, 2010 IEEE Fifth International Conference on (pp. 1585-1592). IEEE.
- [38] Deng, J., Berg, A., Satheesh, S., Su, H., Khosla, A., & Fei-Fei, L. (2012). Imagenet large scale visual recognition competition 2012 (ILSVRC2012). <http://www.image-net.org/challenges/LSVRC/2012/>
- [39] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).
- [40] Pierre Sermanet, Koray Kavukcuoglu, Soumith Chintala, y Yann LeCun. Pedestrian detection with unsupervised multi-stage feature learning. 2013.

- 
- [41] P. Dollar. Caltech's pedestrian dataset. 2014. URL [http://www.vision.caltech.edu/Image\\_Datasets/CaltechPedestrians/](http://www.vision.caltech.edu/Image_Datasets/CaltechPedestrians/)
- [42] Navneet Dalal. Inria's pedestrian dataset. 2014. URL <http://pascal.inrialpes.fr/data/human/>.
- [43] Concha, G., & Naser, A. (2012). El desafío hacia el gobierno abierto en la hora de la igualdad. Comisión Económica para América Latina y el Caribe (CEPAL).
- [44] Geiger, C. P., & von Lucke, J. (2011). Open Government Data1. In Conference for E-Democracy and Open Government (p. 183).
- [45] Thelwall, M. (2001). A web crawler design for data mining. *Journal of Information Science*, 27(5), 319-325.
- [46] Congedo, G., Dimauro, G., Impedovo, S., & Pirlo, G. (1995, August). Segmentation of numeric strings. In *Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on* (Vol. 2, pp. 1038-1041). IEEE.
- [47] Khan, S. A. (1998). Character segmentation heuristics for check amount verification (Doctoral dissertation, Massachusetts Institute of Technology).
- [48] Hilera González, J. R., & Martínez Hernando, V. J. (2000). *Redes neuronales artificiales: fundamentos modelos y aplicaciones*. Madrid, España: Alfaomega Ra-Ma.
- [49] Sietsma, J., & Dow, R. J. (1991). Creating artificial neural networks that generalize. *Neural networks*, 4(1), 67-79.
- [50] Ordoñez, C. R., Santiago, J. J., & Jaimes, J. F. (2009). Reconocimiento de caracteres por medio de una red neuronal artificial. *Respuestas*, 14(1), 30-39.
- [51] Dean, J., Corrado, G., Monga, R., Chen, K., Devin, M., Mao, M., Ranzato, M., Senior, A., Tucker, P., Yang, K., Le, Q. V., and Ng, A. Y. Large scale distributed deep networks. In *NIPS*, pp. 1232–1240, 2012

- 
- [52] Sander Dieleman, Jan Schlüter, Colin Raffel, Eben Olson, Sren Kaae Sn-derby, Daniel Nouri, Daniel Maturana, Martin Thoma, Eric Battenberg, Jack Kelly, Jeffrey De Fauw, Michael Heilman, diogo149, Brian McFee, Hen-drik Weideman, takacsg84, peterderivaz, Jon, instagibbs, Dr. Kashif Rasul, CongLiu, Britefury, and Jonas Degraeve, Lasagne: First release., Aug. 2015. <https://github.com/Lasagne/Lasagne>
- [53] F. Bastien, P. Lamblin, R. Pascanu, J. Bergstra, I. J. Goodfellow, A. Bergeron, N. Bouchard, and Y. Bengio. Theano: new features and speed improvements. Deep Learning and Unsupervised Feature Learning NIPS 2012 Workshop, 2012.
- [54] Dahl, G. E., Sainath, T. N., & Hinton, G. E. (2013, May). Improving deep neural networks for LVCSR using rectified linear units and dropout. In Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on (pp. 8609-8613). IEEE.
- [55] Hu, B., Lu, Z., Li, H., & Chen, Q. (2014). Convolutional neural network ar- chitectures for matching natural language sentences. In Advances in Neural Information Processing Systems (pp. 2042-2050).
- [56] Scherer, D., Müller, A., & Behnke, S. (2010). Evaluation of pooling operations in convolutional architectures for object recognition. In Artificial Neural Net- works–ICANN 2010 (pp. 92-101). Springer Berlin Heidelberg.
- [57] Ciresan, D. C., Meier, U., Gambardella, L. M., & Schmidhuber, J. (2010). Deep, big, simple neural nets for handwritten digit recognition. *Neural computation*, 22(12), 3207-3220.
- [58] Rafiq, M. Y., Bugmann, G., & Easterbrook, D. J. (2001). Neural network design for engineering applications. *Computers & Structures*, 79(17), 1541-1552.
- [59] Enachescu, C., & Miron, C. D. (2009). Handwritten digits recognition using neural computing. *Scientific Bulletin of the "Petru Maior" University of Targu Mures*, 6, 17.

- 
- [60] Li, M., Zhang, T., Chen, Y., & Smola, A. J. (2014, August). Efficient mini-batch training for stochastic optimization. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 661-670). ACM.
- [61] Wilson, D. R., & Martinez, T. R. (2003). The general inefficiency of batch training for gradient descent learning. *Neural Networks*, 16(10), 1429-1451.
- [62] Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep sparse rectifier neural networks. In International Conference on Artificial Intelligence and Statistics (pp. 315-323).
- [63] Jolías, L. Open Data a nivel local en Argentina: ¿ transparencia o innovación democrática?.
- [64] Apache Hadoop. Available at <http://hadoop.apache.org>
- [65] La Nación: El partido que sacó cero votos en las elecciones: ni sus candidatos se eligieron. Nov 15 2013. URL: <http://www.lanacion.com.ar/1635796-el-partido-que-saco-cero-votos-en-las-elecciones-ni-sus-candidatos-se-eligieron>
- [66] La Gaceta: Convergencia Popular Porteña, el partido al que ni sus candidatos votaron. Nov 15 2013. URL: <http://www.lagaceta.com.ar/nota/568353/politica/convergencia-popular-portena-partido-al-ni-sus-candidatos-votaron.html>
- [67] La Voz: El partido político que sacó cero voto en las elecciones porteñas. Nov 15 2013. URL: <http://www.lavoz.com.ar/politica/el-partido-politico-que-saco-cero-voto-en-las-elecciones-portenas>
- [68] Datos tipeados de Resultados Electorales Elecciones Legislativas 2013 en Argentina. 27 Oct 2013. URL: <https://drive.google.com/open?id=0B3S6fgRmEp2GbHVUemt5SmY4Mlk>

- 
- [69] Buenos Aires Data. Resultados definitivos de las elecciones a legisladores porteños desarrollados en el 2013. URL: <http://data.buenosaires.gob.ar/dataset/elecciones-2013>
- [70] Yan, Z., & Shao, Y. Asynchronous Distributed Data Parallelism for Machine Learning.