

# Incorporating Natural Language Processing models in Mexico City's 311 Locatel

**Alejandro Molina-Villegas**

Conacyt - Center for Research in Geospatial Information Sciences

**Oscar S Siordia**

Center for Research in Geospatial Information Sciences

**Edwin Aldana-Bobadilla**

Conacyt - IPN Center for Research and Advanced Studies

**Jorge Luis Perez**

Digital Agency for Public Innovation Mexico City Government

## Abstract

Natural Language Processing based technologies are transforming various sectors by facilitating new ways of providing services through Artificial Intelligence (AI). In this paper, we describe the methodology and present the challenges encountered during the creation of a Deep Learning-based model for classifying citizen service requests. Our system is able to effectively recognize among 48 categories of public services with an accuracy of 97% and was integrated into Mexico City's 311, significantly increasing the government's ability to provide better services.

## 1 Introduction

The Unified Citizen Service System operated by Mexico City's Digital Agency for Public Innovation (ADIP) is a government mechanism to serve citizens' needs. Requests for assistance are registered directly by citizens via telephone or via free text through a Web portal<sup>1</sup>. If a citizen detects an urban infrastructure problem, he or she can enter a request on the portal stating, for example: "*the storm drain in front of my house has garbage in it and it clogs up when it rains*". Previously, a request had to be manually cataloged by an operator from any of the 79 responsible agencies in order to be attended, which could take up to 5 days. Thus, manual classification became a bottleneck. In a metropolitan area with almost 20 million inhabitants, automatic requests classification became a necessity. Faced with this scenario, the objective of this project was to develop a model capable of evaluating requests in real time so that the corresponding unit would be notified immediately. Thanks to the proposed IA solution, a volume of 33,000 requests per month is now handled quickly, providing a 24/7 help desk service, and reducing the number of failures in request assignment and review.

<sup>1</sup><https://311locatel.cdmx.gob.mx/>

## 2 Real data, real challenges

The most challenging aspect of integrating AI into Mexico City's new 311 Locatel service was intersecting NLP research with the existing specific problems of the public sector and citizen's needs. In this sense, unlike the corpora usually used for scientific research, our data featured many inconsistencies which posed several challenges that had to be continuously addressed until satisfactory results were obtained. The original database (DB) had 970,125 records in total, but there were many *null* values in the text field of interest (only 761,200 records contained non-null values). In addition, there were catalog and record inconsistencies. Catalog inconsistencies occur when apparently some of the available classes correspond to the same concept but, for some reason, they are not aggregated together in the DB. On the other hand, registration inconsistencies are those requests that have been incorrectly classified. Both types of inconsistencies were reported to DB administrators who, after a review, delivered a corrected version of the DB from which the methods and algorithms presented in section 3 were applied.

## 3 Methodology

### 3.1 Regularization of text length by automatic summarization

Another problem with our data is the great variability in text length. On the one hand, a large number of texts that are too short, some even lacking semantic content ("I", "Y", "A", "c", 3, "?", ".", "-"). On the other hand, there is a significant number of extremely long texts (some with more than 5,000 characters) that would be impossible, even for the operators of the 311 service, to assign to a category. The impact of processing texts that are too short can be seen at the time of the convolution of the neural network described in section 3.3: when a filter is moved over the text encoding matrix but

the matrix is too small, a stride error will occur (the filter cannot be moved without the matrix overflowing); conversely, the impact of too long texts affects the amount of memory required to host text encoding. In practice, input to the neural network should be a vector of fixed size; when taking the maximum length of the training set as a reference, the memory requirement is unnecessarily amplified.

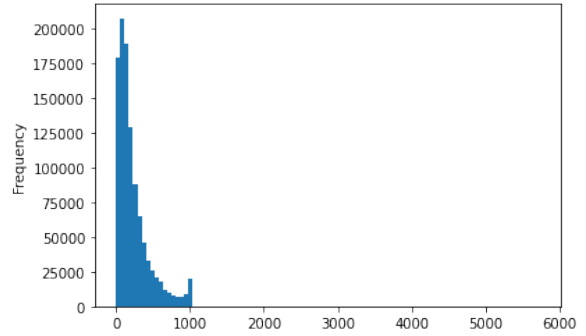
To avoid both effects, it was necessary to regulate the length distribution of original texts. Based on the statistical values of the distribution and a qualitative exploration, we filtered out the training texts consisting of less than 40 characters, and synthesized the long texts into a maximum of 1000 characters by using the [Barrios et al. \(2016\)](#) variant of the automatic summarization algorithm TextRank. Table 1 shows the statistical summary of the distribution of original (left) and corrected (right) lengths in number of characters. It can be seen that after the automatic summary, the mean and the standard deviation are not so affected, but the minimum and maximum values were adjusted according to the criteria established for the application. Figure 1 shows the histograms ( $nbins = 100$ ) corresponding to both distributions.

	Original lengths		Corrected lengths
<i>mean</i>	$2.329 \times 10^2$	<i>mean</i>	$2.492 \times 10^2$
<i>std</i>	$2.287 \times 10^2$	<i>std</i>	$2.126 \times 10^2$
<i>min</i>	$1.000 \times 10^0$	<i>min</i>	$4.000 \times 10^1$
<i>25%</i>	$8.300 \times 10^1$	<i>25%</i>	$1.030 \times 10^2$
<i>50%</i>	$1.590 \times 10^2$	<i>50%</i>	$1.760 \times 10^2$
<i>75%</i>	$2.980 \times 10^2$	<i>75%</i>	$3.180 \times 10^2$
<i>max</i>	$5.733 \times 10^3$	<i>max</i>	$1.000 \times 10^3$

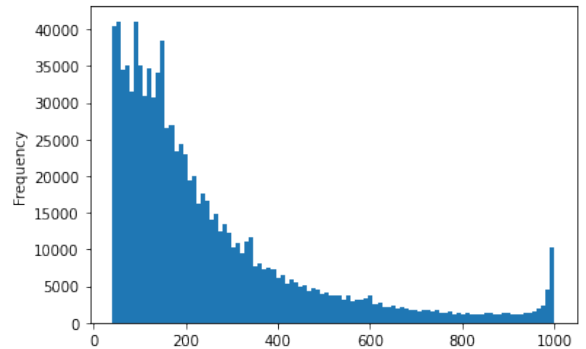
Table 1: Statistical summary of the distribution of text lengths after regularization with automatic summarization.

### 3.2 Lexical-orthographic variability correction encoding by Byte Pair Encoding

Another significant aspect to consider in the design of the solution is related to the high variability of styles that citizens have when writing the messages in the free text form. There is no restriction for users to correctly use upper or lower cases nor a requirement for them to properly use spelling, grammar, syntax, or semantic rules. That is, people can write however they want. Consider, for example, the word *socavón* (sinkhole), which presented all these variants: (*socabon*, *socabón*, *socavon*,



(a) Original length distribution.



(b) Length distribution with automatic summary.

Figure 1: Distribution of text lengths before and after regularization by automatic summarization.

*zocabon*, *zocavon*, *zocabones*). It is clear that although there are some collections of pre-trained embeddings that include Spanish (usually the European variant) ([Bojanowski et al., 2017](#); [Cardellino, 2019](#)), definitely none of them were trained with the specific (or at least similar) characteristics for the development of our model, focused on citizen attention requests, in Mexican Spanish, and with high variability. In view of this, we decided to train our own embeddings by encoding the text using a model based on Byte Pair Encoding (BPE) proposed by [Gage \(1994\)](#) for the generation of subwords.

Thanks to the generation of our BPE-based encoding model, the mentioned variants are decomposed into smaller parts which, being the new base variables, record existing patterns better. Using the previously mentioned example word, the encoding of its variants will be something similar to the following subwords: ("s" "o" "ca" "v" "ón", "s" "o" "ca" "b" "on", "s" "o" "ca" "b" "ón", "s" "o" "ca" "v" "on", "z" "o" "ca" "b" "on", "z" "o" "ca" "v" "on", "z" "o" "ca" "b" "on" "es"). Thus, by including the subword encoding model, we optimize the chance of finding common patterns despite the high

lexical-orthographic variability of the data. It is worth mentioning that each subword is determined automatically by considering its frequency, in a language-agnostic manner (Sennrich et al., 2015). It is also important to note that in order to generate the subword encoding model we did not need to have the classes associated with the text and thanks to this, we were able to take advantage of the text of all the records in the DB, including those without classes.

### 3.3 Automatic classification of citizen service requests using Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are widely used in applications such as object in image recognition (Goodfellow et al., 2016; Krizhevsky et al., 2012). However, this same architecture has also proven to be a powerful tool when adapted to NLP and there are even some works specifically dedicated to Mexican Spanish processing (Molina-Villegas et al., 2021; Cattin et al., 2022). The convolution layer is at the core of CNNs and is where most of the deep learning takes place.

In this layer, there are a series of feature detectors that move from the beginning to the end of the text verifying the patterns that co-occur in the classes. This process is known as 1D convolution. In practice, the feature detectors traverse a matrix of word embeddings, or in our case, a matrix of *subword embeddings*. The deep part of CNN is that during the learning process, guided by back-propagation, feature detectors can also update their weights and in this way, the network learns to select which patterns are useful. The best results of our CNN<sup>2</sup> implementation were obtained with the following criteria: {vocabulary size:  $2 \times 10^{15}$ , embeddings dimensions: 128, number of filters: 50, feedforward units: 512, dropout rate: 0.1, number of splits for stratified K-Fold: 5, batch size: 100, epochs: 30}.

## 4 Results and Discussion

Table 2 shows the results of the global evaluation of the best model, which was carried out on a test set consisting of 212, 539 records (not used during the training phase). The accuracy of the resulting model is higher than 0.8 and this allowed it to be used in production. In addition, we verified the ef-

<sup>2</sup>A total of 2262854 parameters (including sub-word embeddings) were trained in 144 hours.

	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>support</i>
<i>macro avg</i>	0.96	0.95	0.96	212539
<i>weighted avg</i>	0.97	0.97	0.97	212539

Table 2: Results of the 311 citizen attention request classification model.

fective recognition of 48 different utility categories (Table 3).

The code with the solution can be downloaded from a github repository (<https://github.com/alemol/locatel>) which includes the models generated for both the subword encoding and the convolutional network among other modules focused on text homogenization and error management. It is worth mentioning that, although the transferred models were highly effective in detecting the 48 classes, the solution provided is rigid in terms of the classes it recognizes (i.e. the incorporation of new classes would need retraining). For this reason, an important part of the technology transfer is that, besides providing the code and models, personnel training was also provided for those who will maintain the AI system so that they have the opportunity to modify and scale the solution.

## 5 Conclusion

We presented an AI solution that has already been integrated into the operation of the Mexico City government in terms of citizen service. Thanks to these types of solutions, we are adding value by improving citizen-centered services. We believe that this contribution will prevent researchers or technicians working on related problems from having to start from scratch.

## References

- Federico Barrios, Federico López, Luis Argerich, and Rosa Wachenchauzer. 2016. *Variations of the similarity function of textrank for automated summarization. arXiv preprint arXiv:1602.03606.*
- Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. *Enriching word vectors with subword information. Transactions of the association for computational linguistics*, 5:135–146.
- Cristian Cardellino. 2019. *Spanish Billion Words Corpus and Embeddings.*
- Thomas Cattin, Alejandro Molina-Villegas, Julieta Fuentes, and Oscar Siordia. 2022. *The geopolitical repercussions of us anti-immigrant rhetoric on mexican online speech about migration: A transdisciplinary approach. In Advances in Geospatial Data*

*Science*, Cham (Switzerland). Springer International Publishing.

Philip Gage. 1994. [A new algorithm for data compression](#). *C Users Journal*, 12(2):23–38.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT Press.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. [Imagenet classification with deep convolutional neural networks](#). In *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc.

Alejandro Molina-Villegas, Victor Muñiz-Sanchez, Jean Arreola-Trapala, and Filomeno Alcántara. 2021. [Geographic named entity recognition and disambiguation in mexican news using word embeddings](#). *Expert Systems with Applications*, 176:114855.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2015. [Neural machine translation of rare words with subword units](#). *arXiv preprint arXiv:1508.07909*.

	<i>precision</i>	<i>recall</i>	<i>support</i>
NEIGHBORHOOD WATCH ALARMS	0.96	0.97	151
EARTHQUAKE ALARMS	0.99	0.99	8149
STREET LIGHTING	0.99	0.99	32516
FUNERAL SERVICES SUPPORT	1.00	0.69	71
ACCREDITED THIRD PARTY COUNSEL	0.99	1.00	67
LEGAL COUNSELING	0.99	0.99	224
SOCIAL WELFARE	0.92	0.95	395
VETERINARY CARE	0.94	0.99	167
POTHOLE PATCHING	0.97	0.99	8928
ROAD SURFACE MARKING	0.96	0.98	1088
WEEDING	1.00	1.00	1795
SCHOLARSHIPS	0.98	0.96	812
COVID	0.92	0.94	6076
HOUSING CREDIT	0.97	0.95	59
STORM DRAIN CLEANING	0.95	0.99	11484
WATER SHORTAGE	0.98	0.99	20992
WATER LEAK	0.99	0.98	14652
STREET CLEANING	0.96	0.87	500
LLAVE CODE CDMX	0.98	0.93	685
STORM DRAIN MAINTENANCE	0.98	0.87	5761
SEWAGE MAINTENANCE	0.91	0.92	3664
PARK MAINTENANCE	0.89	0.94	3208
TRAFFIC LIGHT MAINTENANCE	0.96	0.89	1519
ROAD MAINTENANCE	0.94	0.94	10594
ASPHALTING	0.93	0.92	3218
TREE PRUNING-REMOVAL	0.98	0.98	24358
CARETAKER SUPPORT PROGRAM	1.00	1.00	21
CIVIL PROTECTION	0.90	0.94	1445
PUBLIC TRANSPORTATION	0.98	0.98	4600
COMPLAINT CIVIL SERVANT	0.93	0.94	488
TRASH COLLECTION	0.92	0.94	6013
NATIVE POPULATIONS FAIR	1.00	0.91	95
CENDI REGISTRY	0.96	0.91	1025
PAVING REPAIR	0.96	0.95	94
STREET MERCHANT REMOVAL	0.96	0.92	205
CONSTRUCTION WASTE	0.96	0.95	11149
HEARING REQUEST	0.95	0.89	455
NEIGHBORHOOD CONSULTATION	0.99	0.96	71
POLICE SURVEILLANCE	0.94	0.97	2403
LEAFLET	0.99	0.99	543
SOCIOECONOMIC STUDY	1.00	0.98	241
RISK ASSESSMENT	0.97	0.97	340
UNEMPLOYMENT INSURANCE	0.95	0.97	1231
MOTOR VEHICLE REGISTRY	0.98	0.98	10239
LAND USE	0.98	0.95	1686
ABANDONED VEHICLES REMOVAL	0.98	0.98	3255
DRUG DEALING	0.93	0.91	2690
ADMINISTRATIVE VERIFICATION	0.92	0.98	3117
<b>macro avg</b>	0.96	0.95	212539
<b>weighted avg</b>	0.97	0.97	212539

Table 3: Appendix with individual evaluation results with stratified sampling for the 48 categories of citizen service requests using the automatic classification model.

# Incorporando modelos de Procesamiento de Lenguaje Natural en el 311 locatel de la Ciudad de México

**Alejandro Molina-Villegas**

Conacyt-Centro de Investigación en Ciencias de Información Geoespacial

**Oscar S Siordia**

Centro de Investigación en Ciencias de Información Geoespacial

**Edwin Aldana-Bobadilla**

Conacyt-Centro de Investigación y de Estudios Avanzados del IPN

**Jorge Luis Perez Hernández**

Agencia Digital de Innovación Pública Gobierno de la Ciudad de México

## Abstract

Las tecnologías basadas en Procesamiento de Lenguaje Natural (PLN) están transformando diversos sectores al facilitar nuevas formas de prestar servicios mediante Inteligencia Artificial (IA). En este artículo, presentamos los desafíos encontrados y los métodos utilizados durante la generación de un modelo de clasificación de solicitudes de atención ciudadana basado en Deep Learning. La IA es capaz de reconocer eficazmente entre 48 categorías de servicios públicos con un accuracy de 97% y fue integrada al 311 de la Ciudad de México aumentando de manera importante la capacidad del gobierno para brindar mejores servicios.

## 1 Introducción

El Sistema Unificado de Atención Ciudadana operado por la Agencia Digital de Innovación Pública (ADIP) de la Ciudad de México es un mecanismo gubernamental para atender a la ciudadanía. Las solicitudes de atención son registradas directamente por los ciudadanos por vía telefónica o a través de texto libre mediante un portal Web<sup>1</sup>. Si un ciudadano detecta un problema de infraestructura urbana, puede ingresar una solicitud en el portal escribiendo por ejemplo: *"la coladera que está frente a mi casa tiene basura y se tapa cuando llueve"*. Anteriormente, para que la solicitud pudiera ser atendida por alguna de las 79 dependencias responsables, primero debía ser catalogada manualmente por un operador, lo cual tomaba hasta 5 días. Así, la clasificación manual llegó a originar un cuello de botella. En una zona metropolitana con casi veinte millones de habitantes, la clasificación automática de las solicitudes se convirtió en una necesidad. Ante este panorama, el objetivo planteado en este proyecto fue elaborar un modelo capaz de evaluar las solicitudes en tiempo real para que, de manera instantánea, la dependencia correspondiente a la atención fuera notificada. Gracias a la solución de

<sup>1</sup><https://311locatel.cdmx.gob.mx/>

IA propuesta, hoy en día se gestiona rápidamente un volumen de 33 mil solicitudes mensuales, lo que permite proporcionar servicio de ayuda las 24 horas del día, todos los días del año, y reduce el número de fallos en la delegación y la revisión de solicitudes.

## 2 Datos reales, desafíos reales

El aspecto más desafiante de integrar IA al nuevo servicio 311 locatel de la Ciudad de México fue que en este proyecto se cruzan las promesas de la investigación en PLN con los problemas específicos del sector público y las necesidades de la ciudadanía. En este sentido, a diferencia de los corpus que se suelen utilizar para la investigación científica, los datos que utilizamos presentaron muchas inconsistencias además de varios retos que tuvimos que atender hasta obtener resultados satisfactorios. La base de datos (BD) original tenía 970 125 registros en total pero había muchos valores nulos en el campo de texto de interés. Del total, solamente 761 200 registros contenían valores no nulos. Además, había inconsistencias de catálogo y de registro. Las inconsistencias de catálogo ocurren cuando aparentemente algunas de las clases disponibles corresponden al mismo concepto pero por alguna razón están desagregadas en la BD. Las inconsistencias de registro consisten en la clasificación incorrecta de algunas solicitudes. Ambos tipos de inconsistencias fueron reportadas ante los administradores de la BD quienes, luego de una revisión entregaron una segunda versión corregida de la BD a partir de la cual se aplicaron los métodos y algoritmos, presentados en la sección 3.

## 3 Metodología

### 3.1 Regularización de la extensión del texto mediante resumen automático

Otro problema con los datos es la gran variabilidad en la extensión de los textos. Por un lado, un

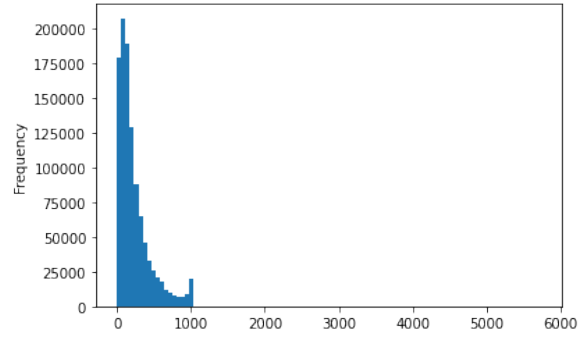
gran número de textos demasiado cortos, algunos incluso carentes de contenido semántico ("I", "Y", "A", "c", 3, "?", ".", "-"). Por otro lado, un número importante de textos extremadamente largos (algunos con más de 5 mil caracteres) que resultaría imposible, incluso para los operadores del servicio 311, asignarles a una categoría. El impacto de los textos demasiado cortos ocurre al momento de la convolución de la red neuronal descrita en la sección 3.3. Cuando un filtro se desplaza sobre la matriz que codifica al texto, pero esta es demasiado pequeña ocurrirá un error de stride (el filtro no se puede desplazar sin desbordar la matriz). El impacto de los textos demasiado largos es la cantidad de memoria necesaria para alojar la codificación del texto. En la práctica, el input de la red neuronal debe ser un vector de tamaño fijo; de manera que al tomar como referencia la extensión máxima del conjunto de entrenamiento la exigencia de memoria se amplifica innecesariamente.

Para evitar ambos efectos, fue necesario regularizar la distribución de longitudes de los textos originales. Basándonos en los valores estadísticos de la distribución y una exploración cualitativa, filtramos fuera del entrenamiento los textos de menos de 40 caracteres y sintetizamos los textos largos a un máximo de 1000 caracteres usando la variante de Barrios et al. (2016) del algoritmo de resumen automático TextRank.

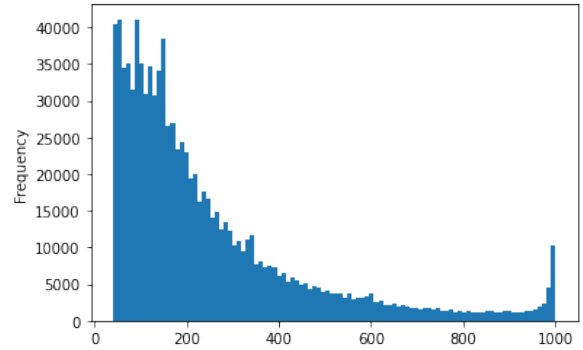
En la tabla 1 se presentan el resumen estadístico de la distribución de longitudes originales (izquierda) y corregidas (derecha) en número de caracteres. Se observa que, después del resumen automático, la media y la mediana no resultan tan afectadas pero los valores mínimo y máximo fueron ajustados de acuerdo con los criterios establecidos para la aplicación. En la figura 1 se presentan los histogramas ( $n_{bins} = 100$ ) correspondientes a ambas distribuciones.

longitudes originales		longitudes corregidas	
<i>mean</i>	$2.329 \times 10^2$	<i>mean</i>	$2.492 \times 10^2$
<i>std</i>	$2.287 \times 10^2$	<i>std</i>	$2.126 \times 10^2$
<i>min</i>	$1.000 \times 10^0$	<i>min</i>	$4.000 \times 10^1$
<i>25%</i>	$8.300 \times 10^1$	<i>25%</i>	$1.030 \times 10^2$
<i>50%</i>	$1.590 \times 10^2$	<i>50%</i>	$1.760 \times 10^2$
<i>75%</i>	$2.980 \times 10^2$	<i>75%</i>	$3.180 \times 10^2$
<i>max</i>	$5.733 \times 10^3$	<i>max</i>	$1.000 \times 10^3$

Table 1: Resumen estadístico de la distribución de longitudes de los textos después de la regularización con resumen automático.



(a) Distribución de longitudes original.



(b) Distribución de longitudes con resumen automático.

Figure 1: Distribución de longitudes de los textos antes y después de la regularización mediante resumen automático.

### 3.2 Corrección de Variabilidad léxico-ortográfica mediante codificación Byte Pair Encoding

Otro problema de suma importancia en el diseño de la solución está relacionado con la manera, tan diversa, de escribir que tiene la ciudadanía. Recordemos que los valores de los textos que nos interesa procesar provienen de un formulario en texto libre. No hay restricción alguna para que los usuarios hagan uso correcto de mayúsculas o minúsculas y mucho menos para que hagan uso correcto de la gramática, la sintaxis o la semántica. Es decir, la gente escribe como quiere. Consideremos por ejemplo la palabra "socavón", la cual presentó todas estas variantes: ("socabon", "socabón", "socavon", "zocabon", "zocavon", "zocabones"). Es claro que aunque existen algunas colecciones de embeddings preentrenados que incluyen español (generalmente la variante de Europa) (Bojanowski et al., 2017; Cardellino, 2019), definitivamente ninguna de ellas fue entrenado con las características específicas (o al menos parecidas) para el desarrollo de nuestro modelo enfocado en solicitudes de atención ciudadana, en español de México, con alta vari-

abilidad léxico-ortográfica. Ante esto, decidimos entrenar nuestros propios *embeddings* codificando el texto mediante la un modelo basado en Byte Pair Encoding (BPE) propuesto por Gage (1994) para la generación de subpalabras.

Gracias a la generación de nuestro modelo de codificación basado en BPE, las variantes mencionadas son desagregadas en partes más pequeñas que al ser las nuevas variables, registran mejor los patrones existentes. Usando la palabra de ejemplo, la codificación de sus variantes será algo similar a las siguientes subpalabras: ("s" "o" "ca" "v" "ón", "s" "o" "ca" "b" "on", "s" "o" "ca" "b" "ón", "s" "o" "ca" "v" "on", "z" "o" "ca" "b" "on", "z" "o" "ca" "v" "on", "z" "o" "ca" "b" "on" "es"). Así, al incluir el modelo de codificación en subpalabras, optimizamos la oportunidad de encontrar patrones comunes a pesar de la alta variabilidad léxico-ortográfica de los datos. Cabe mencionar que cada subpalabra es determinada de manera automática considerando su frecuencia, de manera agnóstica al idioma (Sennrich et al., 2015). También es importante señalar que para generar el modelo de codificación en subpalabras no necesitábamos tener las clases asociadas al texto y gracias a esto, pudimos aprovechar el texto de todos los registros en la BD, incluidos aquellos sin clase.

### 3.3 Clasificación automática de solicitudes de atención ciudadana usando Redes Neuronales Convolucionales

La Red Neuronal Convolutiva (CNN) es ampliamente utilizada en aplicaciones como el reconocimiento de objetos en imágenes (Goodfellow et al., 2016; Krizhevsky et al., 2012). No obstante, esta misma arquitectura también ha demostrado ser una potente herramienta cuando se adapta al PLN y existen incluso algunos trabajos dedicados específicamente al procesamiento de español de México (Molina-Villegas et al., 2021; Cattin et al., 2022). La capa de convolución está en el core de las CNN y es donde recae mayormente el aprendizaje profundo. En dicha capa, se tiene una serie de detectores de características que se mueven desde el inicio hasta el fin del texto verificando los patrones que co-ocurren en las clases. Este proceso es conocido como convolución 1D. En la práctica los detectores de características recorren una matriz de *word embeddings*, o en nuestro caso, una matriz de *subword embeddings*. La parte profunda de la CNN es que durante el proceso de apren-

	<i>precision</i>	<i>recall</i>	<i>f1-score</i>	<i>support</i>
<i>macro avg</i>	0.96	0.95	0.96	212539
<i>weighted avg</i>	0.97	0.97	0.97	212539

Table 2: Resultados globales del modelo de clasificación de solicitudes de atención ciudadana 311 para la ciudad de México.

dizaje, guiado por *backpropagation*, los detectores de características también pueden actualizar sus pesos y de esta manera, la red aprende a aprender qué patrones son útiles. Los mejores resultados de nuestra implementación de CNN<sup>2</sup> se obtuvieron con los siguientes criterios: {tamaño de vocabulario:  $2 \times 10^5$ , dimensiones de embeddings: 128, número de filtros=50, unidades feedforward=512, tasa de dropout=0.1, número de splits para stratified K-Fold: 5, batch size: 100, épocas: 30}.

## 4 Resultados y Discusión

En la tabla 2 presentamos los resultados de la evaluación global del mejor modelo se realizaron con 212, 539 registros (no utilizados durante la fase de entrenamiento). El *accuracy* del modelo resultante es superior a 0.8 y esto permitió poder usarlo en producción. Además, se verificó el reconocimiento eficaz de 48 categorías de servicios públicos diferentes (tabla 3).

La solución transferida al gobierno se puede descargar en un repositorio github (<https://github.com/alemol/locatel>) el cual incluye los modelos generados tanto para la codificación en subpalabras como para la red convolutiva entre otros módulos enfocados a la homogeneización del texto y la gestión de errores. Cabe mencionar que si bien los modelos transferidos resultaron altamente eficaces detectando las 48 clases, tenemos el problema de que la solución otorgada es rígida en cuanto a la incorporación de nuevas clases. Por esta razón, una parte importante de la transferencia tecnológica es que además de otorgar el código y los modelos, también se brindó capacitación al personal que dará mantenimiento a la IA de manera que les brindamos la oportunidad de modificar y escalar la solución.

## 5 Conclusión

Presentamos una solución de IA que ya se integró a la operación del gobierno de la Ciudad de México en materia de atención ciudadana. Gracias a

<sup>2</sup>Se entrenaron un total de 2, 262, 854 parámetros (incluidos los sub-word embeddings) en 144 horas.

este tipo de soluciones estamos aportando valor al mejorar los servicios centrados en el ciudadano. Consideramos que esta aportación evitará que los investigadores o técnicos, trabajando problemas relacionados, tengan que empezar desde cero.

## References

- Federico Barrios, Federico López, Luis Argerich, and Rosa Wachenchauzer. 2016. [Variations of the similarity function of textrank for automated summarization](#). *arXiv preprint arXiv:1602.03606*.
- Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. [Enriching word vectors with subword information](#). *Transactions of the association for computational linguistics*, 5:135–146.
- Cristian Cardellino. 2019. [Spanish Billion Words Corpus and Embeddings](#).
- Thomas Cattin, Alejandro Molina-Villegas, Julieta Fuentes-Carrera, and Oscar S. Siordia. 2022. The geopolitical repercussions of us anti-immigrant rhetoric on mexican online speech about migration: A transdisciplinary approach. In *Advances in Geospatial Data Science*, pages 41–51, Cham (Switzerland). Springer International Publishing.
- Philip Gage. 1994. [A new algorithm for data compression](#). *C Users Journal*, 12(2):23–38.
- Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. MIT Press.
- Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. [Imagenet classification with deep convolutional neural networks](#). In *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc.
- Alejandro Molina-Villegas, Victor Muñoz-Sánchez, Jean Arreola-Trapala, and Filomeno Alcántara. 2021. [Geographic named entity recognition and disambiguation in mexican news using word embeddings](#). *Expert Systems with Applications*, 176:114855.
- Rico Sennrich, Barry Haddow, and Alexandra Birch. 2015. [Neural machine translation of rare words with subword units](#). *arXiv preprint arXiv:1508.07909*.

	<i>precision</i>	<i>recall</i>	<i>support</i>
ALARMAS VECINALES	0.96	0.97	151
ALERTA SÍSMICA	0.99	0.99	8149
ALUMBRADO	0.99	0.99	32516
APOYO SERVICIOS FUNERARIOS	1.00	0.69	71
ASESORÍA DE TERCEROS ACREDITADOS	0.99	1.00	67
ASESORÍA JURÍDICA	0.99	0.99	224
ASISTENCIA SOCIAL	0.92	0.95	395
ASISTENCIA VETERINARIA	0.94	0.99	167
BACHEO	0.97	0.99	8928
BALIZAMIENTO	0.96	0.98	1088
BARBECHO-CHAPONEO	1.00	1.00	1795
BECAS	0.98	0.96	812
COVID	0.92	0.94	6076
CRÉDITO DE VIVIENDA	0.97	0.95	59
DESAZOLVE	0.95	0.99	11484
FALTA DE AGUA	0.98	0.99	20992
FUGA DE AGUA	0.99	0.98	14652
LIMPIEZA VÍA PÚBLICA	0.96	0.87	500
LLAVE CDMX	0.98	0.93	685
MANTENIMIENTO DE ALCANTARILLA	0.98	0.87	5761
MANTENIMIENTO DRENAJE	0.91	0.92	3664
MANTENIMIENTO PARQUE-ÁREA VERDE	0.89	0.94	3208
MANTENIMIENTO SEMÁFOROS	0.96	0.89	1519
MANTENIMIENTO VÍA PÚBLICA	0.94	0.94	10594
PAVIMENTACIÓN	0.93	0.92	3218
PODA-RETIRO ÁRBOL	0.98	0.98	24358
PROGRAMA APOYO A CUIDADORES	1.00	1.00	21
PROTECCIÓN CIVIL	0.90	0.94	1445
QUEJA DE TRANSPORTE PÚBLICO	0.98	0.98	4600
QUEJA FUNCIONARIO	0.93	0.94	488
RECOLECCIÓN BASURA	0.92	0.94	6013
REGISTRO FERIAS INDÍGENAS	1.00	0.91	95
INSCRIPCIÓN AL CENDI	0.96	0.91	1025
REPARACIÓN DE EMPEDRADO	0.96	0.95	94
RETIRO AMBULANTE	0.96	0.92	205
RETIRO CASCAJO-ESCOMBRO-RAMAS	0.96	0.95	11149
SOLICITUD DE AUDIENCIA	0.95	0.89	455
SOLICITUD DE CONCERTACIÓN VECINAL	0.99	0.96	71
SOLICITUD DE VIGILANCIA	0.94	0.97	2403
SOLICITUD DE VOLANTEO	0.99	0.99	543
SOLICITUD ESTUDIO SOCIOECONÓMICO	1.00	0.98	241
SOLICITUD EVALUACIÓN DE RIESGO	0.97	0.97	340
SOLICITUD SEGURO DE DESEMPLEO	0.95	0.97	1231
TRAMITES INFO VEHICULAR	0.98	0.98	10239
USO DE SUELO	0.98	0.95	1686
VEHÍCULO ABANDONADO-CHATARRA	0.98	0.98	3255
VENTA DE ALCOHOL-DROGA	0.93	0.91	2690
VERIFICACIÓN ADMINISTRATIVA	0.92	0.98	3117
<i>macro avg</i>	0.96	0.95	212539
<i>weighted avg</i>	0.97	0.97	212539

Table 3: **Apéndice** con resultados individuales de evaluación con muestreo estratificado para las 48 categorías de solicitudes de atención ciudadana usando el modelo de clasificación automática.