



Experiencias y desafíos en el uso de codificación automatizada en encuestas de hogares: Codificación de los clasificadores CIUO, CCIF y Establecimientos de compra, en la IX EPF de Chile

Departamento de Estadísticas Demográficas y Sociales
Subdepartamento de Estadísticas Socioeconómicas

Noviembre de 2024

Contenidos





Antecedentes IX EPF

Encuesta de Presupuestos Familiares

Versiones EPF

I EPF	• 1956 - 1957
II EPF	• 1968 - 1969
III EPF	• 1977 - 1978
IV EPF	• 1987 - 1988
V EPF	• 1996 - 1997
VI EPF	• 2006 - 2007
VII EPF	• 2011 - 2012
VIII EPF	• 2016 - 2017
IX EPF	• 2021 - 2022

• Objetivo principal

Identificar la **estructura y características del gasto en consumo final** de los hogares urbanos, en las capitales regionales y sus zonas conurbadas, con un período de referencia de un año.

• Objetivo secundario

Identificar la **estructura del ingreso disponible** de los hogares urbanos, en las capitales regionales y sus zonas conurbadas, con un período de referencia de un año.

Principales usos

Canasta IPC

Ponderaciones IPC

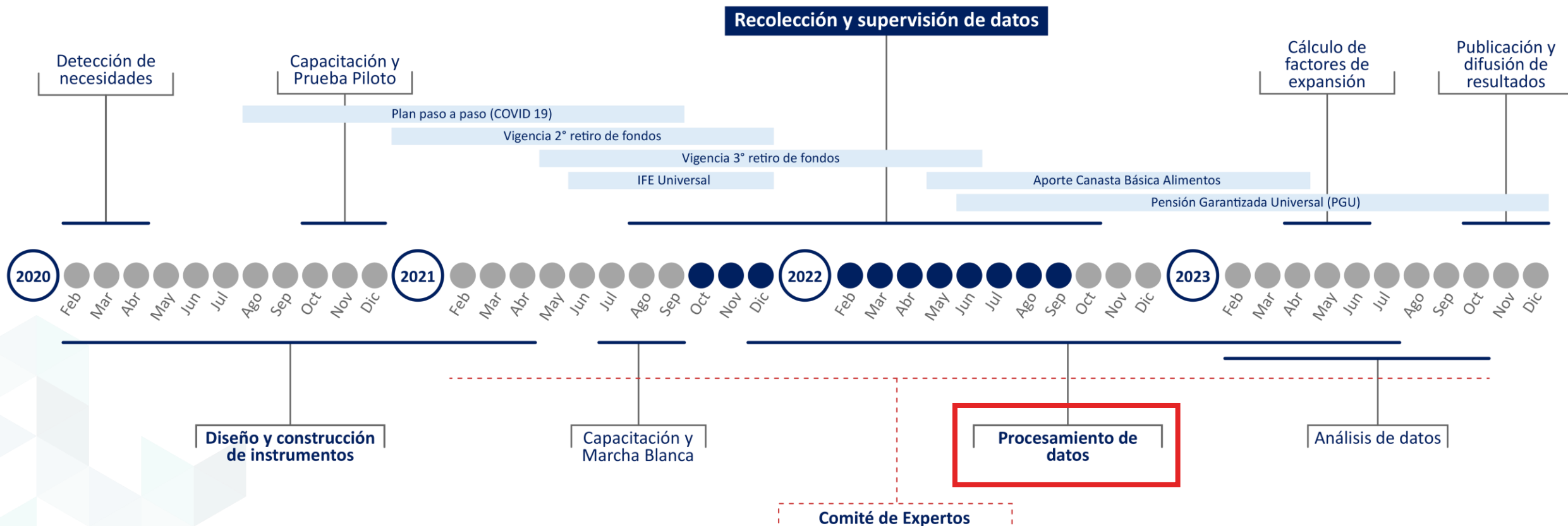
Pobreza

Cuentas Nacionales

Establecimientos

Salud

Alimentación

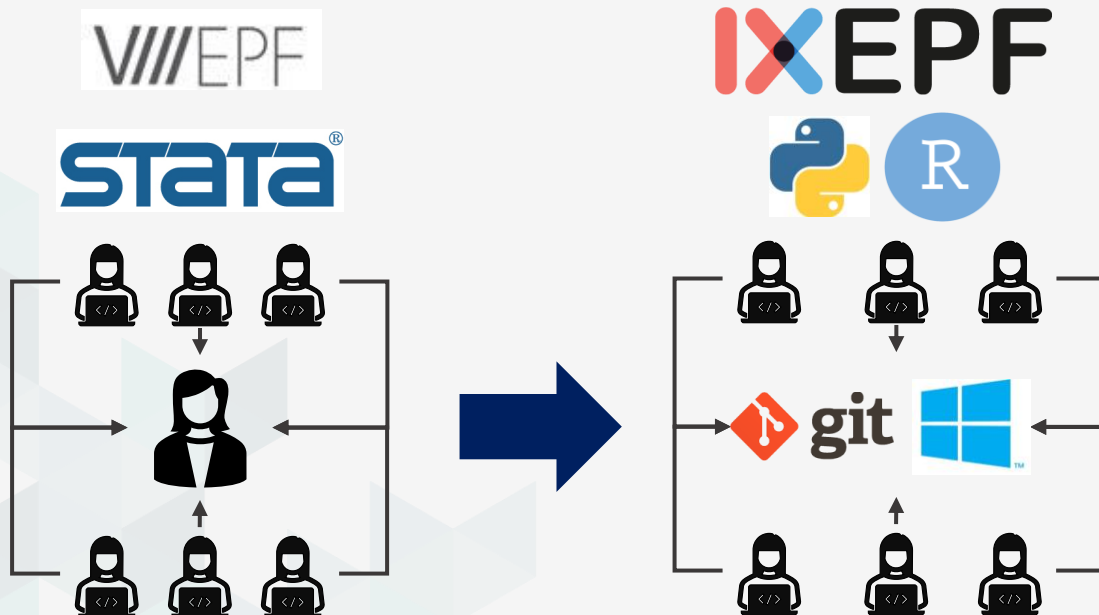




Innovaciones en procesamiento y análisis

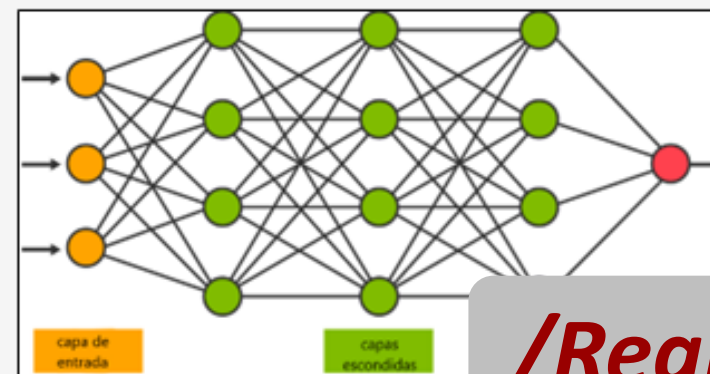
Automatización parcial del procesamiento y análisis:

El uso de un servidor Rstudio para la programación, además de la generación de buenas prácticas para el uso de sistemas integrados, facilitó la implementación de mejoras como la codificación automatizada.



Codificación automatizada:

Uso de redes neuronales y expresiones regulares para la codificación automatizada del clasificador de Ocupaciones (CIUO), Clasificador de Consumo Individual por Finalidades (CCIF) y Clasificador de Establecimientos Comerciales (CEC).



/RegEx. /*



Codificación automatizada

Clasificadores:

Clasificador de Consumo Individual por Finalidades (CCIF 2018.CL)

Pertenece a la familia de clasificaciones estadísticas denominadas “funcionales”, las que permiten analizar el **propósito o finalidad de los gastos** o transacciones que realizan determinados sectores institucionales.

Clasificación Internacional Uniforme de Ocupaciones (CIUO 08.CL)

Permite ordenar jerárquicamente y agrupar, de acuerdo con características similares, **diversas ocupaciones de acuerdo al tipo de tareas realizadas en un puesto de trabajo** y las competencias requeridas para ello, dependiendo del nivel y de la especialización de estas competencias.

Clasificador interno de establecimiento comerciales (CEC)

Clasificador de **puntos de compra** interno que permite identificar los establecimientos para ayudar con la codificación de productos.

Clasificador interno de Establecimientos Comerciales (CEC)

Clasifica los diversos puntos de venta de productos y servicios, contiene 29 categorías que incluyen tres residuales.

Dependiendo del módulo en el cual se registró el gasto el establecimiento viene como una glosa abierta que debe ser codificada.

En la tabla se muestran los primeros diez códigos del clasificador.

Código	Tipo de establecimiento
01	Supermercados
02	Almacén tradicional
03	Ferias Libres
04	Restaurantes y bares
05	Distribuidoras – Mayoristas
06	Vegas – Mercados
07	Comercio Ambulante
08	Multitiendas
09	Ferretería y multiferreterías
10	Farmacias
.	.
.	.
.	.

Clasificador de Consumo Individual por Finalidades (CCIF 2018.CL)

La **CCIF 2018.CL**, corresponde a la adaptación nacional de la “Clasificación del Consumo Individual por Finalidades - CCIF 2018” (COICOP, en inglés) de Naciones Unidas, la cual es parte del sistema de cuentas nacionales desde 1968.



La **IX Encuesta de Presupuestos Familiares** realizó la codificación de productos y servicios con el **CCIF 2018.CL**, siendo una de las primeras encuestas de ingresos y gastos del mundo en contar con una versión actualizada del clasificador.



Estructura del clasificador CCIF

- La CCIF abarca los gastos de consumo individual de tres sectores institucionales: “Hogares”, “Instituciones Sin Fines de Lucro que sirven a los Hogares” y “Gobierno General”. La CCIF está compuesta por quince divisiones, pero para la IX EPF, solo se consideran las trece divisiones correspondientes al sector de los hogares.

DIVISIÓN	GLOSA
01.0.0.00.00	ALIMENTOS Y BEBIDAS NO ALCOHÓLICAS
02.0.0.00.00	BEBIDAS ALCOHÓLICAS, TABACO Y ESTUPEFACIENTES
03.0.0.00.00	VESTUARIO Y CALZADO
04.0.0.00.00	VIVIENDA, AGUA, ELECTRICIDAD, GAS Y OTROS COMBUSTIBLES
05.0.0.00.00	MUEBLES, EQUIPAMIENTO PARA EL HOGAR Y MANTENIMIENTO RUTINARIO DE LA VIVIENDA
06.0.0.00.00	SALUD
07.0.0.00.00	TRANSPORTE
08.0.0.00.00	INFORMACIÓN Y COMUNICACIÓN
09.0.0.00.00	RECREACIÓN, DEPORTES Y CULTURA
10.0.0.00.00	SERVICIOS DE EDUCACIÓN
11.0.0.00.00	SERVICIOS DE RESTAURANTES Y ALOJAMIENTO
12.0.0.00.00	SEGUROS Y SERVICIOS FINANCIEROS
13.0.0.00.00	CUIDADO PERSONAL, ASISTENCIA SOCIAL Y BIENES Y SERVICIOS DIVERSOS

Estructura del clasificador CCIF

- Estas trece divisiones se subdividen en grupos, y estos a su vez se subdividen en clases y subclases. Estos cuatro niveles de división del clasificador son establecidos por las Naciones Unidas y permiten la comparabilidad internacional de manera estandarizada. Sin embargo, las subclases se subdividen nuevamente en producto, las clasificaciones a este nivel utilizados en la IX EPF han sido adaptadas a la realidad de Chile.
- La cantidad total de productos es 1111.

Nivel del clasificador	Código	Glosa
División	03.0.0.00.00	VESTUARIO Y CALZADO
Grupo	03.1.0.00.00	VESTUARIO
Clase	03.1.2.00.00	VESTUARIO PARA HOMBRES O NIÑOS, MUJERES O NIÑAS Y LACTANTES
Subclase	03.1.2.04.00	VESTUARIO PARA MUJERES O NIÑAS
Producto	03.1.2.04.02	ABRIGOS, CHAQUETAS Y PARKAS PARA MUJERES

Clasificador
CCIF

DIVISIÓN

GRUPO

CLASE

SUBCLASE

PRODUCTO



Clasificador de Consumo Individual
por Finalidades (CCIF 2018.CL)

Clasificador interno de
Establecimientos Comerciales (CEC)

La programación para la codificación automatizada del clasificador CCIF y el de Establecimientos comerciales se diseñó utilizando **expresiones regulares** que detectan **patrones comunes** en la manera de referirse a los productos o establecimientos.

Las expresiones regulares detectan patrones en las glosas abiertas que identifican el texto como un producto en el clasificador y a la vez detecta patrones de texto que no corresponden con el producto.

Por ejemplo:

Producto que intentamos codificar: 1 kg de carne molida

Patrón que buscamos: (?=CARNE.*) (?=MOLIDA.*)

Patrón que dejamos fuera: BISTEC/POLLO/ETC...

Clasificador de Consumo Individual por Finalidades (CCIF 2018.CL)

¿Qué debe incluir la glosa?

- Debe incluir la palabra HUEVO
- Puede incluir palabras como BANDEJA – GALLINA – BLANCO - FRESCO

¿Qué NO debe incluir la glosa?

- No debe incluir ningún tipo de pasta
- No debe incluir preparaciones
- No debe incluir funciones para el huevo
- Palabras específicas: PASCUA - MAYONESA

Descripción del gasto
SPAGUETTI CON HUEVO
FIDEO CON HUEVO 400 G
BANDEJA HUEVO
HUEVO
PAILA HUEVOS
HUEVOS PARA COMIDA DE PERRO
HUEVOS FRITOS
HUEVITOS DE PASCUA
MAYONESA TRADICIONAL SUPREME DE HUEVO NORMAL
BANDEJA DE HUEVOS DE GALLINA
HUEVO BLANCO
HUEVO FRESCO
HUEVOS

Clasificador de Consumo Individual por Finalidades (CCIF 2018.CL)

¿Qué debe incluir la glosa?

- Debe incluir la palabra HUEVO
- Puede incluir palabras como BANDEJA – GALLINA – BLANCO - FRESCO

¿Qué NO debe incluir la glosa?

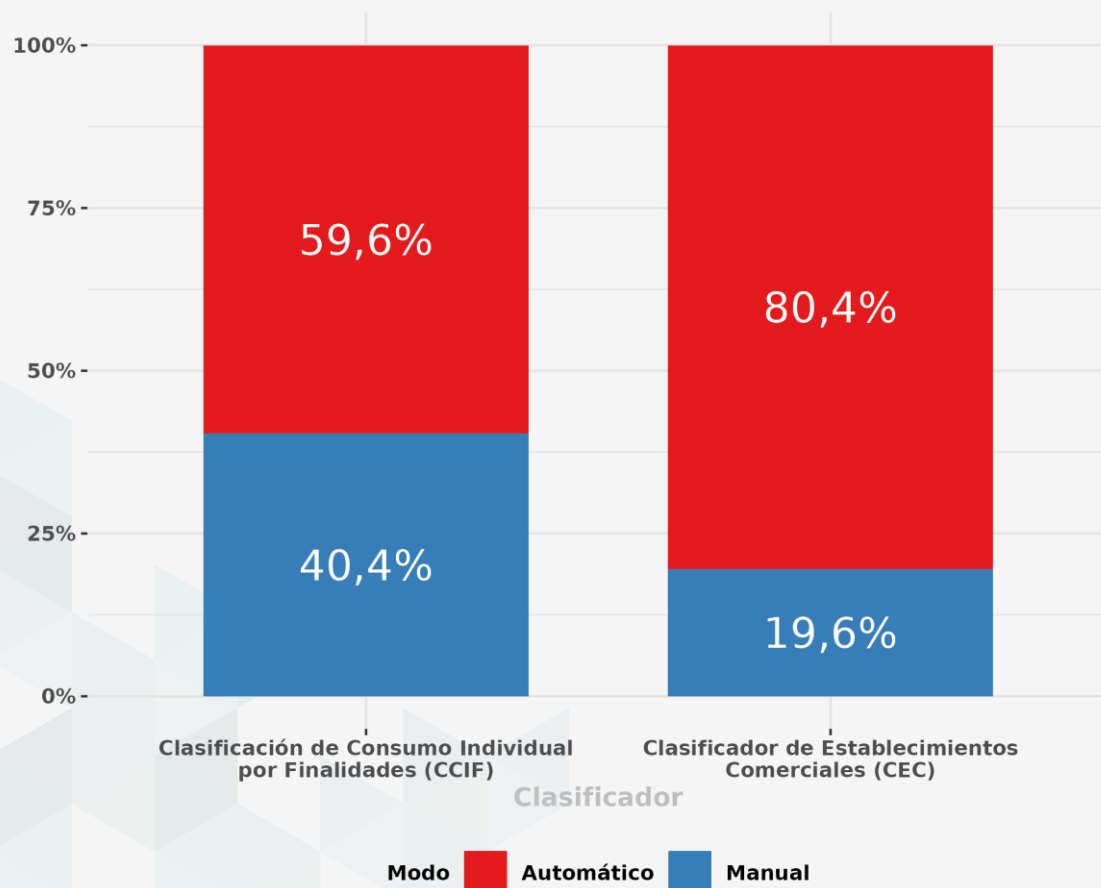
- No debe incluir ningún tipo de pasta
- No debe incluir preparaciones
- No debe incluir funciones para el huevo
- Palabras específicas: PASCUA - MAYONESA

Se le asigna el código **CCIF 01.1.4.08.01** a las glosas detectadas

Descripción del gasto
SPAGUETTI CON HUEVO
FIDEO CON HUEVO 400 G
BANDEJA HUEVO
HUEVO
PAILA HUEVOS
HUEVOS PARA COMIDA DE PERRO
HUEVOS FRITOS
HUEVITOS DE PASCUA
MAYONESA TRADICIONAL SUPREME DE HUEVO NORMAL
BANDEJA DE HUEVOS DE GALLINA
HUEVO BLANCO
HUEVO FRESCO
HUEVOS

Codificación automatizada de la CCIF por reglas

Clasificador de Consumo Individual por Finalidades (CCIF 2018.CL)



Clasificador interno de Establecimientos Comerciales (CEC)

Del total de registros enviados al equipo de codificación, un **59,6%** fue codificado automáticamente con éxito en su glosa de descripción del gasto con la clasificación **CCIF**.

Del total de registros enviado al equipo de codificación, un **80,4%** fue codificado automáticamente con éxito en su glosa de establecimiento comercial con el clasificador **CEC**.

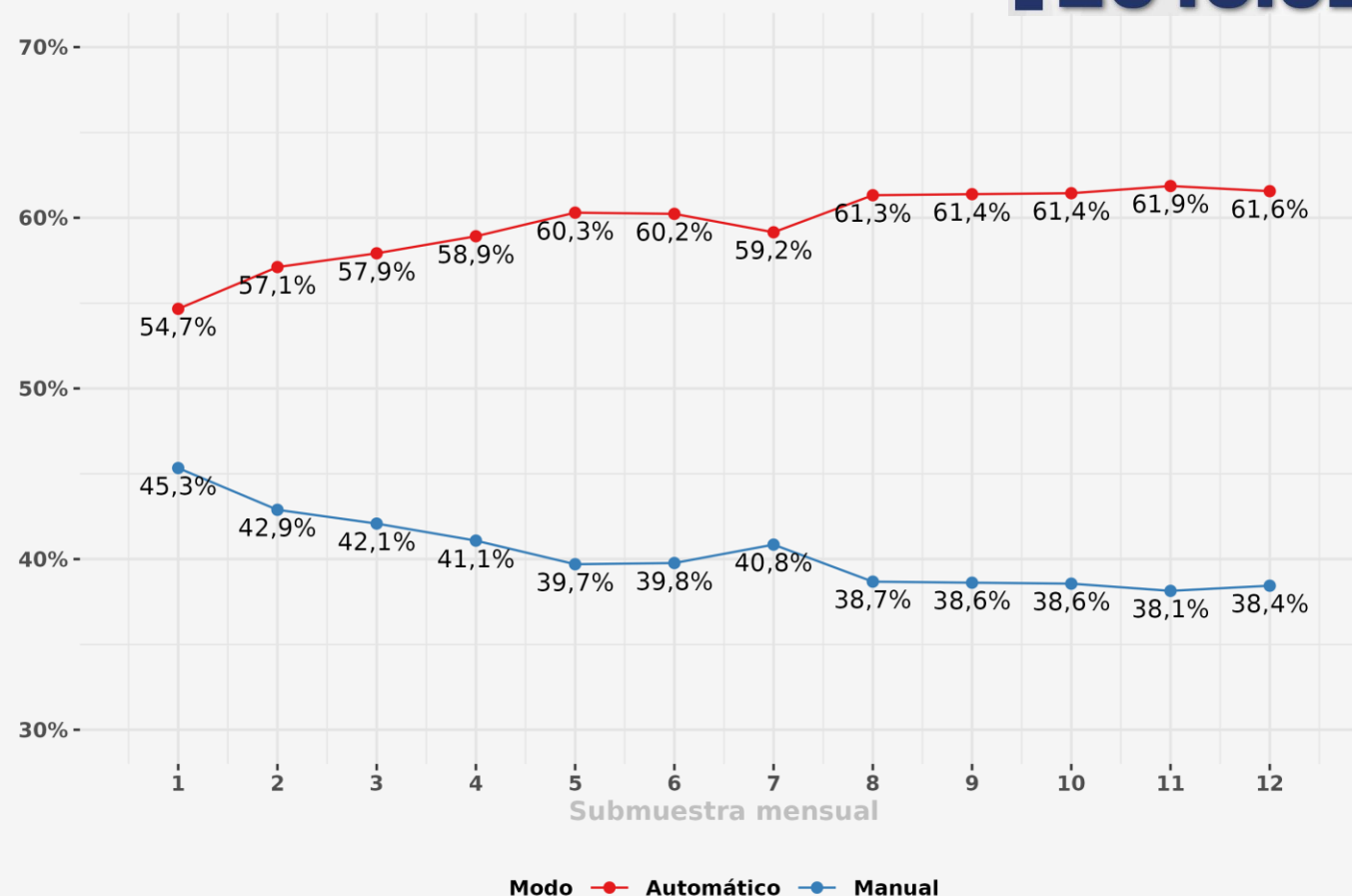
Lo que no fue codificado automáticamente fue codificado de forma manual por el equipo de codificación del proyecto.

Clasificador de Consumo Individual por Finalidades (CCIF 2018.CL)

Al revisar la codificación por submuestra mensual se puede ver una tendencia a mejorar la codificación dado los constantes ajustes realizados al código lo que mejoró la codificación en las últimas submuestras.

CCIF 2018.CL

CCIF 2018.CL



Clasificación Internacional Uniforme de Ocupaciones (CIUO 08.CL)

La **Clasificación Internacional Uniforme de Ocupaciones** (CIUO) permite **ordenar jerárquicamente y agrupar**, de acuerdo con características similares, diversas **ocupaciones** de acuerdo con el tipo de **tareas** realizadas en un puesto de trabajo y las **competencias** requeridas para ello, dependiendo del nivel y de la especialización de estas competencias.

A la fecha, se han publicado cuatro versiones de esta clasificación (CIUO 58, CIUO 68, CIUO 88 y CIUO 08), esta última versión de la clasificación no tan solo recoge las experiencias adquiridas luego del uso de su antecesora (CIUO 88), sino también los cambios y evolución del mundo del trabajo. La entidad encargada de la actualización de la clasificación es la **Organización Internacional del Trabajo** (OIT)

Fuente: CIUO 08.CL Clasificador Chileno de Ocupaciones.

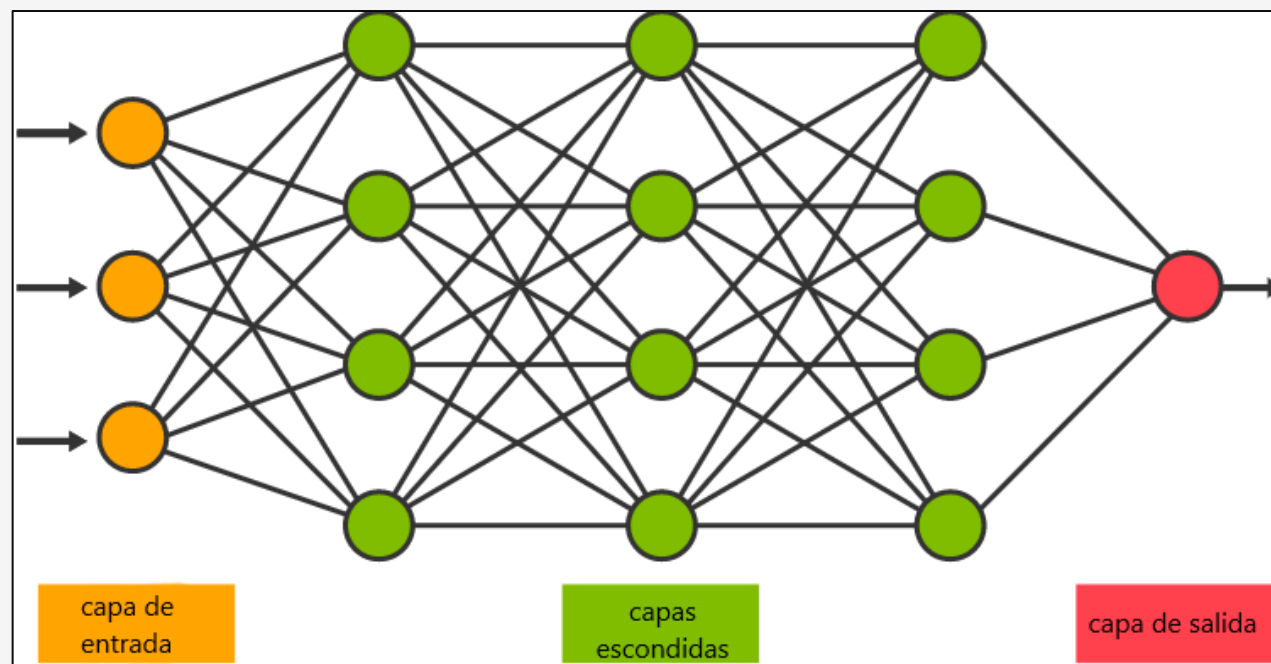
CIUO 08.CL
Clasificador Chileno
de Ocupaciones

 **IXEPF**

Clasificación Internacional Uniforme de Ocupaciones (CIUO 08.CL)

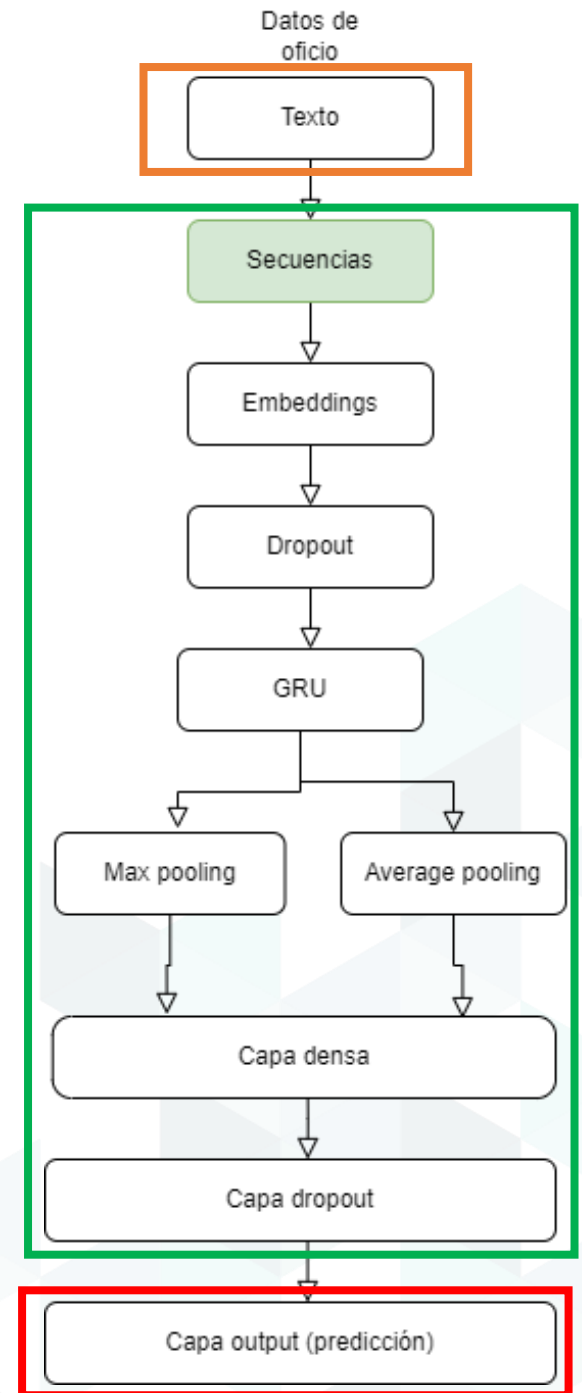
La codificación automatizada se realizó utilizando una **red neuronal** recurrente con puertas (GRU por sus siglas en inglés).

En términos simples, una red neuronal consiste en un modelo predictivo que se configura en 3 tipos de capas: la **capa de entrada**, que contiene los inputs del modelo, las **capas escondidas** que realizan transformaciones sucesivas a los datos y finalmente la **capa de salida**, que contiene las predicciones del modelo



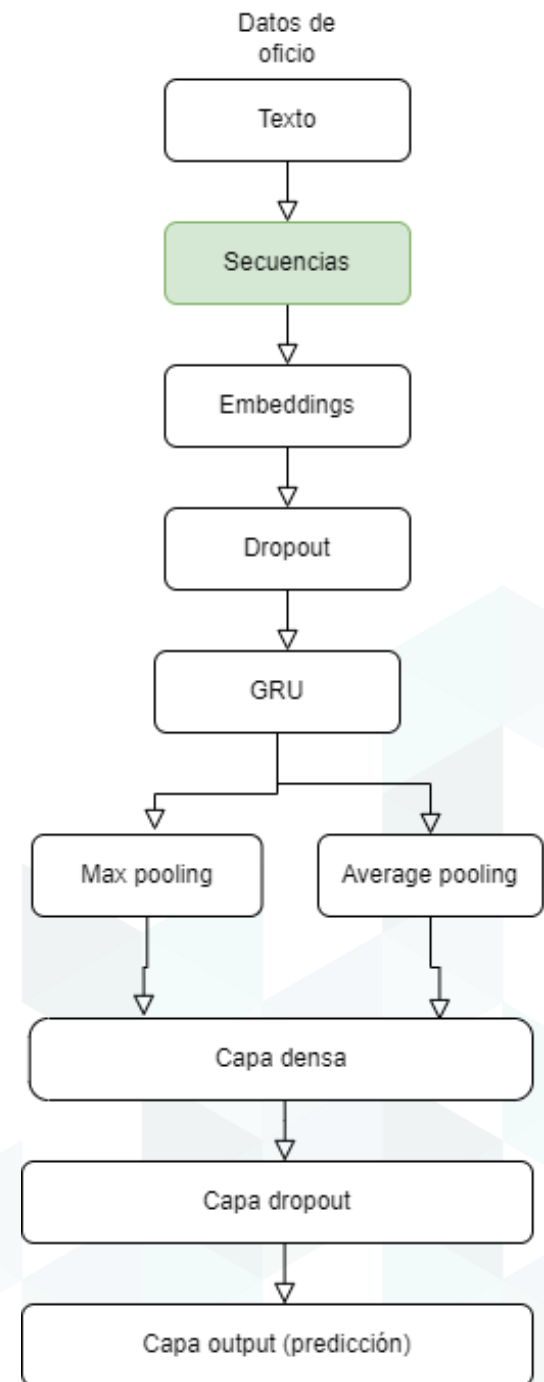
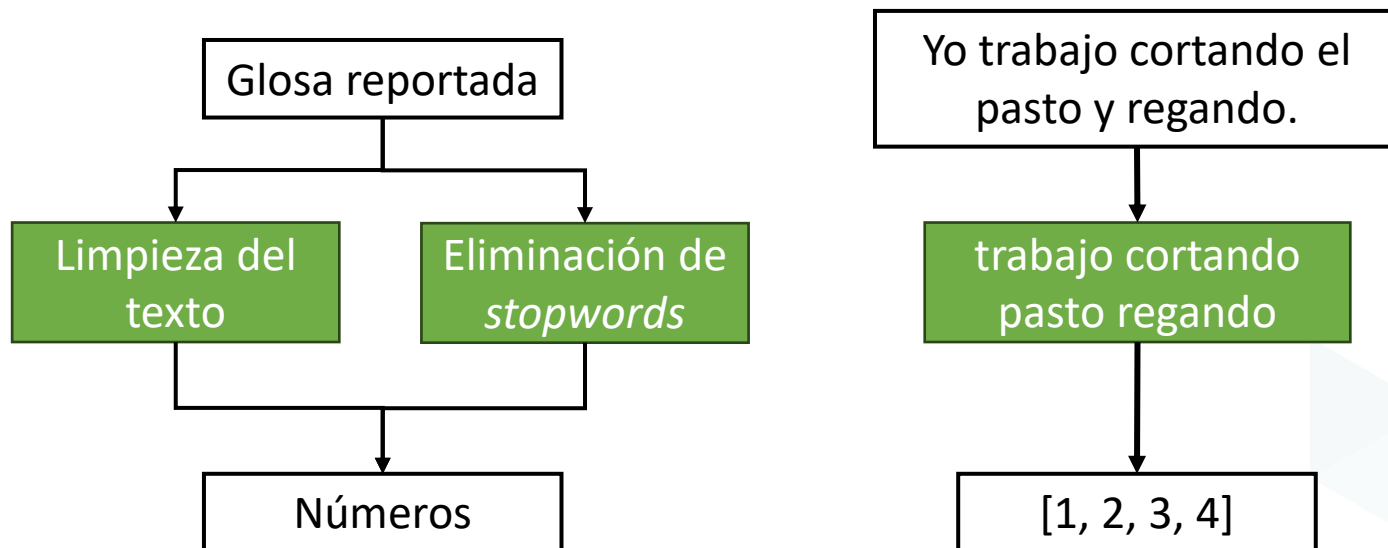
Clasificación Internacional Uniforme de Ocupaciones (CIUO 08.CL)

En el caso de CIUO, la capa de entrada contiene las glosas de oficio y tareas de cada informante que haya entregado información, las capas escondidas realizan transformaciones que finalmente llevan a las predicciones: por cada combinación de oficio y tarea se obtendrá un vector que contiene la probabilidad de que la glosa pertenezca a cada uno de los códigos CIUO-08.CL.



Clasificación Internacional Uniforme de Ocupaciones (CIUO 08.CL)

En el caso de CIUO, la capa de entrada contiene las glosas de oficio y tareas de cada informante que haya entregado información, las capas escondidas realizan transformaciones que finalmente llevan a las predicciones: por cada combinación de oficio y tarea se obtendrá un vector que contiene la probabilidad de que la glosa pertenezca a cada uno de los códigos CIUO-08.CL.

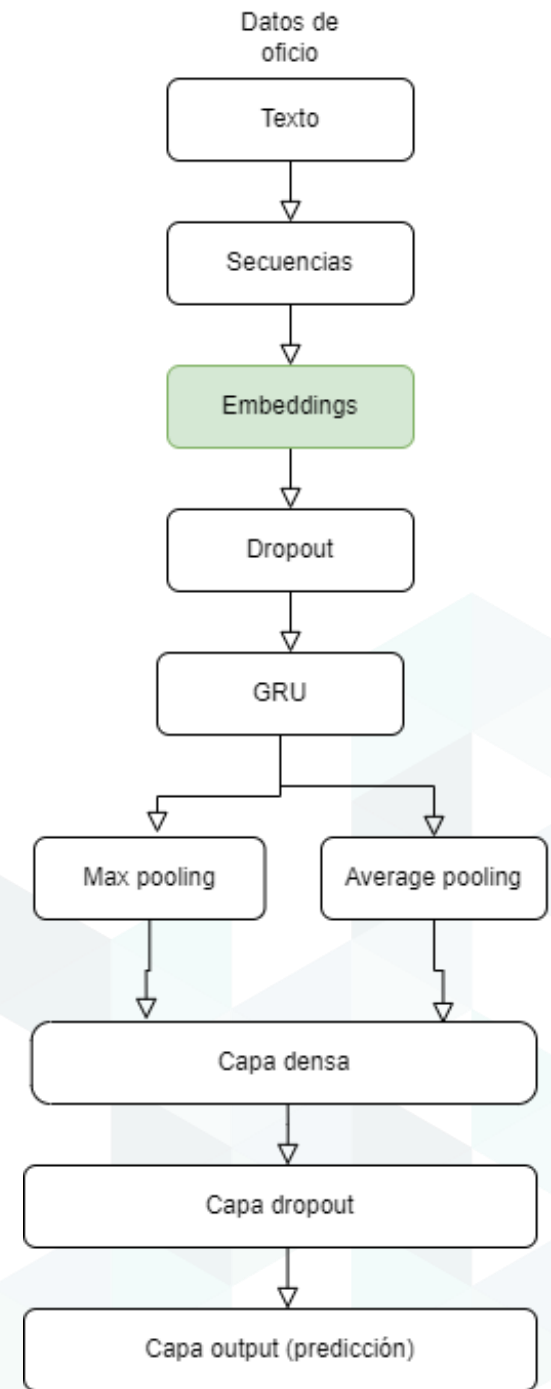


Clasificación Internacional Uniforme de Ocupaciones (CIUO 08.CL)

Para darle sentido al formato numérico se utilizó un modelo disponibilizado por el Profesor Jorge Pérez, que está entrenado con *corpora* que contiene aproximadamente 3 mil millones de palabras en español obtenidas de Wikipedia, ParaCrawl, OpenSubtitles, entre otros.



Usando Embeddings provenientes del modelo se máxima la información extraída de cada palabra en las descripciones

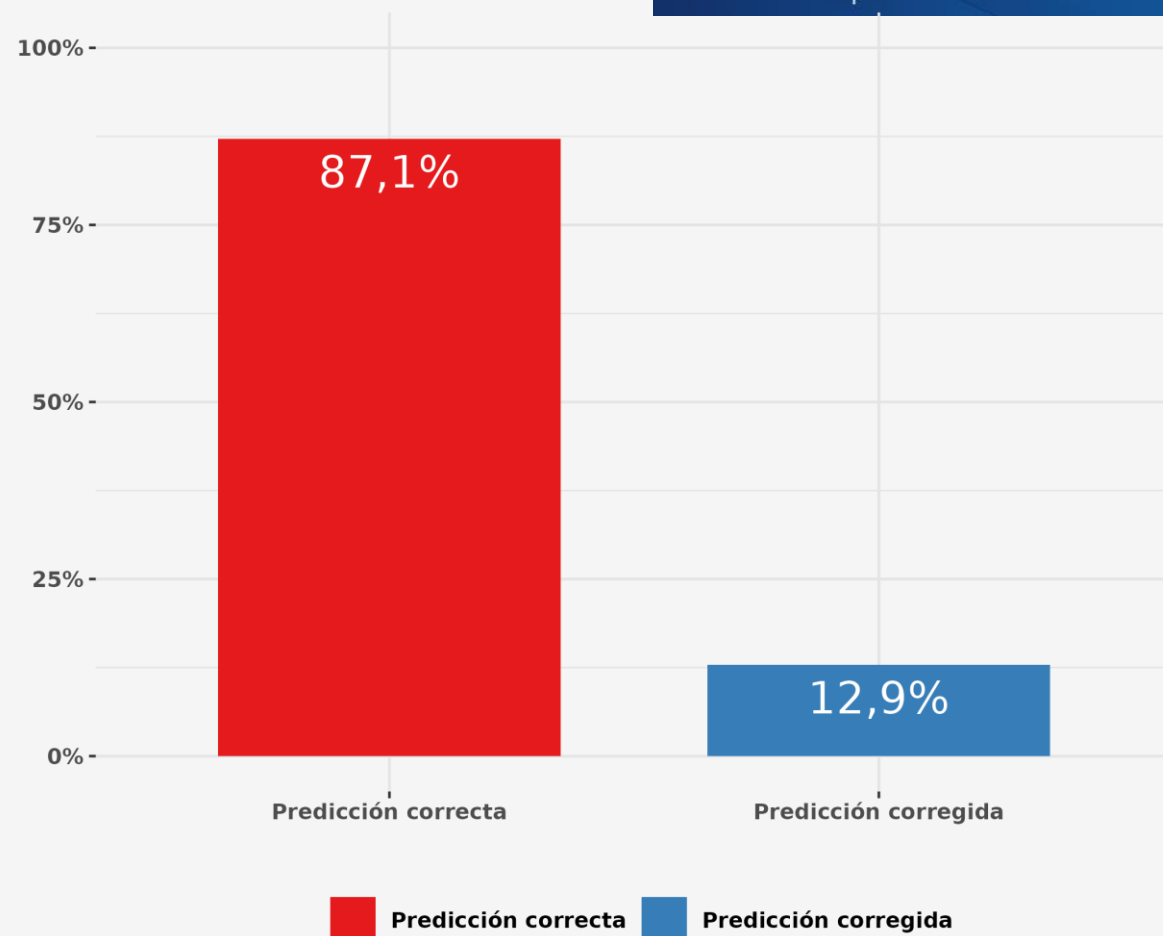


Clasificación Internacional Uniforme de Ocupaciones (CIUO 08.CL)

Todos los registros fueron codificados por el modelo y revisados por el equipo de analistas o el equipo de codificadores.

De esta revisión se concluye que 87,1% de los registros fue **codificado correctamente** por la predicción del modelo y 12,9% tuvo que ser **corregido** por el equipo de codificación.

CIUO 08.CL
Clasificador Chileno
de Ocupaciones



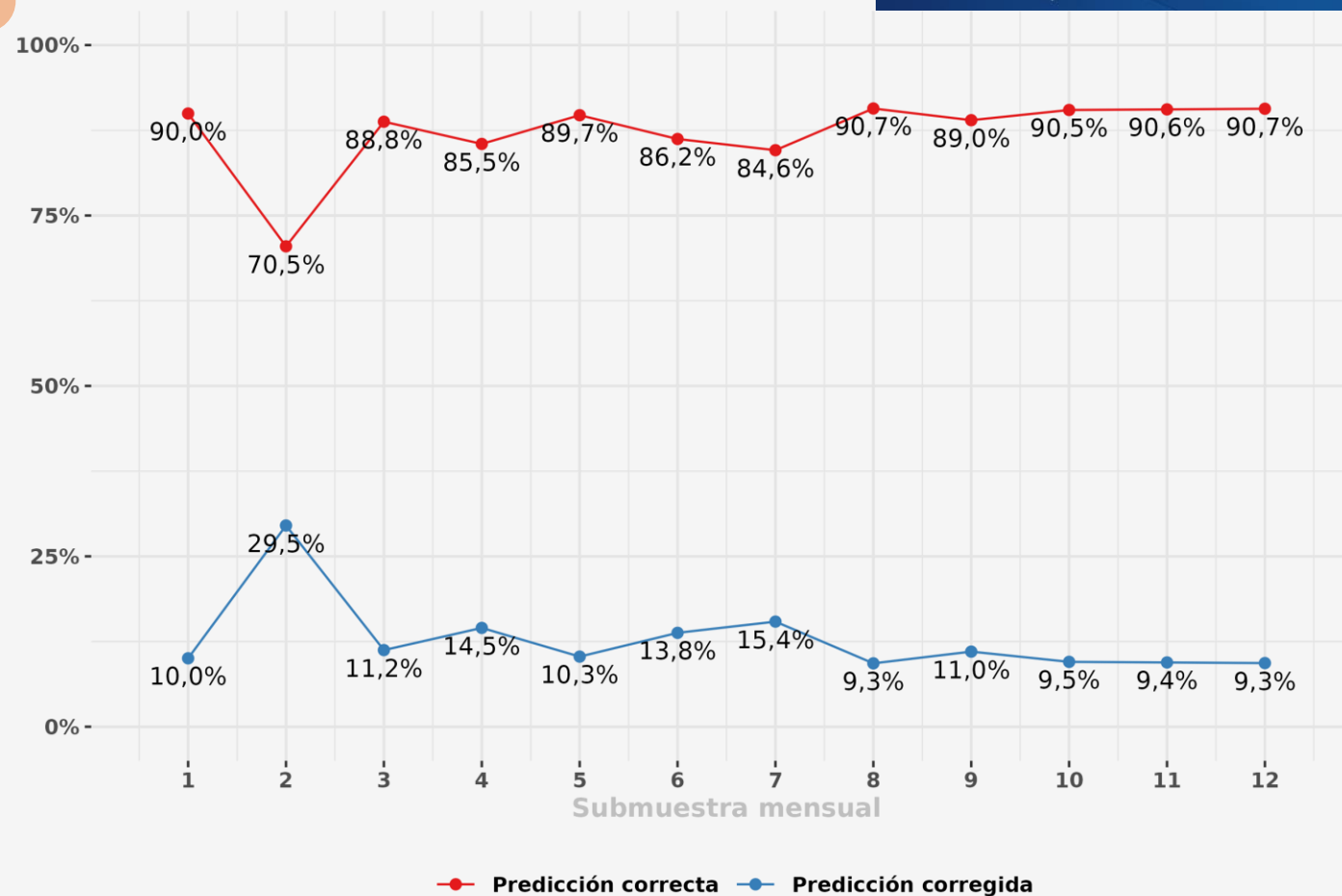
Clasificación Internacional Uniforme de Ocupaciones (CIUO 08.CL)

Al ver la distribución por submuestra mensual se puede ver que el modelo se comportó de forma similar durante todo el proceso **mejorando la predicción** desde el mes 8 en adelante.

En el segundo mes se ve una baja en la predicción correcta debido a ajustes internos del modelo y la falta de datos para el entrenamiento en un comienzo.

CIUO 08.CL

Clasificador Chileno de Ocupaciones





Experiencias y aprendizajes

Luego de la experiencia utilizando codificación automatizada esta se volvió **parte esencial del procesamiento** de los datos de la EPF.

- ✓ La codificación por redes neuronales solo puede mejorar en la medida que aumentan los datos codificados.
- ✓ La codificación por reglas tiene la **flexibilidad suficiente** para adaptarse a los cambios en como las personas nombran los distintos productos en el paso de 5 años.

El trabajo constante con el equipo de codificación fue fundamental tanto para la optimización de los distintos algoritmos de codificación automatizada como para alcanzar el objetivo de codificar todos los clasificadores.

Por esto, las mejoras a estos algoritmos de codificación automatizada no deben ser en desmedro del trabajo de codificación y revisión manual.



GRACIAS



ENCUESTA DE PRESUPUESTOS FAMILIARES 2021 - 2022

ine.gob.cl

