



Estudio Temático  
**Especial Demanda**  
**Laboral**



## Instituto Nacional de Estadística y Censos

### Autoridades:

#### Dirección Ejecutiva

Roberto Castillo Añazco.

#### Subdirección General

Jorge García-Guerrero

#### Coordinación General Técnica de Innovación en Métricas y Análisis de la Información

Lorena Moreno Enríquez

#### Dirección de Estudios y Análisis de la Información

Cristhian Rosales Castillo

### Revisión

Lorena Moreno Enríquez

Cristhian Rosales Castillo

### Elaboración:

Laboratorio de Dinámica Laboral y Empresarial – Proyecto de Reconversión de la Educación Técnica y Tecnológica Superior Pública del Ecuador (LDLE – PRETT):

#### Gestión del subcomponente de demanda laboral:

María Isabel García Mosquera

Natalia Garzón Durango

Cristhian Rosales Castillo

#### Equipo técnico

##### Especialista de demanda laboral

David Puebla Robles (Consultor PRETT)

##### Analistas de demanda laboral

Miguel Coral Carvajal (Consultor PRETT)

Bryan Barragán Pazmiño (Consultor PRETT)

### Diseño y Diagramación:

Esteban Palacios Jaramillo



## Volumen II

# Documento metodológico de la construcción de la matriz de empleo basada en registros administrativos

on the greatest opportunities to increase sales  
to increase sales and achieve the advantage  
activities of marketing that has to do with the  
objectives will be based on how you gain

	2016
	39,912
	108,287
	91,938
	125,819
	276,161
	11,627
	107,812

%Growth
-11%
+37%
+42%
+78%
+18%

2013 - 2016
June
45,556
1,876
190
134
47,756

Company's Growth

- Place
- Department Store
- Super Center
- Shopping Center
- E-commerce
- Specialty

\*According to sales

Opportunities

## Tabla de contenidos

---

1. Antecedentes .....	6
2. Introducción .....	6
3. Experiencias internacionales .....	8
4. Construcción de la matriz de empleo basada en registros administrativos.....	10
4.1.    Consideraciones generales.....	10
4.2.    Construcción de la variable de cargo ocupacional .....	11
4.2.1.    Sector privado.....	15
4.2.1.1.    Creación del cargo en registros con información disponible .....	15
4.2.1.2.    Creación del cargo en registros sin información disponible .....	19
4.2.1.3.    Composición ocupacional con registros completos.....	22
4.2.2.    Sector público: GADs y EPs.....	23
4.2.2.1.    Creación del cargo en registros con información disponible .....	23
4.2.2.2.    Creación del cargo en registros sin información disponible .....	24
4.2.2.3.    Composición ocupacional con registros completos.....	29
4.2.3.    Sector público: Ministerio de Finanzas .....	29
4.2.3.1.    Creación del cargo en registros con información disponible .....	29
4.2.3.2.    Creación del cargo en registros sin información disponible .....	31
4.2.3.3.    Composición ocupacional con registros completos.....	34
4.2.4.    Evolución de la composición ocupacional con información completa .....	35
4.3.    Matriz de empleo basada en registros administrativos .....	35
5. Conclusiones .....	39
6. Bibliografía .....	40
7. Anexos.....	42



## Índice de Figuras

---

Figura 1. Resumen de la construcción de la variable de cargo de ocupación .....	12
Figura 2. Esquema de arrastres del cargo de la base de Avisos de Entradas y Salidas a las bases mensuales de Afiliados del IESS.....	17
Figura 3. Periodos con información disponible de cargo para el sector privado del IESS .....	19
Figura 4. Distribución ocupacional con información disponible del cargo. Sector privado IESS. ..	20
Figura 5. Errores de clasificación individual de las predicciones para valores de k. Sector privado IESS. ....	21
Figura 6. Error relativo promedio de la composición ocupacional de las predicciones para valores de k. Sector privado IESS. ....	22
Figura 7. Error absoluto promedio de la composición ocupacional de las predicciones para valores de k. Sector privado IESS. ....	22
Figura 8. Distribución ocupacional con información completa del cargo. Sector privado IESS.....	23
Figura 9. Periodos con información disponible de cargo para GADs y EPs .....	25
Figura 10. Errores de clasificación individual de las predicciones para valores de k. GADs. ....	27
Figura 11. Error relativo promedio de la composición ocupacional de las predicciones para valores de k. GADs .....	27
Figura 12. Error absoluto promedio de la composición ocupacional de las predicciones para valores de k. GADs. ....	28
Figura 13. Distribución ocupacional con información completa del cargo. GADs municipales y provinciales. ....	29
Figura 14. Periodos con bases disponibles del Distributivo de Personal del MEF .....	31
Figura 15. Error de clasificación individual de las predicciones para valores de k. Sector Público.33	
Figura 16. Error relativo promedio de la composición ocupacional de las predicciones para valores de k. Sector público. ....	33
Figura 17. Error absoluto promedio de la composición ocupacional de las predicciones para valores de k. Sector público. ....	34
Figura 18. Distribución ocupacional con información completa del cargo. Sector público.....	34
Figura 19. Distribución ocupacional con información completa del cargo. Universo de afiliados al IESS. ....	35

## Índice de Tablas

---

Tabla 1. Variables incluidas en los modelos de clasificación del sector privado del IESS .....	20
Tabla 2. Grupos de ocupación para GADs y EPs .....	24
Tabla 3. Variables incluidas en los modelos de clasificación de GADs y EPs .....	26
Tabla 4. Variables incluidas en los modelos de clasificación para el sector público.....	32
Tabla 5. Matriz de empleo basada en registros administrativos - 2020.....	37
Tabla 6. Matriz de empleo basada en registros administrativos - 2014.....	38

## 1. Antecedentes

El 22 de diciembre de 2016, la República del Ecuador a través del Ministerio de Economía y Finanzas, y el Banco Internacional de Reconstrucción y Fomento (BIRF) suscribieron el Convenio de Préstamo LOAN NUMBER 8667-EC, para la realización del Proyecto de Reconversión de la Educación Técnica y Tecnológica Superior Pública del Ecuador – PRETT, el cual es ejecutado por la Secretaría Nacional de Educación Superior, Ciencia Tecnología e Innovación (SENESCYT) a través de la Entidad Operativa Desconcentrada PRETT (EOD PRETT), y en el que el Instituto Nacional de Estadística y Censos (INEC) actúa como co-ejecutor dentro del Componente 3.2b. El proyecto tiene como objetivo reconvertir la formación técnica y tecnológica superior pública del país mediante la dotación de infraestructura física, equipamiento y una oferta académica pertinente a los institutos técnicos y tecnológicos, con el fin de aportar al cambio de la matriz productiva del país y el fortalecimiento estratégico de los institutos técnicos y tecnológicos.

Con fecha 3 de marzo de 2017, la SENESCYT y el INEC suscriben el Convenio de Cooperación Interinstitucional Nro. 20170006, para el diseño e implementación de un sistema de información estadística sobre demanda laboral en el Marco del Proyecto “Apoyo a la Reconversión de la Educación Técnica y Tecnológica Pública del Ecuador”. El INEC, como co-ejecutor del proyecto PRETT, tiene como objetivo realizar una estimación de la demanda del mercado laboral ecuatoriano mediante información de encuestas y registros administrativos del Laboratorio de Dinámica Laboral y Empresarial (LDLE).

Con la finalidad de comprender adecuadamente la estructura del mercado laboral ecuatoriano, el primer gran objetivo fue la construcción de una matriz de empleo con información de registros administrativos. En este documento se expone la metodología utilizada para la obtención de dicha matriz, lo cual implicó un análisis a profundidad de la información disponible y un proceso de depuración que tuvo como fin contar con todas las variables necesarias, siendo la categoría ocupacional la variable más importante que no se encontraba disponible previo a este ejercicio.

## 2. Introducción

Las matrices de empleo tienen como fin resumir la estructura del mercado laboral específico de una economía en base a características como la distribución por sectores económicos, cargos de ocupación, características demográficas, nivel de calificación, horas trabajadas, salarios, entre otras. La disponibilidad de la información define el alcance que puede llegar a tener una matriz de empleo, tanto en cantidad de variables como cobertura del empleo, por lo que resulta necesario evaluar toda la información disponible y definir la mejor estrategia para la construcción de una matriz de empleo que represente el mercado laboral general.

Las fuentes de información que usualmente se utilizan para la construcción de una matriz de empleo son encuestas de hogares, encuestas empresariales y registros administrativos, las cuales

pueden reportarse desde el lado de la oferta cuando son los propios trabajadores quienes proveen información sobre su condición de empleo, y desde el lado de la demanda cuando las empresas reportan información relacionada a sus trabajadores. En el caso de las encuestas, es claro que la información proviene desde el lado de la oferta cuando se trata de una encuesta a hogares y de la demanda cuando se trata de una encuesta a empresas. Los registros administrativos, por su parte, se pueden construir a partir de la información específica de las personas y por lo tanto de los oferentes de empleo (e.g. información demográfica, nivel de instrucción), y desde el lado de la demanda cuando las empresas reportan el registro de un nuevo trabajador e información sobre los términos de contratación.

Este documento tiene como objetivo describir el proceso de construcción de una matriz de empleo basada en registros administrativos para el caso ecuatoriano a través de las fuentes de información disponibles en el Laboratorio de Dinámica Laboral y Empresarial (LDLE) del INEC con periodicidad mensual desde el 2008. Las fuentes utilizadas para este fin son, por parte de los empleadores (demanda laboral), la información de Avisos de Entradas y Salidas, y Afiliados del Instituto Ecuatoriano de Seguridad Social (IESS), Contratos del Ministerio del Trabajo (MDT), el Distributivo de Personal del Ministerio de Economía y Finanzas (MEF), la información de transparencia de los Gobiernos Autónomos Descentralizados (GADs) y Empresas Públicas (EPs), y las bases del Directorio de Empresas y Establecimientos (DIEE) del INEC y de Contribuyentes del SRI<sup>1</sup>. Por parte de los trabajadores (oferta laboral) se utilizó información de los Cedulados del Registro Civil.

Es necesario tomar en cuenta que la matriz de empleo resultante de este ejercicio cubre solamente al sector formal de la economía, definido como el empleo registrado en la seguridad social, del cual forma parte uno de cada tres trabajadores a nivel nacional. El desarrollo de una matriz de empleo nacional que cubre al sector informal de la economía y que se insume de otras fuentes complementarias de información como encuestas de hogares y empresariales se expone en el documento de metodología y resultados de las proyecciones de empleo, debido a que es usada específicamente para este fin. Un aspecto adicional que distingue a la matriz de empleo basada en registros administrativos y la matriz de empleo nacional es la posibilidad de obtener estadísticas desagregadas por varios grupos de interés, ya que esto se puede hacer sin problema en el primer caso, mientras que en el segundo la incorporación de información de encuestas impone una limitación en este sentido.

El uso estadístico y para fines analíticos de los registros administrativos requiere un tratamiento previo de la información, ya que es usual que la información inicialmente tenga campos no llenados, campos incompletos, errores de digitación, formatos inadecuados, etc. Adicionalmente, la calidad de la información de los registros administrativos suele mejorar en el tiempo, lo cual impone un reto al momento de contar con series de tiempo largas y comparables. La experiencia

---

<sup>1</sup> Las bases del DIEE del INEC y Contribuyentes del SRI permiten obtener la información de la actividad económica de las empresas. Ambas fuentes son complementarias, ya que en el DIEE se realiza una validación de la actividad económica de las empresas registradas el SRI, principalmente en base a las encuestas levantadas por el INEC.

generada en el manejo de registros administrativos dentro del LDLE<sup>2</sup> ha sido enormemente aprovechada para construir una matriz de empleo, destacándose principalmente el uso de variables ya disponibles de características personales, información específica de las empresas y relaciones laborales entre empleado y empleador. A pesar de ello, previamente no existía una variable de cargo ocupacional en los registros administrativos disponibles, lo cual limitaba la posibilidad de construir una matriz de empleo con esta dimensión. Por tal razón, las actividades del equipo técnico se centraron principalmente en la creación de una variable de cargo de ocupación a partir de información descriptiva disponible en varias fuentes de datos, proceso en el cual no solamente se completó la información del cargo en los registros con información disponible, sino también en los registros que no contaban con alguna descripción de cargo mediante el uso de modelos de clasificación y de otras variables disponibles.

El resto del documento está dividido en tres secciones. En la sección 3 se hace una síntesis breve de los esfuerzos realizados para la construcción de matrices de empleo en otros países. La sección 4 presenta una descripción de las fuentes de información utilizadas para el caso ecuatoriano, así como el tratamiento realizado para construir una variable de cargo ocupacional en los registros con información descriptiva del cargo, y los modelos de clasificación empleados para completar la información del cargo en los registros faltantes. Finalmente, en la sección 5 se presentan las conclusiones y recomendaciones que surgieron de esta actividad.

### 3. Experiencias internacionales

Las matrices de empleo por lo general se conciben como una parte de un proceso integral de los modelos de proyección de empleo, principalmente porque resumen su estructura y posibilitan la obtención de información desagregada en el tiempo. La Oficina de Estadísticas Laborales de Estados Unidos (Bureau of Labor Statistics-BLS en inglés), desarrolla periódicamente un conjunto de proyecciones de empleo resumidas en formatos de matrices de empleo a nivel nacional. Estas matrices se construyen con datos de empleo por ocupación, salarios, y tipo de empleado (asalariado, independiente y del hogar), a partir de varias fuentes: (1) Estadísticas de Empleo Ocupacional (Occupational Employment Statistics), información que se levanta semestralmente para producir estimaciones de empleo y salarios para 800 ocupaciones, (2) el programa de Estadísticas de Empleo (Current Employment Statistics), que recolecta información de los establecimientos sobre empleo no-agricultor (asalariado y por hora) y horas de trabajo semanales, (3) el Censo Trimestral de Empleo y Salarios (Quarterly Census of Employment and Wages) y (4) las Encuestas de Hogares (Current Population Survey) que recolectan información sobre los trabajadores asalariados, independientes y del hogar, incluyendo el sector agricultor. Con estos

---

<sup>2</sup> El LDLE del INEC se ha convertido en un referente a nivel nacional del uso de información enlazada de varias entidades públicas del país, lo cual ha requerido un proceso de varios años con la generación de convenios interinstitucionales para el traspaso de la información, y generación de procesos para la transformación de registros administrativos a registros estadísticos. Desde su creación, el sistema de información basado en registros administrativos que posee el LDLE se ha fortalecido de manera importante, sin embargo, un proceso de esta envergadura se debe fortalecer continuamente.



insumos y el uso de técnicas econométricas, el BLS construye su modelo de proyección de matrices de empleo, así como proyecciones del salario mensual y por hora (BLS, 2018<sup>3</sup>).

Zukersteinova, Strietska-Illina (eds.), 2007<sup>4</sup> realizan una descripción detallada de los procesos que siguen 14 países de Europa para construir matrices de empleo y proyectarlas en el tiempo. Dependiendo del país considerado, las bases de datos empleadas para construir la matriz de empleo y obtener la composición ocupacional pueden ser encuestas, censos o registros administrativos. Por ejemplo, en el Reino Unido se construyen matrices de empleo utilizando datos del Censo de Empleo (Census of Employment), la Encuesta Anual de Empleo (Annual Employment Survey), los datos de encuestas a empresas (Business Register and Employment Survey y Annual Business Inquiry) y la información del Censo Poblacional y de las Encuestas de Hogares para 75 industrias y 25 grupos ocupacionales desagregadas por género (Wilson, et al., 2016<sup>5</sup>).

A nivel regional, se pueden destacar experiencias como la del Departamento Administrativo Nacional de Estadística (DANE) de Colombia. Las fuentes de información empleadas para la construcción de la matriz de empleo son encuestas de hogares, encuestas empresariales y estadísticas de cuentas nacionales. Las etapas que se siguen para la construcción de la matriz de empleo son cinco: i) la reclasificación de los ocupados según tipo de contratación y naturaleza del empleo por rama de actividad; ii) el cálculo de empleo equivalente por rama de actividad, categoría ocupacional y zona geográfica; iii) la imputación de empleo para fuentes sin ocupación asociada; iv) el contraste entre oferta y demanda laboral (DANE, 2011<sup>6</sup>). Otra experiencia regional es la que efectúa el Instituto Nacional de Estadística e Informática (INEI) de Chile, en la que se puede destacar que, al igual de Colombia, la construcción de la matriz se basa en el cálculo de las horas trabajadas, el empleo equivalente, y la clasificación del empleo en categorías de ocupación y sectores institucionales.

Cada país tiene su propia infraestructura estadística, lo cual afecta directamente al alcance de las matrices de empleo que pueden construir. Por un lado, hay países como Estados Unidos que vienen trabajando con este tipo de información por más de 60 años, lo cual le permite contar con un proceso consolidado y resultados con alto nivel de desagregación. Sin embargo, por otro lado, también hay países que aún presentan limitaciones en la disponibilidad de la información y los niveles de desagregación, como es el caso de los países de la región.

---

<sup>3</sup> Ver [más](#).

<sup>4</sup> Ver [más](#).

<sup>5</sup> Ver [más](#).

<sup>6</sup> Ver [más](#).

## 4. Construcción de la matriz de empleo basada en registros administrativos

### 4.1. Consideraciones generales

En el caso ecuatoriano, la construcción de la matriz de empleo nacional se realizó en dos etapas. En primer lugar, se construyó una matriz de empleo formal basada en registros administrativos, en donde las tareas se enfocaron en la construcción de la variable del cargo de ocupación, y posteriormente una matriz de empleo nacional incorporando información de encuestas de hogares y empresariales. En este documento se expone el proceso llevado a cabo en la primera etapa, en donde el uso exclusivo de registros administrativos conlleva ventajas frente al uso de fuentes tradicionales de encuestas en aspectos como la periodicidad de la información, la sostenibilidad en el tiempo, bajo costo y posibilidad de obtener resultados desagregados por grupos de interés al no depender de un diseño muestral, aunque con la principal limitación de que solo cubre al sector formal de la economía y la no disponibilidad de variables relevantes como horas y días trabajados<sup>7</sup>.

Hasta el momento, la fuente empleada por analistas e investigadores para describir el mercado laboral ecuatoriano ha sido la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU), la cual a pesar de ser la fuente oficial de las estadísticas de empleo del país y de permitir identificar a los trabajadores del sector formal e informal, su diseño muestral no necesariamente garantiza que las desagregaciones de los resultados sean precisas<sup>8</sup>. En este sentido, el hecho de usar fuentes de información alternativas que permitan conocer con mayor precisión la estructura del mercado laboral de un subconjunto de trabajadores como aquellos registrados en la seguridad social se convierte en una mejor alternativa que usar una fuente única de información como la ENEMDU.

En el sentido más básico, una matriz de empleo debe permitir describir la estructura del mercado laboral en dos dimensiones: el sector económico de las empresas donde se encuentran los trabajadores, y su cargo de ocupación<sup>9</sup>. Para el caso ecuatoriano, los sectores económicos se clasificaron en base a la Clasificación Industrial Internacional Uniforme Rev.4 (CIIU), y los grupos ocupacionales en base a la Clasificación Internacional Uniforme de Ocupaciones 2008 (CIUO) (Anexos 1 y 2).

---

<sup>7</sup> Si bien en la base de afiliados al IESS hay información de días, estos no necesariamente corresponden a los días trabajados durante el mes. Por ejemplo, una persona que trabajó todos los días del mes a medio tiempo puede tener un registro de 15 o 30 días trabajados.

<sup>8</sup> La variable de diseño de la ENEMDU es el desempleo, lo cual no garantiza la precisión de las estimaciones de otros indicadores laborales como el empleo formal, o el empleo por cargo de ocupación y sector económico.

<sup>9</sup> Adicionalmente, existen otras variables que pueden complementar la lectura del mercado laboral como características individuales de los trabajadores (e.g. sexo, edad o nivel de instrucción), características de la empresa (e.g. ubicación geográfica) y características de la relación laboral (e.g. sueldo, días y horas trabajadas).

Si se contase con una base de datos de trabajadores con variables de rama de actividad y cargo de ocupación, la matriz de empleo se podría obtener directamente con el cruce de ambas variables. Con relación a la variable de sector económico, el LDLE insume esta información del Directorio de Empresas y Establecimientos (DIEE) del INEC y la complementa con el catastro de Contribuyentes del Servicio de Rentas Internas (SRI)<sup>10</sup>. Sin embargo, en el caso del cargo ocupacional no se contaba previamente con una variable, razón por la cual se trabajó en una estrategia para construirla.

En la siguiente subsección se describe el proceso de construcción de la variable de cargo ocupacional y posteriormente se muestra el resultado obtenido al realizar el cruce con la variable de la rama de actividad, lo cual da como resultado una matriz de empleo basada en registros administrativos.

## 4.2. Construcción de la variable de cargo ocupacional

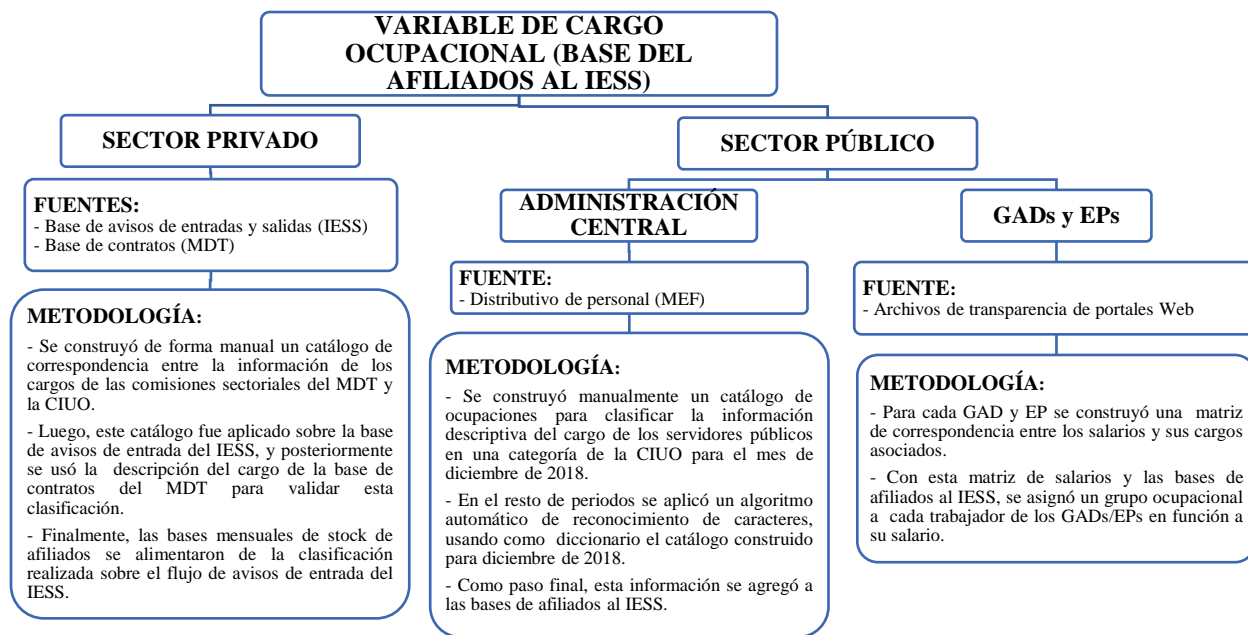
La información de grupo ocupacional no se encontraba disponible en el LDLE previo al desarrollo de este proyecto. Con la finalidad de atender esta necesidad, se realizó un trabajo intensivo de procesamiento, limpieza, integración y utilización de varias bases de datos que cuentan con información de cargo ocupacional. El objetivo fue usar varias fuentes de información (Anexo 3) para construir una variable de grupo ocupacional que se integró al sistema de información del LDLE a través de una variable adicional en las bases mensuales de Afiliados al IESS de 2008 a 2020.

La Figura 1 resume en un esquema el proceso que se siguió para recuperar la información del cargo de ocupación, destacándose principalmente que el proceso de recuperación de la información del cargo ocupacional se realizó a través de dos estrategias independientes para el sector privado y público.

---

<sup>10</sup> Si bien, una empresa puede dedicarse a más de una actividad económica, para fines analíticos el LDLE toma en cuenta solamente la actividad principal.

Figura 1. Resumen de la construcción de la variable de cargo de ocupación



*Fuente: Elaboración propia.*

Una vez que se realizó el proceso de construcción de la variable de cargo ocupacional con la información disponible, hubo un conjunto de trabajadores afiliados al IESS del sector público y privado para los cuales no se pudo conocer el grupo ocupacional. Adicionalmente, el hecho de que algunas fuentes de información no hayan estado disponibles para todos los meses desde el año 2008 también generó una pérdida de información que impidió en principio contar con la variable de cargo ocupacional en todos los periodos. Con la finalidad de construir una variable de ocupación para los trabajadores afiliados al IESS en todos los meses desde el año 2008 se aplicaron algoritmos de clasificación basados en árboles (Box 1).

Las aplicaciones de los modelos de clasificación, incluyendo los basados en árboles, se extienden a varios campos. Para el sector financiero, se han probado modelos de k-vecinos, regresión logística, *Naïve-Bayes*, árboles de decisión y *Random Forest*, con el fin de predecir la probabilidad de que se produzca una transacción fraudulenta con tarjeta de crédito de clientes europeos, alcanzando una tasa de precisión de 85,1% (Khatri, Arora y Agrawal, 2020). La genética es otro campo de aplicación donde ha cobrado relevancia la aplicación de modelos de *machine learning*, en donde la aplicación de modelos de árboles de decisión ha dado como resultado una precisión de 73,3% de exactitud para predecir un evento de malaria en Kenia a partir de la secuencia de ARN de este parásito (Arowolo et al., 2020). En el campo de la salud, se ha logrado predecir el fallecimiento por enfermedades cardíacas con base en variables morfológicas de las personas (edad, altura, peso y género), otras referentes a su estado de salud (presión sistólica, diastólica, colesterol, glucosa, y si tiene alguna enfermedad cardíaca) y sus hábitos diarios (fuma, bebe

alcohol y si realiza ejercicio físico), con una tasa de exactitud de 73% al trabajar con modelos de árboles de decisión (Princy et al., 2020).

En el caso ecuatoriano se puede destacar la aplicación de métodos de clasificación en campos como la demografía para predecir la densidad poblacional de la Amazonía del Norte del Ecuador en base a la degradación de imágenes satelitales con un porcentaje de casos correctamente clasificados de 74% (Santos, Graw y Bonilla, 2019). En el campo medioambiental, los modelos de clasificación han sido empleados para categorizar la cobertura vegetal de las provincias de Loja y Zamora Chinchipe a través de imágenes satelitales, logrando niveles de precisión del 90% (Tapia-Armijos et al., 2015). En la salud, se han probado modelos de clasificación basados en máquinas de soporte vectorial, regresión lineal múltiple, redes neuronales y *Random Forest*, utilizando imágenes satelitales y variables medioambientales para determinar la prevalencia de enfermedades crónicas respiratorias en hospitales de la ciudad de Quito (Alvarez-Mendoza et al., 2020). En el área de la educación superior se destaca un estudio en el cual se crea un modelo de clasificación basado en árboles de decisión para predecir la deserción estudiantil a partir de variables sociales y educativas de los estudiantes, alcanzando un nivel de exactitud de 91,7% de exactitud (Albán y Sánchez, 2018).

### Box 1. Métodos de clasificación basados en árboles

Los métodos basados en árboles involucran segmentaciones del espacio en  $J$  regiones:  $R_1, R_2, \dots, R_J$ . Las segmentaciones se originan a partir de la fijación de puntos de corte con las variables explicativas elegidas, los cuales se determinan a través de la minimización de una función de costo. Una opción es minimizar la tasa de error de clasificación:

$$E = 1 - \max_k(\hat{p}_{mk})$$

donde  $\hat{p}_{mk}$  es la proporción de observaciones de una muestra de entrenamiento en la región  $m$  que pertenecen a la clase  $k$ .

Otras funciones de costo se basan en medidas de “pureza” que penalizan a los criterios que no aportan información significativa al modelo como es el caso del Índice de Gini:

$$G = \sum_{k=1}^K \hat{p}_{mk}(1 - \hat{p}_{mk})$$

Así, cuando  $\hat{p}_{mk} = 0,5$ , el clasificador divide a las observaciones por partes iguales sin brindar información relevante al modelo, y por ello el valor de la función de costo es el más alto.

Dentro de la familia de métodos basados en árboles, existen varias metodologías como *Bagging* y *Random Forest*<sup>11</sup>:

- *Bagging*: Se generan “B” bases de datos “*Bootstrapped*”, es decir, se obtiene B muestras con reemplazos de la base de entrenamiento, y por cada base muestreada, se realiza un árbol, para finalmente realizar promedios de las predicciones obtenidas por cada muestra. Esto se realiza para disminuir la varianza del árbol y tener estimaciones más consistentes.
- *Random Forest*: Este caso, sigue la misma idea de *Bagging*, pero además de obtener una muestra de forma “*Bootstrapped*”, también genera un muestreo de  $k$  variables que utiliza.

<sup>11</sup> *Tree Pruning* y *Gradient Boosting* son otras metodologías que pertenecen a la misma familia.



De este modo, un modelo *Random Forest* que considera a todas las variables es equivalente a un modelo *Bagging*.

Para fines del presente trabajo se realizó la predicción del cargo ocupacional de manera separa para cada subuniverso (Figura 1) evaluando modelos *Random Forest* con una cantidad distinta de variables que el modelo elige de forma aleatoria para determinar los criterios de clasificación a fin de elegir el parámetro con los menores errores de clasificación<sup>12</sup>. La validación de los modelos se realizó dividiendo la población total de cada subgrupo con información observada en una muestra de entrenamiento (75%) y prueba (25%). Con los cargos ocupacionales predichos se evaluó tanto la precisión a nivel de microdato como la distribución ocupacional agregada. El primer caso involucra una comparación directa de los cargos clasificados con el modelo y los cargos observados (fórmula 1), mientras que en el segundo se agregó la información observada y predicha en vectores de distribución ocupacional a fin de comparar su similitud en términos absolutos (fórmula 2) y relativos (fórmula 3) siguiendo un enfoque de error cuadrático medio.

**Error de clasificación individual para cada valor  $k$  en el año  $t$ :**

$$Error\_Individual_{kt} = \frac{1}{N} \sum_{i=1}^N I(\hat{y}_i = y_i)$$

Donde  $N$  es el total de registros e  $I(\cdot)$  es una función indicadora que toma el valor de 1 si el valor observado y predicho de la ocupación del trabajador  $i$  es el mismo y 0 en caso contrario.

**Error absoluto para cada valor  $k$  en el año  $t$ :**

$$Error\_Absoluto\_Medio_{kt} = \sqrt{\frac{1}{9} \sum_{i=1}^9 (N\_pred_{ikt} - N\_obs_{it})^2}$$

**Error relativo para cada valor  $k$  en el año  $t$ :**

$$Error\_Relativo\_Medio_{kt} = \sqrt{\frac{1}{9} \sum_{i=1}^9 (N\_pred_{ikt}/N\_obs_{it} - 1)^2}$$

Donde  $N\_pred_{ikt}$  es el total de registros clasificados en el grupo ocupacional  $i$  por el modelo y  $N\_obs_{it}$  es el total de registros clasificados en el grupo  $i$  según la información observada.

Finalmente, con estos insumos se identificó el modelo que arroja los mejores resultados en términos de precisión a nivel de microdato y agregado, y se evaluó el trade-off entre precisión a nivel de microdato y en términos agregados cuando no existió un modelo superior en ambos criterios.

Las siguientes subsecciones describen el proceso realizado para obtener la variable del cargo ocupacional en cada subuniverso de trabajadores. En primer lugar, se resume el proceso de

---

<sup>12</sup> La elección de los modelos también se hizo en base a un análisis de consistencia temporal de la información y priorizando estrategias que no demanden una alta carga computacional por la cantidad de registros disponibles en las fuentes de datos utilizadas.

recuperación del cargo para aquellos registros con información disponible del cargo. Posteriormente se muestran los resultados del proceso de recuperación del cargo ocupacional para aquellos registros en los que no se pudo asignar un cargo en la primera etapa. Finalmente se muestra el resultado obtenido una vez que se completa la información del cargo para cada subuniverso y el universo total de trabajadores afiliados al IESS.

#### 4.2.1. Sector privado

##### 4.2.1.1. Creación del cargo en registros con información disponible

###### *Depuración y uso de la base de Avisos de Entradas y Salidas del IESS*

La base de datos de Avisos de Entradas y Salidas es una base histórica que contiene todos los registros de entradas y salidas que los empleadores del sector privado han reportado desde enero de 2006 hasta diciembre de 2020. Cada registro de esta fuente de información se genera cuando un empleador declara que un trabajador se integra a su empresa o cuando uno deja de formar parte de esta<sup>13</sup>. En el proceso de registro, el empleador debe proveer información del cargo que ocupará la persona contratada a partir de una lista de ocupaciones, las cuales provienen de los acuerdos logrados en las comisiones sectoriales que mantiene el Ministerio de Trabajo con otros actores civiles<sup>14</sup> en cada año con la finalidad de definir el salario mínimo por cada cargo ocupacional.

La clasificación de los cargos de ocupación elaborada por las comisiones de cada sector en las tablas de salarios mínimos contiene la descripción del cargo, la jerarquía del grupo (estructura ocupacional) y algunos comentarios adicionales<sup>15</sup>. Sin embargo, el principal objetivo de las comisiones sectoriales no es precisamente clasificar las ocupaciones en base a una nomenclatura homologada internacionalmente como la CIUO, sino fijar los salarios mínimos para cada una de las ocupaciones identificadas en las mesas técnicas. Esto representa un inconveniente para la generación de una variable codificada del cargo ocupacional, ya que los grupos ocupacionales definidos en las comisiones sectoriales no responden a una metodología de clasificación determinada y no son estrictamente comparables con los grupos ocupacionales de la CIUO.

La falta de correspondencia exacta entre los cargos identificados por las comisiones sectoriales y la CIUO generó varias dificultades al momento de construir una variable de cargo de ocupación codificada. Algunas descripciones de ocupación de las comisiones sectoriales no poseen suficiente información para poder clasificarlas con facilidad en alguna categoría de la CIUO: por ejemplo, la

---

<sup>13</sup> El Art. 2 de La Ley de Seguridad Social determina que todas las personas que perciben ingresos por la ejecución de una obra o la prestación de un servicio físico o intelectual, con relación laboral o sin ella, son sujetos "obligados a solicitar la protección" del Seguro General Obligatorio. Esto implica que todos los trabajadores deberían estar afiliados a la Seguridad Social según la Ley ([link](#)).

<sup>14</sup> El Consejo Nacional del Trabajo y Salarios (Consejo Nacional de Salarios antes de 2016) está conformado por i) El Ministro del Trabajo o por delegación el Viceministro de Trabajo y Empleo; ii) dos representantes de las organizaciones de empleadores -confederaciones o federaciones legalmente reconocidas; y iii) dos representantes de las organizaciones más representativas de las personas trabajadoras a nivel nacional (Acuerdo MDT-2015-0240).

<sup>15</sup> Ver salarios mínimos sectoriales de 2021 ([link](#)).

descripción “vendedor” no cuenta con información precisa para determinar si pertenece a los grupos de técnicos (grupo 3 de la CIUO) o vendedores o trabajadores de los servicios (grupo 5 de la CIUO). Adicionalmente, existe otro tipo de descripciones de ocupaciones que son bastante extensas y que poseen características consistentes con más de una clasificación: por ejemplo, en las ocupaciones relacionadas a labores agrícolas, las descripciones de las ocupaciones tienen elementos de agricultura básica que se categorizarían en el grupo de ocupaciones elementales (grupo 9 de la CIUO), pero también actividades más tecnificadas según las cuales deberían clasificarse en el grupo de agricultores (grupo 6 de la CIUO).

Tomando en cuenta estas limitaciones, se elaboró un diccionario de clasificación para asignar un grupo ocupacional de la CIUO a cada una de las descripciones de los cargos ocupacionales de las comisiones sectoriales. No obstante, debido a los inconvenientes presentados al momento de realizar la codificación de ciertos cargos ocupacionales, sobre todo con las descripciones que no pueden ser asignadas con precisión a un grupo de la CIUO, se emplearon fuentes complementarias como la base de contratos del MDT<sup>16</sup>.

### *Proceso de arrastre de la información del cargo en las Bases de Afiliados al IESS*

El registro de Afiliados al IESS es una base de stock con periodicidad mensual que contiene el registro de todas las personas que han cotizado a la seguridad social en un determinado mes, ya sea bajo relación de dependencia, independientes, voluntarios, trabajo no remunerado del hogar o semicontribuyentes<sup>17</sup>, empleo doméstico o seguro campesino.

La base de Afiliados al IESS define el universo del empleo formal sobre el cual el LDLE realiza un monitoreo continuo y en el cual se agregan variables de otras fuentes a fin de realizar una caracterización más amplia de los individuos o empresas<sup>18</sup>. La construcción de la variable del cargo ocupacional fue pensada justamente con la idea de agregar una dimensión adicional a esta fuente de información y que pueda ser empleada para realizar análisis históricos y coyunturales. Sin embargo, para lograrlo fue necesario partir de la construcción de la variable del cargo ocupacional realizada para cada relación laboral registrada en la base de Avisos de Entradas y Salidas, y luego replicar esta información en la base de Afiliados al IESS en los meses en los cuales esta relación laboral siguió vigente, ya que la información del cargo de ocupación solo se encuentra disponible en la primera (Figura 2).

---

<sup>16</sup> Otra fuente empleada para validar la clasificación del cargo fue el Manual de Codificación de la Encuesta Nacional de Empleo, Desempleo y Subempleo – ENEMDU 2019, el cual contiene los criterios para codificar la información descriptiva del cargo levantada en campo a la clasificación CIUO.

<sup>17</sup> Este régimen se creó mediante La Ley Orgánica para la Justicia Laboral y Reconocimiento del Trabajo no remunerado, publicada en el Suplemento de Registro Oficial No. 483 de 20 de abril de 2015

<sup>18</sup> Previo al uso de las bases de Afiliados al IESS, el LDLE realiza un tratamiento inicial que consiste en: (a) eliminar los afiliados voluntarios; (b) priorizar la relación laboral con el sueldo más alto en caso de que una persona tenga más de una relación laboral; (c) eliminar los registros de empleo doméstico cuando una persona tenga una relación de dependencia bajo esta modalidad y una adicional en otra actividad económica. Este primer paso permite contar con una base de datos depurada a nivel de relación laboral con periodicidad mensual.

Figura 2. Esquema de arrastres del cargo de la base de Avisos de Entradas y Salidas a las bases mensuales de Afiliados del IESS



Fuente: Elaborado por los autores.

### Base de Contratos del Ministerio de Trabajo

El MDT es la institución rectora de las políticas públicas de mercado laboral del país que regula y controla el cumplimiento de las obligaciones laborales establecidas en las normas laborales vigentes. Su actividad de control establece la necesidad de contar con sistemas de información que permitan monitorear el cumplimiento de las obligaciones de los empleadores y trabajadores, razón por la cual gestiona registros relacionados a los pagos del décimo tercer y décimo cuarto sueldo, la entrega de utilidades, el registro de contratos, actas de finiquito, entre otros.

Para efectos de la construcción de la variable del cargo ocupacional de los trabajadores del sector privado afiliados al IESS, la base de Contratos que gestiona el MDT fue usada como un recurso adicional para complementar la información del cargo de la base de Avisos de Entradas y Salidas del IESS. Esta base de datos es un registro histórico que posee todos los contratos del sector privado que han sido registrados en el MDT<sup>19</sup> y que provee información de varias características contractuales de una relación laboral empleado-empleador<sup>20</sup>. Al igual que la base de Avisos de Entradas y Salidas del IESS, un registro en la base de Contratos del MDT se origina cuando el empleador registra el ingreso de un trabajador, proceso en el cual reporta en un campo abierto la información del cargo que el nuevo trabajador ocupará. Esto último permite suponer que la información del cargo que se reporta en esta fuente es más precisa y detallada, razón por la cual se usó para solucionar las descripciones ambiguas obtenidas previamente a partir de la base de Avisos de Entradas y Salidas del IESS. El proceso de validación de la información disponible originalmente en el registro de Avisos de Entradas y Salidas consistió en cuatro partes:

<sup>19</sup> A pesar de ser un registro histórico, la cobertura de la base de Contratos mejora en el tiempo. Entre 2015 y 2020, el número promedio de contratos registrados fue de aproximadamente 800.000, mientras que de 2010 a 2014, el número se situaba en alrededor de 80.000. Dada esta evolución, la base de Contratos fue empleada como una fuente de datos auxiliar para depurar ciertos cargos ambiguos de la base de Avisos de Entradas y Salidas del IESS, y con ello mejorar la calidad de la variable del cargo ocupacional.

<sup>20</sup> Bajo el Acuerdo Ministerial MDT-2015-0098, todos los empleadores del sector privado están obligados a registrar los contratos del personal en el sistema SUT del MDT.

- 1. Trabajadores con cargos de “trabajador en general”:** Inicialmente, los trabajadores con esta descripción de ocupación se clasificaron en el grupo 9 de la CIUO (ocupaciones elementales) y representaron el 16% de los trabajadores registrados del sector privado en diciembre de 2020. Al verificar la información de estos casos con la información del MDT para el mes de diciembre de 2020 se logró modificar solamente el 1% de los casos clasificados como trabajadores en general.
- 2. Trabajadores clasificados en el grupo 1 (directores y gerentes) y que en la base de Contratos del MDT poseen un cargo ocupacional diferente:** Para corregir la clasificación de este grupo se utilizó un algoritmo de *machine learning*<sup>21</sup> para predecir la clasificación de ocupación usando la descripción de ocupación de la base del MDT, y tomando como información de aprendizaje la clasificación de la Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU). Las observaciones que según el algoritmo permanecieron el grupo 1 de la CIUO se acogieron automáticamente. Para aquellas en que se predijo un grupo ocupacional diferente, se realizó una revisión manual antes de asignar una categoría según las descripciones de la base de Contratos del MDT. En total se modificó el cargo codificado del 0,35% de trabajadores previamente clasificados como directores en diciembre de 2020.
- 3. Trabajadores clasificados en el grupo 6 (agricultores y trabajadores calificados agropecuarios, forestales y pesqueros):** Desafortunadamente, para la mayor parte de los casos (90.897 o 5% de trabajadores de diciembre de 2020) no se contó con suficiente información para distinguir qué trabajadores agrícolas pueden pertenecer a los grupos 6 (agricultores) o 9 (ocupaciones elementales). Sin embargo, a fin de guardar consistencia con la distribución ocupacional en el sector de agricultura de la ENEMDU, estos casos fueron clasificados en el grupo 6.
- 4. Trabajadores con cargos relacionados a vendedores:** La CIUO clasifica a los trabajadores relacionados con actividades comerciales en el grupo 5 (trabajadores de los servicios y vendedores de comercios y mercados), y a los vendedores de productos técnicos y/o ventas a nivel corporativo en el grupo 3 (técnicos y profesionales de nivel medio). Consecuentemente, y al no poder realizar una distinción apropiada entre ambos grupos de vendedores con la información de Avisos de Entradas y Salidas, se asumió que la mayor parte de vendedores se ubican en actividades de distribución final, es decir en el grupo 5. De forma similar, los supervisores de ventas fueron clasificados en el grupo 5 bajo el supuesto de que los supervisores de ventas se encuentran mayoritariamente en almacenes en lugar de actividades gerenciales. En total, de los 121.740 trabajadores que en diciembre de 2020 tuvieron algún cargo relacionado a vendedores, el 81% fue clasificado al grupo 5. Finalmente, los asesores de postventas fueron clasificados en el grupo ocupacional 3 (técnicos y profesionales de nivel

---

<sup>21</sup> Para esta actividad se utilizó un algoritmo basado en vecinos más cercanos adaptado a la codificación de ocupaciones. Para más información véase Gweon et al. (2017) ([link](#)).



medio), representando solamente el 1% del grupo de trabajadores analizado. El resto de los trabajadores fueron clasificados en el grupo 1 (directores y gerentes) y grupo 4 (trabajadores administrativos), ya que presentaron descripciones de cargo relacionados a estos grupos.

#### 4.2.1.2. Creación del cargo en registros sin información disponible

##### Motivación

La Figura 3 ilustra que, si bien las bases de Afiliados al IESS se disponen desde el año 2006, en todos los meses existen personas para las cuales se pudo conocer su cargo (cuadros azules) y un subconjunto para las cuales no puede conocerse (cuadros blancos).

Figura 3. Periodos con información disponible de cargo para el sector privado del IESS

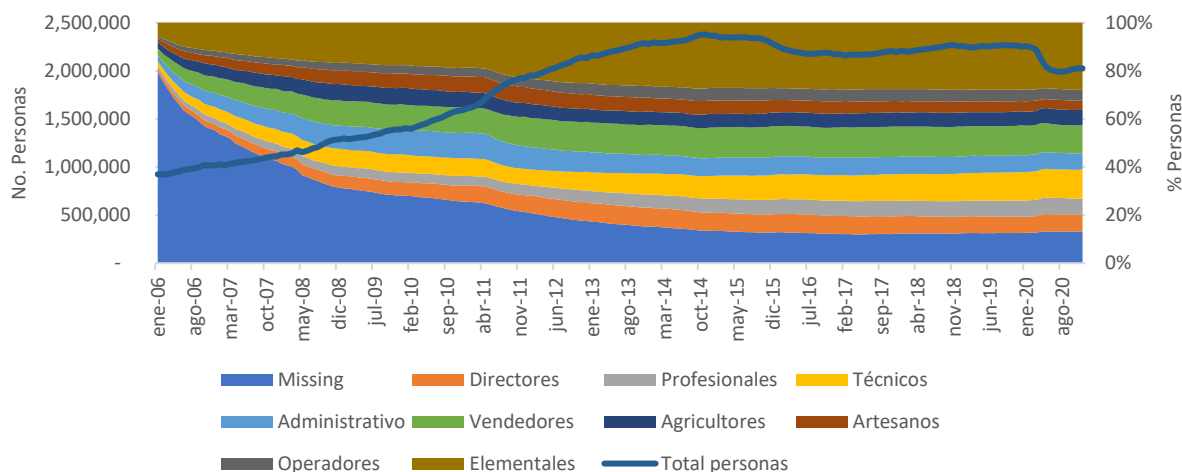
Mes/Año	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
1															
2															
3															
4															
5															
6															
7															
8															
9															
10															
11															
12															

*Fuente:* Elaborado por los autores.

*Nota:* La información disponible se presenta en cuadros azules. La figura ilustra que, si bien se cuenta con datos de Afiliados al IESS para todos los periodos, siempre existieron registros sin información de cargo, aunque no necesariamente en el mismo porcentaje.

La dinámica del proceso de construcción del cargo ocupacional de los trabajadores del sector privado afiliados al IESS hace que la cobertura de la variable aumente a medida que se registran nuevas relaciones laborales en la base de Avisos de Entradas y Salidas. Sin embargo, a pesar de este aumento gradual de la cobertura, el porcentaje de datos perdidos ha permanecido estable en un nivel de alrededor del 12% desde el 2015 (Figura 4).

Figura 4. Distribución ocupacional con información disponible del cargo. Sector privado IESS.



Fuente: Información de registros administrativos del LDLE-INEC. Elaborado por los autores.

### Prueba de algoritmos de clasificación

Para completar la información del cargo en todas las bases desde 2006 a 2020 se emplearon modelos de clasificación basados en árboles, haciendo uso de los cortes de diciembre de cada año para entrenar los modelos con los registros con información disponible y posteriormente prediciendo el cargo de los registros sin información tanto de diciembre como del resto de meses del mismo año. Las variables usadas para predecir el cargo se encuentran en la Tabla 1.

Tabla 1. Variables incluidas en los modelos de clasificación del sector privado del IESS

No.	Variable	Observaciones
1	Sueldo	Variable del IESS
2	Edad	Variable del IESS-Registro Civil
3	Género	Variable del IESS-Registro Civil
4	Estado Civil	Variable del Registro Civil
5	Nivel de instrucción	Variable de SENESCYT
6	Tipo de empleador	Variable del IESS
7	Provincia*	Variable del IESS
8	Sector económico (CIU)	Variable del DIEE y SRI
9	Estado de la empresa	Variable del SRI
10	Tipo de empresa	Variable del SRI
11	Tamaño de la empresa	Variable del IESS. Tamaño según el empleo.
12	Población provincial/cantonal**	Proyecciones poblacionales del INEC

Fuente: Elaborado por los autores.

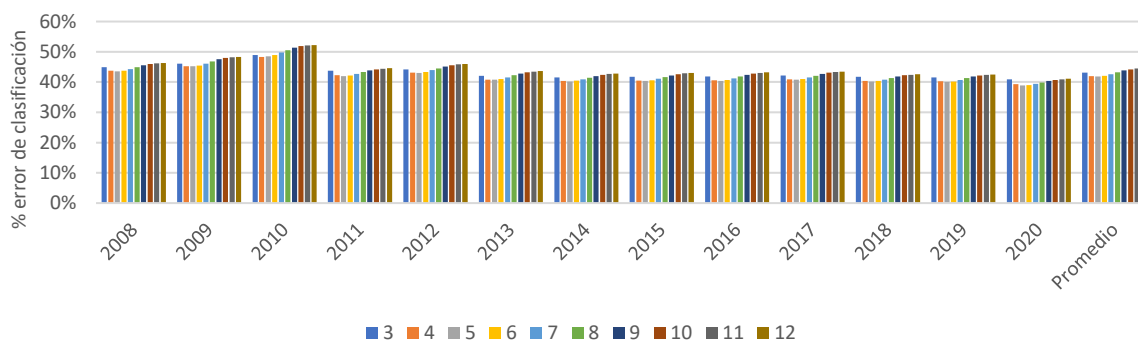
A partir del conjunto de variables de la Tabla 1 se estimaron varios modelos de clasificación que se diferencian por el número de  $k$  variables que se eligen aleatoriamente para determinar los criterios de clasificación de los modelos de árboles. Una vez que se estimaron los  $k$  modelos se logró determinar que el modelo con  $k = 5$  tiene el menor nivel de error, con un promedio de

41,8% casos clasificados erróneamente entre 2008 y 2020 (Figura 5). Es importante tomar en cuenta que desde el año 2014 el porcentaje de casos perdidos es de menos del 15%, por lo que a partir de este periodo los casos erróneamente clasificados corresponderían al 6,3% del total de trabajadores del sector privado.

Posterior a esta verificación se comparó la distribución del empleo por grupo ocupacional predicho con los  $k$  modelos de clasificación, frente a la distribución ocupacional observada. Tanto en términos de errores relativos (

Figura 6) como absolutos (Figura 7), el modelo con  $k = 12$  presenta las menores diferencias una vez que se agrega el empleo por grupos ocupacionales. Considerando este resultado y el hecho de que el porcentaje de errores de clasificación a nivel de microdato asciende a 44,6%, es decir 2,7 puntos porcentuales frente al modelo con  $k = 5$ , se optó por trabajar con el modelo que considera a las  $k = 12$  variables.

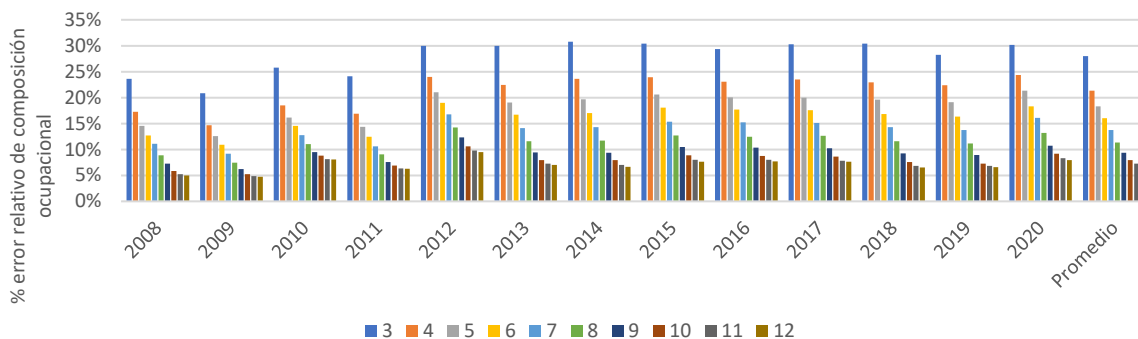
Figura 5. Errores de clasificación individual de las predicciones para valores de  $k$ . Sector privado IESS.



**Fuente:** Información de registros administrativos del LDLE-INEC. Elaborado por los autores.

**Nota:** Se eligió aleatoriamente el 75% de la muestra para entrenamiento y el 25% para prueba. El parámetro  $k$  define la cantidad de variables que el modelo elige aleatoriamente para estimar el modelo de entrenamiento. No se presentan los resultados para los valores de  $k=1$  y  $k=2$  por los altos niveles de error que poseen.

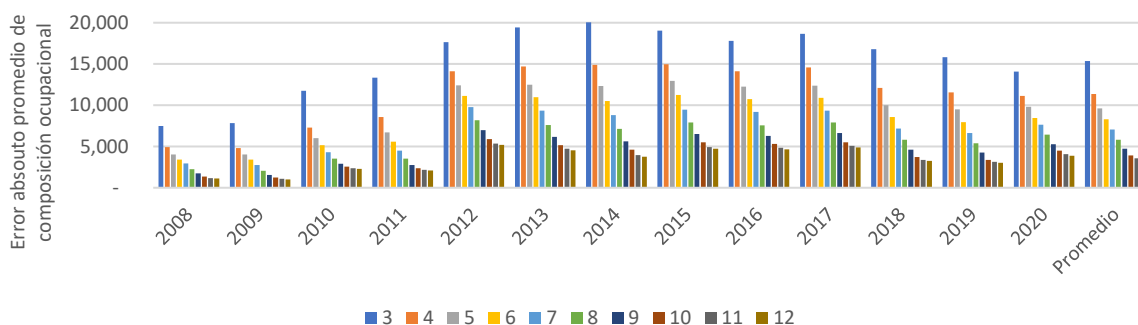
**Figura 6. Error relativo promedio de la composición ocupacional de las predicciones para valores de k. Sector privado IESS.**



**Fuente:** Información de registros administrativos del LDLE-INEC. Elaborado por los autores.

**Nota:** Se eligió aleatoriamente el 75% de la muestra para entrenamiento y el 25% para prueba. El parámetro *k* define la cantidad de variables que el modelo elige aleatoriamente para estimar el modelo de entrenamiento. No se presentan los resultados para los valores de *k*=1 y *k*=2 por los altos niveles de error que poseen.

**Figura 7. Error absoluto promedio de la composición ocupacional de las predicciones para valores de k. Sector privado IESS.**



**Fuente:** Información de registros administrativos del LDLE-INEC. Elaborado por los autores.

**Nota:** Se eligió aleatoriamente el 75% de la muestra para entrenamiento y el 25% para prueba. El parámetro *k* define la cantidad de variables que el modelo elige aleatoriamente para estimar el modelo de entrenamiento. No se presentan los resultados para los valores de *k*=1 y *k*=2 por los altos niveles de error que poseen.

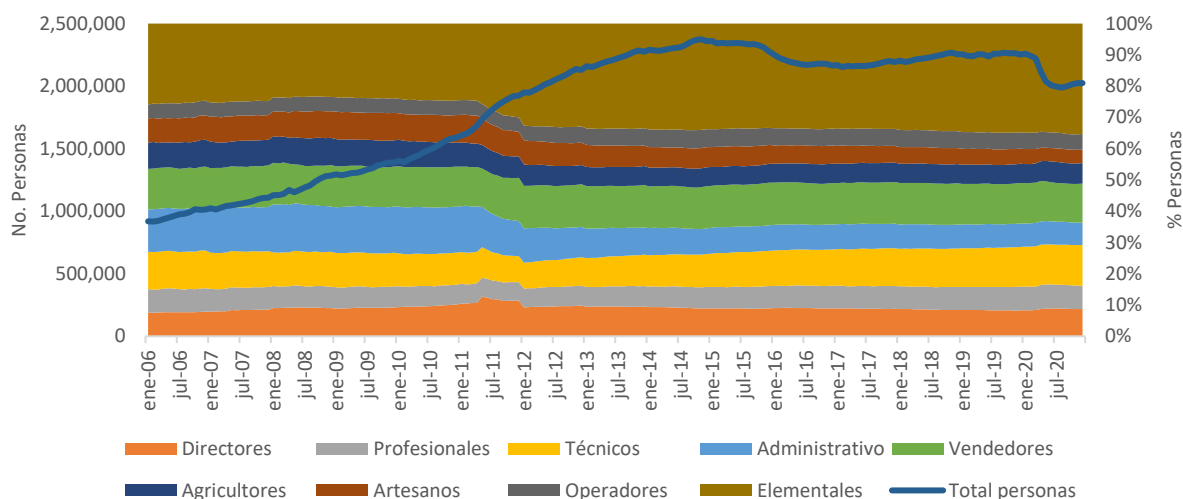
#### 4.2.1.3. Composición ocupacional con registros completos

Una vez identificado el modelo que genera los mejores resultados en términos de precisión de la distribución ocupacional fue posible completar el cargo en aquellos registros que previamente no contaban con esta información<sup>22</sup>. La Figura 8 muestra justamente el resultado de todo el proceso,

<sup>22</sup> Para la construcción de la variable del cargo ocupacional en las bases históricas se emplearon herramientas de paralelización, lo cual permitió reducir el tiempo de procesamiento de alrededor de 244 a 62 minutos por cada año.

pudiéndose notar que los registros con información faltante del cargo prácticamente desaparecen en todo el periodo comprendido entre 2006 y 2020.

**Figura 8. Distribución ocupacional con información completa del cargo. Sector privado IESS.**



*Fuente: Información de registros administrativos del LDLE-INEC. Elaborado por los autores.*

## 4.2.2. Sector público: GADs y EPs

### 4.2.2.1. Creación del cargo en registros con información disponible

#### *Archivos de transparencia*

Si bien los trabajadores de los GADs y EPs cotizan a la seguridad social y por ende se encuentran registrados en la base de Afiliados al IESS, la información del cargo ocupacional no estaba disponible en ninguno de los registros administrativos del LDLE, lo cual llevó a emplear una estrategia diferente a las realizadas para el sector privado y la administración central del sector público.

A diferencia de las entidades públicas de la administración central, las escalas salariales de los GADs y EPs no están sujetas a un estándar único ya que cuentan con autonomía para fijarlas. Esto imposibilita el uso de la información de los servidores públicos de la administración central del distributivo del Ministerio de Finanzas para realizar alguna clasificación en base a la variable construida del cargo (proceso descrito en la sección 4.2.3.1) y otras adicionales. Por esta razón se trabajó en una estrategia que permitió completar la información del cargo ocupacional de los trabajadores de los GADs y EPs<sup>23</sup> considerando la heterogeneidad salarial entre entidades.

<sup>23</sup> En el conjunto de EPs se encuentran las siguientes entidades: CNT, CNEL, CELEC, IESS, BANECUADOR, PETROAMAZONAS, PETROECUADOR y Banco Central del Ecuador. Es importante tomar en cuenta que varias de estas entidades no son en estricto rigor una empresa pública, pero al no disponer de información del cargo en los registros del LDLE se priorizaron a las entidades que concentran la mayor cantidad de empleo. Entre las ocho entidades acumularon un total de 39.932 empleos en diciembre de 2020, lo cual en comparación a los empleos que generan los



En el registro de afiliados al IESS se puede identificar en qué GAD o EP trabaja una persona y el sueldo que posee. Por otra parte, todas las entidades públicas, incluyendo los GADs y EPs, deben publicar en sus páginas web la nómina de sus trabajadores con el cargo que ocupan y su respectiva remuneración<sup>24</sup>. Con ello fue posible realizar una equivalencia del cargo ocupacional que un trabajador ocupa con el sueldo que percibe en cada GAD o EP<sup>25</sup>, para luego identificar el cargo de cada trabajador registrado en el IESS según el rango salarial en el que se encuentre<sup>26</sup>. La Tabla 2 resume los grupos ocupacionales que se construyeron en cada GAD y EP a través de información de rangos salariales y su correspondencia con la CIUO.

Tabla 2. Grupos de ocupación para GADs y EPs

No.	Descripción	Correspondencia CIUO
1	Personas con nivel jerárquico superior.	Grupo 1: Directores y gerentes
2	Profesionales universitarios o especializados (abogados, ingenieros, médicos, entre otros).	Grupo 2: Profesionales científicos e intelectuales
3	Personal de nivel técnico.	Grupo 3: Técnicos y profesionales de nivel medio
4	Personal de apoyo de oficina y administrativo.	Grupo 4: Personal de apoyo administrativo
5	Personal de servicios.	Grupo 9: Ocupaciones elementales

*Fuente: Elaborado por los autores.*

La recuperación de la información se realizó solamente para los GADs municipales y provinciales<sup>27</sup>. En el primer caso se pudo recuperar la información para 191 de los 218 GADs (88%), mientras que en el segundo se pudo obtener información para los 24 GADs (100%) provinciales existentes.

#### 4.2.2.2. Creación del cargo en registros sin información disponible

##### *Motivación*

El proceso de construcción de la variable del cargo ocupacional para los GADs y EPs requirió un trabajo intensivo de búsqueda manual y sistematización de información. El esfuerzo realizado para

---

242 GADs municipales y provinciales representa el 37%; es decir el 37% de los empleos que generan los GADs municipales y provinciales.

<sup>24</sup> El Art. 7 de la Ley Orgánica de Transparencia y Acceso a la Información Pública (LOTAIP) dispone que todas las instituciones del Estado que conforman el sector público difundirán a través de un portal de información o página web, entre otra información, el distributivo de personal y la remuneración mensual por puesto ([link](#)).

<sup>25</sup> Este proceso se realizó solamente sobre los meses de diciembre entre 2014 y 2019 por el alto componente manual que conlleva. El ejercicio consistió en identificar en cada GAD o EP conjuntos de cargos que pueden clasificarse en los distintos grupos de la CIUO y determinar el rango salarial que perciben.

<sup>26</sup> Esta estrategia no está exenta de limitaciones ya que dentro de un mismo GAD dos personas con igual salario pueden ocupar cargos diferentes. En estos casos, para armar las correspondencias se consideraron los cargos ocupacionales que se asocian a un salario con más frecuencia.

<sup>27</sup> Los GADs parroquiales fueron descartados de este proceso debido a su baja participación sobre el empleo según el registro de Afiliados al IESS (8% del empleo de todos los GADs incluyendo provinciales y municipales en diciembre de 2020) y la falta de información de transparencia publicada en los portales web (el 43% de los GADs parroquiales tuvieron información publicada de transparencia).

construir las bases de correspondencia de rango salarial y grupo ocupacional por cada GAD y EP permitió completar la información del cargo para los meses de diciembre entre 2014 y 2019 (Figura 9).

Figura 9. Periodos con información disponible de cargo para GADs y EPs

Mes/Año	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
1													
2													
3													
4													
5													
6													
7													
8													
9													
10													
11													
12													

Fuente: Elaborado por los autores.

Nota: La información disponible se presenta en cuadros azules.

Por la intensidad de trabajo manual que requiere este proceso, no es sostenible ni recomendable repetirlo en cada mes. Sin embargo, contar con información del cargo para todos los meses desde el año 2008 y en los periodos futuros es de suma utilidad, y por tal razón se trabajó en una estrategia de recuperación del cargo ocupacional para estos periodos.

### Prueba de algoritmos de clasificación

#### 1) Construcción del cargo ocupacional en los meses de enero a noviembre entre 2014 y 2019.

Una vez que se obtuvo la variable del cargo de ocupación para todos los trabajadores de los GADs y EPs con información disponible de los meses de diciembre de 2014 a 2019, se procedió a predecir el cargo de los trabajadores para el resto de los meses mediante algoritmos de clasificación basados en árboles. En total se incluyeron nueve variables al modelo de clasificación de GADs municipales y provinciales y EPs (Tabla 3).

Tabla 3. Variables incluidas en los modelos de clasificación de GADs y EPs

No.	Variable	Observaciones
1	Sueldo	Variable del IESS
2	Edad	Variable del IESS-Registro Civil
3	Género	Variable del IESS-Registro Civil
4	Estado civil	Variable del Registro Civil
5	Nivel de instrucción	Variable de SENESCYT
6	Relación de trabajo	Variable del IESS
7	Provincia	Variable del IESS
8	Tamaño del GAD	Variable del IESS. Tamaño según el empleo.
9	Población provincial/cantonal*	Proyecciones poblacionales del INEC

*Fuente:* Elaborado por los autores.

*Nota:* (\*) La población se tomó dependiendo del nivel geográfico sobre el cual se aplicó el algoritmo.

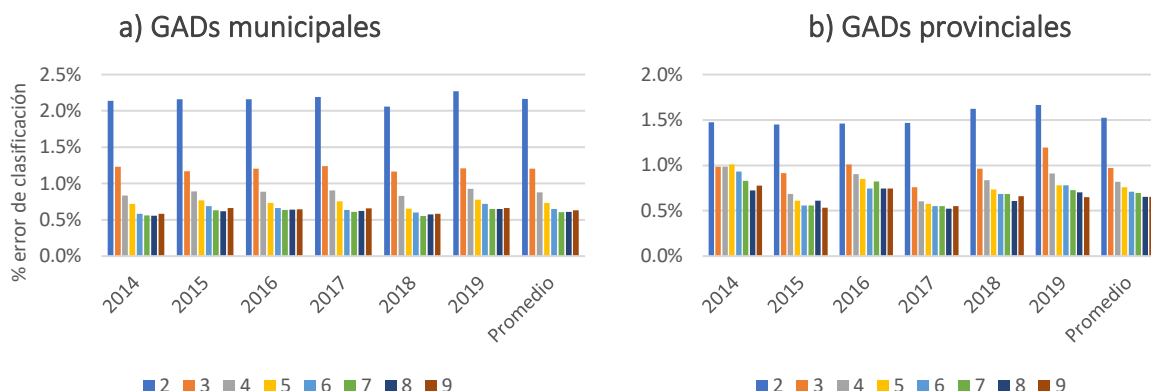
Para elegir el modelo de clasificación del cargo ocupacional se evaluó la precisión de la clasificación con varios modelos que se diferencian por el número de  $k$  variables que se eligen aleatoriamente para determinar los criterios de clasificación. Dado que no todas las variables usadas en los GADs municipales y provinciales son las mismas, se realizaron ejercicios independientes. Adicionalmente, es importante recordar que para este subconjunto de trabajadores, el cargo fue construido en base a los rangos salariales que perciben personas con cargos similares dentro de la misma entidad, y que por lo tanto la precisión de los modelos es alta al considerar al salario como una de las variables predictoras del cargo.

En el caso de los GADs municipales, el error de clasificación individual cae a medida que se considera un valor de  $k$  mayor con un valor promedio mínimo de 0,6% para  $k = 7$  entre 2014 y 2019, aunque con niveles similares desde  $k = 5$ . Para los GADs provinciales el error de clasificación se estabiliza desde  $k = 4$  y se alcanza un mínimo para  $k = 8$  con un error promedio de 0,7% (Figura 10).

Cuando se evalúa la precisión de la distribución ocupacional en términos relativos (Figura 11), no se logra identificar un patrón claro entre los valores de  $k$  y el nivel de precisión en ninguno de los periodos. Sin embargo, si se toman los errores relativos promedio para todos los años entre 2014 y 2019, se observa una forma de U invertida a partir de  $k = 3$ . Al analizar los errores absolutos promedio de las diferencias en las composiciones ocupacionales (Figura 12), estos decrecen en el caso de los GADs municipales cuando se toma un valor más grande de  $k$ . Con los GADs provinciales no se logra identificar un patrón claro a lo largo de los años, aunque en promedio de 2014 a 2019 se observa una forma de U invertida entre  $k = 3$  y  $k = 9$ .

Tomando en consideración los errores de predicción de los modelos con distintos valores de  $k$  a nivel de microdato y distribución ocupacional, se optó por trabajar con un valor de  $k = 7$  para GADs municipales y  $k = 8$  para GADs provinciales, parámetros que dan como resultado el error de predicción mínimo a nivel de microdato. Si bien estos parámetros no dan los menores niveles de error cuando se consideran las diferencias relativas y absolutas en las distribuciones ocupacionales, las diferencias no son significativas en comparación a los valores mínimos.

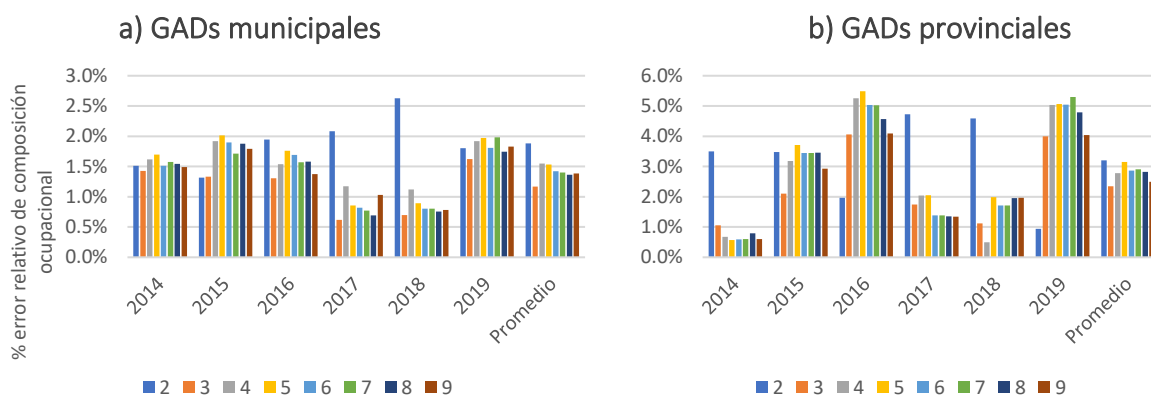
Figura 10. Errores de clasificación individual de las predicciones para valores de k. GADs.



**Fuente:** Información de registros administrativos del LDLE-INEC. Elaborado por los autores.

**Nota:** Se eligió aleatoriamente el 75% de la muestra para entrenamiento y el 25% para prueba. El parámetro  $k$  define la cantidad de variables que el modelo elige aleatoriamente para estimar el modelo de entrenamiento. No se presentan los resultados para los valores de  $k=1$  por los altos niveles de error que poseen.

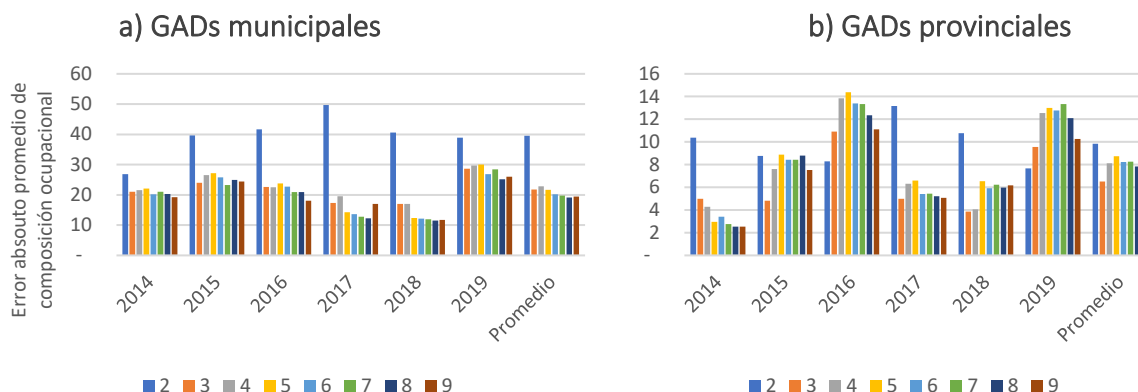
Figura 11. Error relativo promedio de la composición ocupacional de las predicciones para valores de k. GADs



**Fuente:** Información de registros administrativos del LDLE-INEC. Elaborado por los autores.

**Nota:** Se eligió aleatoriamente el 75% de la muestra para entrenamiento y el 25% para prueba. El parámetro  $k$  define la cantidad de variables que el modelo elige aleatoriamente para estimar el modelo de entrenamiento. No se presentan los resultados para los valores de  $k=1$  por los altos niveles de error que poseen.

Figura 12. Error absoluto promedio de la composición ocupacional de las predicciones para valores de  $k$ . GADs.



*Fuente:* Información de registros administrativos del LDLE-INEC. Elaborado por los autores.

*Nota:* Se eligió aleatoriamente el 75% de la muestra para entrenamiento y el 25% para prueba. El parámetro  $k$  define la cantidad de variables que el modelo elige aleatoriamente para estimar el modelo de entrenamiento. No se presentan los resultados para los valores de  $k=1$  por los altos niveles de error que poseen.

## 2) Construcción del cargo ocupacional entre 2008 y 2013, y el año 2020.

La información correspondiente a los meses de diciembre de 2014 y 2019 fue empleada para realizar el entrenamiento de los modelos, los cuales posteriormente fueron aplicados al resto de los meses del mismo año de referencia del modelo. Para los años en los que no se cuenta con información para diciembre se tomó como base de entrenamiento la del periodo más reciente. Es decir, para los años antes de 2014 se tomó como referencia la información de diciembre de 2014, mientras que para 2020 se empleó la base de diciembre de 2019 para este fin<sup>28</sup>.

## 3) Construcción del cargo ocupacional en los registros faltantes de diciembre entre 2014 y 2019.

En el proceso de búsqueda de información de los archivos de transparencia de los portales web no se logró encontrar información para 27 GADs municipales. Por tal razón, se aplicó el mismo algoritmo del paso 1 a los registros con información faltante de cargo de los meses de diciembre entre 2014 y 2019<sup>29</sup>. Este proceso se hizo al final para evitar tomar a registros con ocupación predicha como parte de las muestras de entrenamiento para el resto de los periodos.

<sup>28</sup> Para los años sin información de diciembre (2008 a 2013, y 2020), el sueldo fue normalizado al mes de diciembre del año más cercano para considerar cambios salariales promedio a nivel de GAD. Por ejemplo, para el año 2013, el salario de los trabajadores de los GADs fue transformado a un salario que tendrían en 2014 considerando el cambio del promedio salarial general.

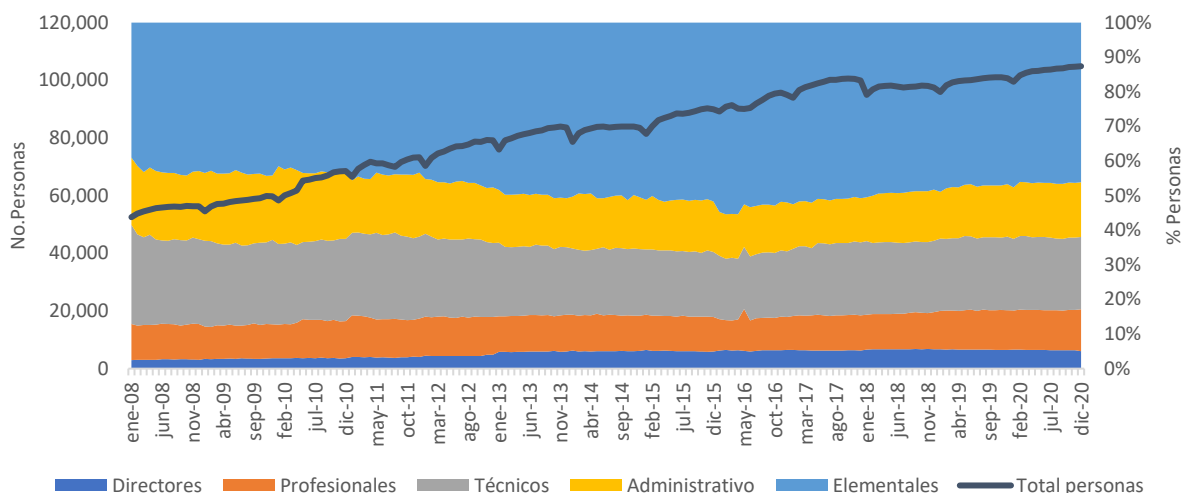
salarial a un salario normalizado a través del promedio salarial de cada GAD fue normalizado con el promedio correspondiente al 2014

<sup>29</sup> Dado que la pérdida de información en este caso ocurre a nivel de GAD, se probó la precisión del modelo simulando pérdidas de información de GADs en su totalidad. Como resultado de este ejercicio, se logró determinar los errores de clasificación son mínimos con  $k = 2$ . Por consiguiente, la información del cargo de los registros faltantes de los meses de diciembre entre 2014 y 2019 se recuperó a través de este modelo.

### 4.2.2.3. Composición ocupacional con registros completos

Como resultado de todo este proceso se logró obtener la evolución de la composición ocupacional entre 2008 y 2020 (Figura 13)<sup>30</sup>. Si bien no se observa una composición ocupacional fija a lo largo del tiempo, los patrones evolucionan de forma estable.

Figura 13. Distribución ocupacional con información completa del cargo. GADs municipales y provinciales.



Fuente: Información de registros administrativos del LDLE-INEC. Elaborado por los autores.

## 4.2.3. Sector público: Ministerio de Finanzas

### 4.2.3.1. Creación del cargo en registros con información disponible

#### Base de distributivo MEF

La información del cargo ocupacional de los trabajadores del sector público se obtuvo de la base del Distributivo de Personal del MEF, la cual posee el listado de las personas que se encuentran trabajando en alguna entidad pública de la administración central al cierre de cada mes<sup>31</sup>. Previo a la construcción de la variable del cargo ocupacional se realizó una delimitación del universo de trabajo<sup>32</sup> de modo que sea compatible con el universo de servidores públicos registrados en la base de Afiliados al IESS, y se priorizó un conjunto de variables útiles para el análisis, entre las cuales se destacan principalmente las que contienen información referente al cargo de los

<sup>30</sup> Para la construcción de la variable del cargo ocupacional en las bases históricas se emplearon herramientas de paralelización, lo cual permitió reducir el tiempo de procesamiento de alrededor de 25 a 8 minutos por cada año.

<sup>31</sup> El Distributivo de Personal del MEF excluye a los trabajadores de las entidades públicas con autonomía y Gobiernos Autónomos Descentralizados (GADs).

<sup>32</sup> Dentro del Distributivo de Personal están los servidores de la Policía Nacional y Fuerzas Armadas. Estos grupos fueron descartados ya que no forman parte del sistema de afiliación del IESS, sino de ISSPOL e ISSFA, respectivamente.

trabajadores<sup>33</sup>. Sin embargo, al igual que en el caso de los trabajadores del sector privado, la información del cargo no se encontraba codificada y por lo tanto se creó un diccionario de correspondencia de las descripciones del cargo disponibles y los grupos ocupacionales de la CIUO<sup>34</sup>.

El proceso de construcción de la variable de ocupación en el universo de servidores públicos del Distributivo del MEF incluyó un componente de automatización, con la finalidad de que esta variable pueda obtenerse de manera eficiente en nuevas entregas de información y evitar en el mayor grado posible un proceso manual de limpieza de la información. En resumen, el proceso que se siguió se puede resumir en tres pasos:

- 1. Elaboración del diccionario de correspondencia:** Con el corte de la base del Distributivo de personal de diciembre de 2018 se creó un diccionario de correspondencia entre las descripciones del cargo y la CIUO. Este proceso permitió contar con una variable de cargo ocupacional completa para todos los trabajadores de diciembre de 2018.
- 2. Recuperación del cargo a partir del uso directo del diccionario:** Con el diccionario de correspondencia de diciembre de 2018 se procedió a recuperar el cargo en los registros de otros meses en los cuales se encontraba una coincidencia exacta de la descripción del cargo ocupacional.
- 3. Uso de modelos de clasificación para los cortes disponibles:** Para recuperar el cargo de los registros que no pudieron clasificarse debido a que su descripción no coincidió exactamente con el diccionario, se emplearon modelos de clasificación de lenguaje natural basados en medidas de similitud<sup>35</sup>.
- 4. Completar el diccionario con los registros que no pudieron ser clasificados:** A pesar de aplicar modelos de clasificación, existen descripciones que aparecen en otros cortes de información (y otras que pueden aparecer en bases futuras) a los cuales no se puede asignar un cargo porque no existen descripciones semejantes en el diccionario que permitan hacerlo. Por tal razón, al final se completa manualmente las descripciones que no pudieron ser clasificadas, a fin de tener una cobertura completa de la variable<sup>36</sup>.

---

<sup>33</sup> El proceso de limpieza se aplicó para los meses con información disponible desde el año 2008. Dado que en este periodo se cuenta con dos estructuras para la base del Distributivo del MEF (eSIPREN entre 2013 y 2018, y SPRYN desde 2014 en adelante), se trabajó en una identificación previa de criterios consistentes en el tiempo desde un punto de vista conceptual, para luego determinar las variables que en ambas estructuras permitieron aplicarlos.

<sup>34</sup> El diccionario de correspondencia se creó a través de la asignación de las descripciones del cargo a los grupos ocupacionales de la CIUO que conceptualmente son más acordes. Para este efecto se utilizó el manual de la CIUO a fin de contar con los criterios adecuados para efectuar este ejercicio.

<sup>35</sup> Se usó el paquete “occupationCoding” de R. Para más información véase Schierholz y Schonlau (2020) ([link](#)).

<sup>36</sup> En la programación realizada se exporta un archivo con el listado de las descripciones de cargo que no pudieron ser clasificadas, con el fin de facilitar al usuario la exploración de estos casos.



### 4.2.3.2. Creación del cargo en registros sin información disponible

#### Motivación

El proceso de recuperación del cargo para las bases del Distributivo del MEF solo puede aplicarse en los cortes para los cuales se tiene información. En la Figura 14 se observa que la información del MEF empieza a estar disponible en el LDLE de manera mensual a partir de julio de 2018 cuando inició el convenio de transferencia de información con el INEC; mientras que, para los años anteriores, el INEC recibió solamente los cortes pertenecientes a los meses de diciembre.

Figura 14. Periodos con bases disponibles del Distributivo de Personal del MEF

Mes/Año	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
1													
2													
3													
4													
5													
6													
7													
8													
9													
10													
11													
12													

Fuente: Elaborado por los autores.

Nota: La información disponible se presenta en cuadros azules.

Adicional a esta falta de información, fue necesario considerar un aspecto relacionado al sector público en general. En primer lugar, es importante partir del hecho de que en la base del Distributivo del MEF todos los trabajadores cuentan con un cargo asignado luego de aplicar el algoritmo de la sección 4.2.3.1. Similarmente, el proceso que se siguió para obtener el cargo de los trabajadores de los GADs municipales y provinciales, y EPs permite contar con una variable de cargo para todos los trabajadores de estas entidades públicas. Sin embargo, dentro del sector público registrado en la base de Afiliados al IESS aún quedan trabajadores sin información del cargo pertenecientes a entidades como GADs parroquiales, empresas públicas del Estado, empresas públicas de los GADs, y otras entidades con autonomía.

Con estos antecedentes, es claro que para el sector público la falta de información no solamente ocurre en meses específicos, sino que también hay un grupo de servidores públicos para los cuales no se tiene información del cargo ocupacional en todo el periodo analizado. Por lo tanto, la estrategia de recuperación del cargo fue diseñada para solventar la falta de información de todos estos casos y al igual que en los procesos anteriores se utilizaron modelos basados en árboles verificando previamente su nivel de precisión.

### Prueba de algoritmos de clasificación

La estrategia de recuperación del cargo fue más sencilla que la realizada para los GADs y EPs. Con el universo de trabajadores que poseen información del cargo para los meses de diciembre entre 2008 y 2020, tanto de la administración central (Distributivo de Personal del MEF) como de los GADs y EPs, se entrenaron modelos basados en árboles y los parámetros fueron aplicados a los registros del resto de los meses del mismo año. El proceso de predicción del cargo solamente se realizó en las observaciones que no tenían previamente un cargo asignado. Las variables usadas para predecir el cargo se encuentran detalladas en la Tabla 4.

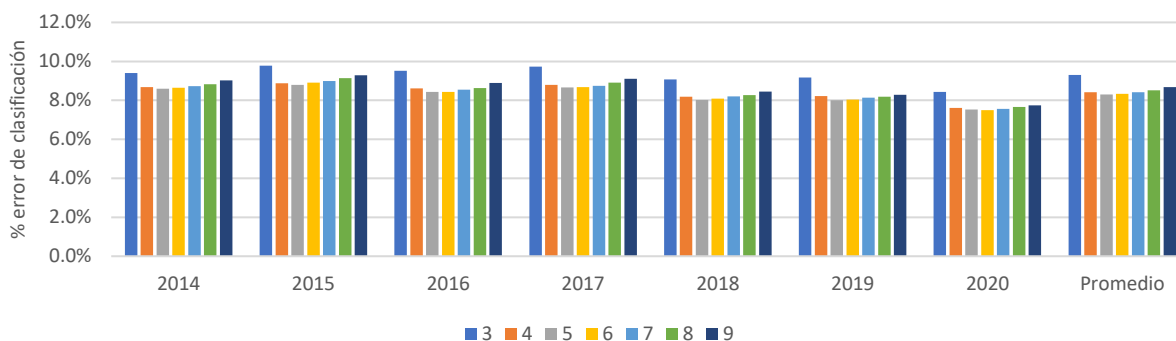
**Tabla 4. Variables incluidas en los modelos de clasificación para el sector público**

No.	Variable	Observaciones
1	Sueldo	Variable del IESS
2	Edad	Variable del IESS-Registro Civil
3	Género	Variable del IESS-Registro Civil
4	Estado civil	Variable del Registro Civil
5	Nivel de instrucción	Variable de SENESCYT
6	Relación de trabajo	Variable del IESS
7	Provincia	Variable del IESS
8	Tamaño del GAD	Variable del IESS. Tamaño según el empleo de los GADs.
9	Población cantonal	Proyecciones poblacionales del INEC

*Fuente: Elaborado por los autores.*

Los errores de clasificación que se obtienen al estimar modelos con distintos valores para  $k$  son estables a partir de  $k = 4$  con un promedio de 8,4% entre 2014 y 2020. Si bien el error mínimo ocurre con  $k = 5$ , la diferencia es de apenas 0,2 puntos porcentuales con respecto a  $k = 9$  (Figura 15). Al contrario, al considerar las diferencias relativas (Figura 16) y absolutas (Figura 17) de las distribuciones ocupacionales agregadas se puede notar que los niveles de error caen cuando el valor de  $k$  aumenta, alcanzando un mínimo en ambos casos cuando  $k = 9$ . Como consecuencia de esto, el modelo que se eligió para estimar el cargo de los registros no observados consideró un valor de  $k = 9$ , ya que por un lado existe un mejor desempeño en comparación a otros valores en términos agregados, y por otro lado se tiene una pérdida de precisión a nivel de microdato marginal.

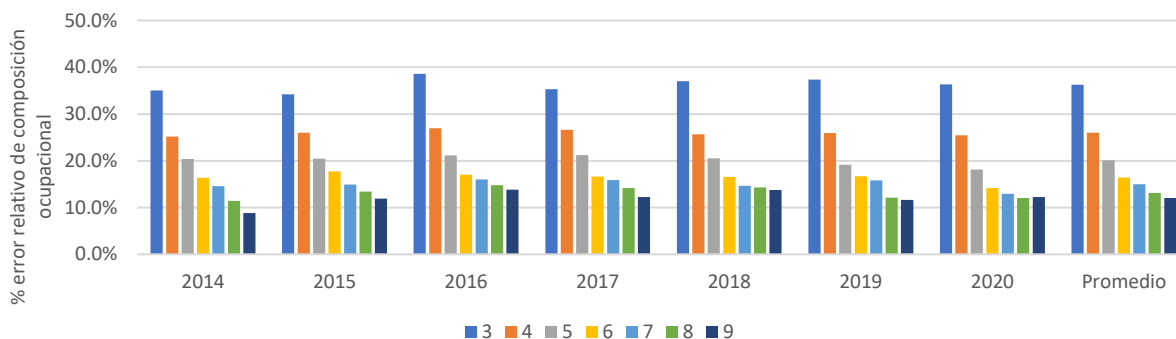
Figura 15. Error de clasificación individual de las predicciones para valores de k. Sector Público.



**Fuente:** Información de registros administrativos del LDLE-INEC. Elaborado por los autores.

**Nota:** Se eligió aleatoriamente el 75% de la muestra para entrenamiento y el 25% para prueba. El parámetro  $k$  define la cantidad de variables que el modelo elige aleatoriamente para estimar el modelo de entrenamiento. No se presentan los resultados para los valores de  $k=1$  y  $k=2$  por los altos niveles de error que poseen.

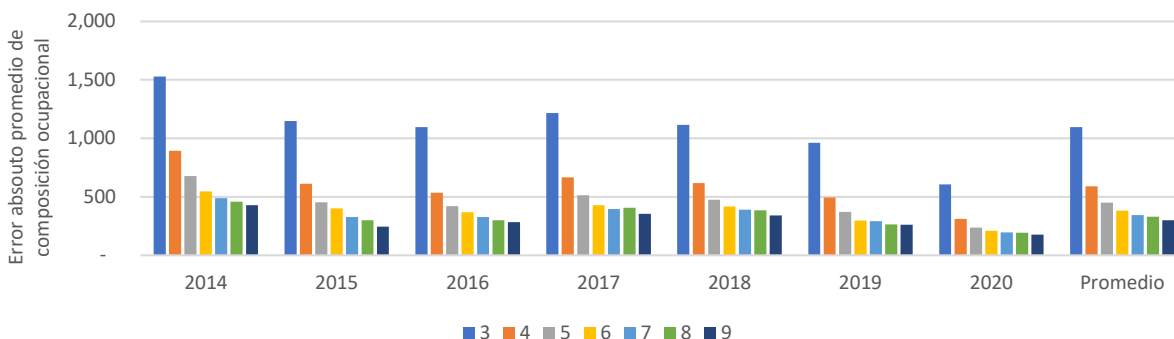
Figura 16. Error relativo promedio de la composición ocupacional de las predicciones para valores de k. Sector público.



**Fuente:** Información de registros administrativos del LDLE-INEC. Elaborado por los autores.

**Nota:** Se eligió aleatoriamente el 75% de la muestra para entrenamiento y el 25% para prueba. El parámetro  $k$  define la cantidad de variables que el modelo elige aleatoriamente para estimar el modelo de entrenamiento. No se presentan los resultados para los valores de  $k=1$  y  $k=2$  por los altos niveles de error que poseen.

Figura 17. Error absoluto promedio de la composición ocupacional de las predicciones para valores de k. Sector público.



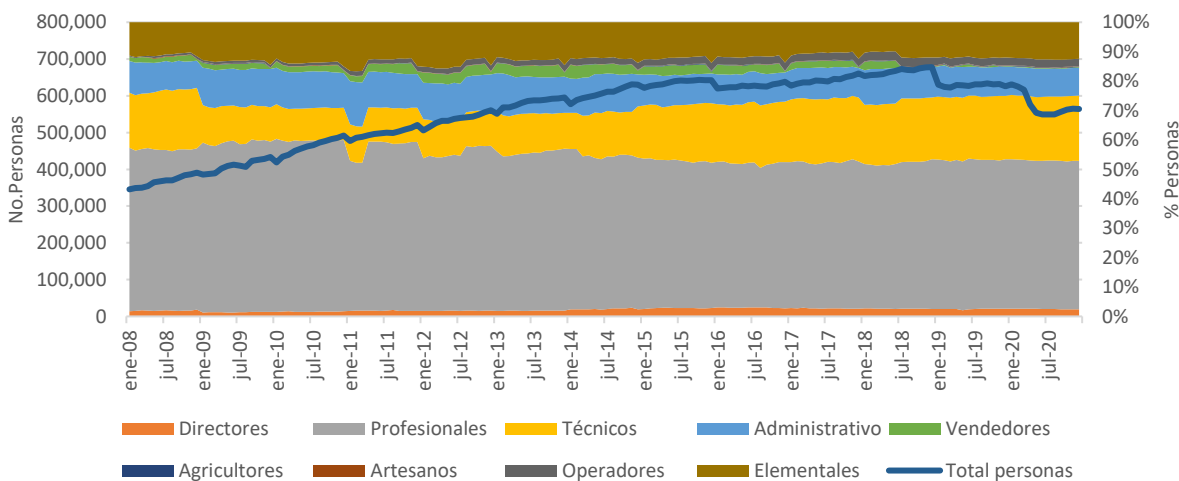
**Fuente:** Información de registros administrativos del LDLE-INEC. Elaborado por los autores.

**Nota:** Se eligió aleatoriamente el 75% de la muestra para entrenamiento y el 25% para prueba. El parámetro k define la cantidad de variables que el modelo elige aleatoriamente para estimar el modelo de entrenamiento. No se presentan los resultados para los valores de k=1 y k=2 por los altos niveles de error que poseen.

#### 4.2.3.3. Composición ocupacional con registros completos

Una vez identificado el modelo con el mejor desempeño en términos de precisión, se completó el cargo para todos los registros faltantes entre 2008 y 2020<sup>37</sup>. El resultado de este ejercicio se muestra en la Figura 18.

Figura 18. Distribución ocupacional con información completa del cargo. Sector público.



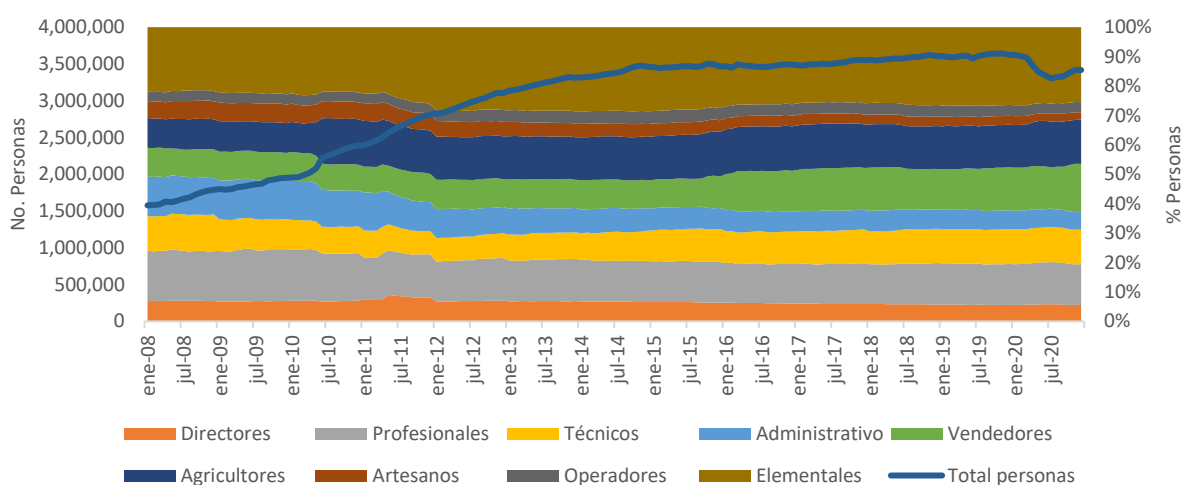
**Fuente:** Información de registros administrativos del LDLE-INEC. Elaborado por los autores.

<sup>37</sup> Para la construcción de la variable del cargo ocupacional en las bases históricas se emplearon herramientas de paralelización, lo cual permitió reducir el tiempo de procesamiento de alrededor de 14 a 6 minutos por cada año.

#### 4.2.4. Evolución de la composición ocupacional con información completa

Finalmente, la variable de cargo de ocupación generada en los tres subuniversos descritos en la anterior sección se integró a las bases de afiliados al IESS en todos los meses desde enero de 2008 a diciembre de 2020. La Figura 19 muestra la evolución de la composición ocupacional para todo el universo de trabajadores afiliados a la seguridad social. Cabe notar que la variable del cargo es una variable adicional a las ya disponibles, lo cual posteriormente abre la posibilidad de hacer cruces de información con variables como el sector económico y obtener entre otros resultados una matriz de empleo basada en registros administrativos.

Figura 19. Distribución ocupacional con información completa del cargo. Universo de afiliados al IESS.



Fuente: Información de registros administrativos del LDLE-INEC. Elaborado por los autores.

#### 4.3. Matriz de empleo basada en registros administrativos

Como resultado del ejercicio de construcción de la variable del cargo ocupacional en el universo de referencia de la base de trabajadores del LDLE para todos los meses desde enero de 2008 y con la disponibilidad de la variable de rama de actividad para el mismo periodo, fue posible obtener matrices de empleo históricas. Las Tabla 5 y 6 muestran las matrices de empleo para los años 2014 y 2020 obtenidas a través del promedio de las matrices de empleo mensuales de cada año. Si bien en todas las ramas de actividad se destaca la presencia de trabajadores en ocupaciones elementales, también hay grupos ocupacionales cuya participación es importante dependiendo de la rama de actividad. Por ejemplo, los profesionales tienen una participación importante en ramas como el Suministro de electricidad, gas, vapor y aire acondicionado (D), Información y comunicación (J), Administración pública y defensa; planes de seguridad social de afiliación obligatoria (O), Enseñanza (P), mientras que los trabajadores de los servicios o vendedores tienen una participación alta en Comercio al por mayor y al por menor; reparación de los vehículos de motor y de las motocicletas (G), Alojamiento y servicios de comida (I), Actividades financieras y de seguros (K) y Actividades administrativas y servicios de apoyo (N).

Entre 2014 y 2018 no se observaron cambios importantes en la estructura ocupacional del empleo. Una de las excepciones ocurrió en el caso de los trabajadores en ocupaciones técnicas, ya que, con respecto al total del empleo, su participación aumentó en 3,2 puntos porcentuales. Las ramas en donde la participación de estos trabajadores creció más fue en Salud (Q), Construcción (F) y Minas (B). Sin embargo, estos cambios en la participación no responden estrictamente a un aumento en términos absolutos en el número de trabajadores técnicos, pues en el caso de Salud hay una caída en la cantidad de técnicos, pero la mayor caída de trabajadores en otras ocupaciones provoca que la participación cambie. En un documento posterior se analizará en detalle los resultados relacionados a la matriz de empleo basada en registros administrativos y otros aspectos de relevancia para entender la dinámica del mercado laboral formal ecuatoriano en los últimos años.

Tabla 5. Matriz de empleo basada en registros administrativos - 2020.

Rama/ Ocupación	Directores	Profesionales	Técnicos	Administrativo	Vendedores	Agricultores	Artesanos	Operadores	Elementales	Total
A	8,329	2,007	7,723	5,565	2,867	484,878	2,555	4,194	75,783	593,901
B	2,745	4,065	4,651	1,911	439	39	1,329	2,736	14,927	32,841
C	23,318	14,072	39,989	19,056	15,092	14,174	33,892	29,003	158,205	346,801
D	939	7,576	4,358	752	109	16	736	842	2,811	18,139
E	673	2,289	6,243	2,049	395	297	305	3,074	6,871	22,196
F	6,134	3,940	12,747	3,738	1,918	192	19,263	6,373	29,798	84,101
G	55,166	20,226	83,843	44,825	100,146	8,565	14,511	14,461	147,786	489,528
H	16,011	4,458	16,187	13,558	4,601	657	1,959	35,751	46,125	139,307
I	9,133	1,804	6,682	4,167	23,909	377	2,538	1,066	45,675	95,350
J	3,813	10,724	12,611	5,367	1,763	30	1,858	783	12,638	49,587
K	5,410	8,024	17,372	17,712	14,284	32	143	247	7,263	70,487
L	4,474	1,976	3,682	2,199	5,984	1,883	1,048	626	15,197	37,069
M	17,096	13,916	19,749	13,698	4,792	1,338	6,250	3,669	51,069	131,577
N	5,127	2,811	10,819	10,366	70,855	804	1,091	2,312	36,765	140,950
O	10,438	247,805	72,205	40,962	13,900	197	227	9,058	50,104	444,897
P	5,688	71,100	10,986	7,345	6,599	487	487	897	15,426	119,014
Q	5,311	42,591	62,621	11,155	3,745	229	1,703	2,166	19,914	149,435
R	1,241	1,791	1,917	1,426	1,095	88	274	198	6,555	14,586
S	5,186	5,273	6,225	4,013	3,241	595	837	2,159	32,301	59,830
T									83,071	83,071
U	12	30	35	54	8		1	18	158	315
<b>Total</b>	<b>186,244</b>	<b>466,477</b>	<b>400,644</b>	<b>209,917</b>	<b>275,741</b>	<b>514,879</b>	<b>91,006</b>	<b>119,633</b>	<b>858,441</b>	<b>3,122,981</b>

Fuente: Información de registros administrativos del LDLE-INEC. Elaborado por los autores.

Nota: Ver en el Anexo 1 la nomenclatura de la CIU Rev.4 de la rama de actividad.



Tabla 6. Matriz de empleo basada en registros administrativos - 2014.

Rama/ Ocupación	Directores	Profesionales	Técnicos	Administrativo	Vendedores	Agricultores	Artesanos	Operadores	Elementales	Total
A	8,697	2,529	4,847	6,733	2,928	463,603	3,064	4,731	72,569	569,701
B	3,126	4,908	3,913	1,346	694	82	2,792	4,621	18,928	40,411
C	27,672	13,603	34,577	24,223	16,532	14,205	52,556	35,282	176,984	395,634
D	1,062	7,361	4,052	996	166	13	676	1,213	4,455	19,994
E	427	2,438	4,276	1,414	1,426	63	339	1,981	5,503	17,867
F	7,848	5,394	14,999	6,154	2,986	333	39,965	14,661	61,917	154,257
G	65,692	17,258	67,860	55,583	116,109	8,401	18,232	16,349	140,895	506,380
H	18,442	5,906	12,514	15,502	6,697	670	2,510	39,706	47,731	149,678
I	10,445	1,965	5,787	6,650	33,215	485	3,385	1,247	57,484	120,663
J	4,385	10,879	14,159	4,911	2,248	47	1,998	991	12,600	52,217
K	5,780	6,487	15,356	20,987	12,696	132	109	357	6,703	68,607
L	5,611	2,397	3,663	3,372	7,843	2,048	2,063	922	16,726	44,643
M	19,195	13,723	18,927	19,293	6,166	1,188	14,532	6,269	55,741	155,032
N	5,789	3,001	8,141	11,911	81,607	797	1,697	2,992	34,757	150,692
O	10,337	208,474	53,550	56,375	13,231	61	197	6,353	54,699	403,276
P	6,965	101,021	8,759	11,347	10,487	262	450	1,417	17,381	158,089
Q	5,076	37,212	37,247	11,551	5,628	211	1,585	1,626	17,494	117,629
R	1,355	2,145	1,557	1,926	1,919	132	305	191	7,176	16,705
S	6,138	5,045	5,247	5,131	4,433	499	1,227	1,766	29,888	59,373
T									101,846	101,846
U	24	26	45	72	9	1	1	26	138	341
<b>Total</b>	214,065	451,774	319,474	265,475	327,019	493,232	147,681	142,698	941,616	3,303,035

Fuente: Información de registros administrativos del LDLE-INEC. Elaborado por los autores.

Nota: Ver en el Anexo 1 la nomenclatura de la CIU Rev.4 de la rama de actividad.

## 5. Conclusiones

Los registros administrativos de las instituciones públicas son fuentes de información que se actualizan constantemente y que al ser transformadas a registros de uso estadístico tienen un alto potencial analítico que se fortalece aún más cuando son enlazados. El LDLE del INEC es un sistema de información que ha avanzado de manera importante hacia este objetivo y que ha permitido analizar temáticas relacionadas al mercado laboral y el sector productivo.

La construcción de la variable del cargo ocupacional es una actividad que nace con la necesidad de construir una matriz de empleo para el caso ecuatoriano. En este documento se resumieron todos los pasos que se siguieron para agregar esta variable a las bases de afiliados al IESS, para lo cual fue necesario emplear fuentes de información de varias entidades aparte del IESS, como el MEF y el MDT.

La primera parte del proceso de construcción de la variable del cargo involucró un alto componente manual enfocado en la codificación de información descriptiva del cargo ocupacional a la CIUO. En este proceso se identificaron algunas limitaciones en la información disponible, principalmente por el hecho de no contar con información sistematizada desde la fuente, lo cual dificulta en varios casos hallar la correspondencia con los grupos ocupacionales definidos en la CIUO. Otra limitación hallada es que en principio la variable del cargo no pudo ser obtenida en un conjunto importante de trabajadores, lo cual limita realizar un análisis histórico y un análisis con registros completos en los años más recientes.

La segunda parte del proceso se enfocó en identificar el cargo de ocupación de aquellas personas que previamente no pudieron ser clasificadas a través de técnicas de clasificación basadas en árboles, priorizando la elección de los parámetros que dan como resultado un mayor nivel de precisión a nivel de microdato y de distribución ocupacional agregada. Por la naturaleza de cada subuniverso del cual parte la construcción del cargo, se realizaron ejercicios independientes a fin de obtener un resultado más preciso para cada universo. Como resultado de todo este proceso se logró generar una variable de cargo ocupacional en las bases de afiliados al IESS en todos los meses de 2008 a 2020. Para fines analíticos esta información resulta importante ya que permite analizar con mayor exactitud los cambios del empleo en el tiempo, así como la distribución ocupacional. La siguiente acción a corto plazo es generar un algoritmo que permita calcular la variable del cargo ocupacional en las bases futuras.

Una vez que se contó con la variable del cargo ocupacional en las bases de trabajo del LDLE se procedió a hacer el cruce con la variable de rama de actividad para de este modo obtener la matriz de empleo basada en registros administrativos. En las matrices de los años 2014 y 2020 se pudo evidenciar que la distribución ocupacional es consistente al interior de las ramas de actividad y que la estructura se mantiene salvo en el caso de los técnicos en donde la participación aumentó entre estos años.

Finalmente, una recomendación que surge de toda esta actividad es la necesidad de mejorar la captación de la información en la etapa de generación del registro, lo cual permitiría contar con información de calidad, precisa y útil para la toma oportuna de decisiones. En el caso específico del cargo ocupacional, también es recomendable que tanto las comisiones sectoriales como los sistemas de entrada de la información del IESS gestionen su trabajo desde un estándar internacional de clasificación de ocupaciones como la CIUO, o un diccionario de ocupaciones con las cuales se encuentren mejor familiarizados los empleadores y que tenga una correspondencia inequívoca con los grupos ocupacionales de la CIUO.

## 6. Bibliografía

- Albán, M.S., y Sánchez, D.M. (2018). Prediction of university dropout through technological factors: a case study in Ecuador. *Revista Espacios*, 39(52).
- Alvarez-Mendoza, C. I., Teodoro, A., Freitas, A., y Fonseca, J. (2020). Spatial estimation of chronic respiratory diseases based on machine learning procedures—an approach using remote sensing data and environmental variables in Quito, Ecuador. *Applied Geography*, 123, 102273.
- Arowolo, M. O., Adebisi, M., Adebisi, A., y Okesola, O. (2020). PCA model for RNA-Seq malaria vector data classification using KNN and decision tree algorithm. En 2020 international conference in mathematics, computer engineering and computer science (ICMCECS) (pp. 1-8). IEEE.
- BLS (2018). *Employment Projections. Handbook of Methods*.
- DANE (2011). *Matriz de empleo en la base 2005 de las cuentas nacionales*. Departamento Administrativo Nacional de Estadística.
- Gweon, H., Schonlau, M., Kaczmirek, L., Blohm, M., y Steiner, S. (2017). Three methods for occupation coding based on statistical learning. *Journal of Official Statistics*, 33(1), 101.
- International Labour Office (2012). *International Standard Classification of Occupations Structure, group definitions and correspondence tables. ISCO–08, Volume I*.
- Khatri, S., Arora, A., y Agrawal, A. P. (2020). Supervised machine learning algorithms for credit card fraud detection: a comparison. En 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 680-683). IEEE.
- Naciones Unidas (2009). *Clasificación Industrial Internacional Uniforme de todas las actividades económicas (CIIU), Revisión 4*. Departamento de Asuntos Económicos y Sociales. Informes estadísticos Serie M, No. 4/Rev. 4
-

- Princy, R. J. P., Parthasarathy, S., Jose, P. S. H., Lakshminarayanan, A. R., y Jeganathan, S. (2020). Prediction of cardiac disease using supervised machine learning algorithms. In 2020 4th international conference on intelligent computing and control systems (ICICCS) (pp. 570-575).
- Santos, F., Graw, V., y Bonilla, S. (2019). A geographically weighted random forest approach for evaluate forest change drivers in the Northern Ecuadorian Amazon. *PloS one*, 14(12)
- Schierholz, M., y Schonlau, M. (2021). Machine Learning for Occupation Coding: A Comparison Study. *Journal of Survey Statistics and Methodology*, 9(5), 1013-1034.
- Tapia-Armijos, M. F., Homeier, J., Espinosa, C. I., Leuschner, C., y de la Cruz, M. (2015). Deforestation and forest fragmentation in South Ecuador since the 1970s—losing a hotspot of biodiversity. *PloS one*, 10(9).
- Wilson R., Barnes, S.A., May-Gillings, M., Patel, S. y Bui, H. (2020). Working Futures 2017-2027: Long-run labour market and skills projections for the UK: Main report. Department for Education, UK Government.
- Zukersteinova, A. y Strietska-Ilina, O. (Eds.) (2007). Panorama: Towards European skill needs forecasting. Cedefop Panorama series, No. 137.
-

## 7. Anexos

### Anexo 1. Clasificación Industrial Internacional Uniforme, Revisión 4.

CIU Rev. 4 Sección	Descripción
A	Agricultura, Silvicultura y Pesca
B	Explotación de Minas y Canteras
C	Industrias Manufactureras
D	Suministro de Electricidad, gas, vapor y aire acondicionado
E	Suministro de agua; alcantarillado, gestión de desechos y actividades de saneamiento
F	Construcción
G	Comercio al por mayor y al por menor; reparación de vehículos automotores y motocicletas
H	Transporte y almacenamiento
I	Actividades de alojamiento y de servicio de comidas
J	Información y Comunicación
K	Actividades financieras y de seguros
L	Actividades inmobiliarias
M	Actividades profesionales, científicas y técnicas
N	Actividades de servicios administrativos y de apoyo
O	Administración pública y defensa; planes de seguridad social de afiliación obligatoria
P	Enseñanza
Q	Actividades de atención de la salud humana y de asistencia social
R	Artes, entretenimiento y recreación
S	Otras actividades de servicios
T	Actividades de los hogares como empleadores; actividades no diferenciadas de los hogares como productores de bienes y servicios para uso propio
U	Actividades de organizaciones y órganos extraterritoriales

*Fuente: Naciones Unidas (2009). Elaborado por los autores.*

### Anexo 2. Clasificación Internacional Uniforme de Ocupaciones 2008

CIUO-08 Grupo	Descripción
1	Directores y Gerentes
2	Profesionales, Científicos e Intelectuales
3	Técnicos y Profesionales de Nivel Medio
4	Personal de Apoyo Administrativo
5	Trabajadores de los Servicios y Vendedores de Comercios y Mercados
6	Agricultores y Trabajadores Calificados Agropecuarios, Forestales y Pesqueros
7	Oficiales, operarios y artesanos de artes mecánicas y de otros oficios
8	Operadores de Instalaciones y Máquinas y Ensambladoras
9	Ocupaciones Elementales

*Fuente: International Labour Office (2012). Elaborado por los autores.*

## Anexo 3. Fuentes de información para la construcción del cargo de ocupación

Base de datos	Tipo	Periodicidad	Detalles
<b>IESS (Avisos de Entradas y Salidas)</b>	Registro administrativo	Mensual	<ul style="list-style-type: none"> <li>• Información del cargo ocupacional solamente de los trabajadores del sector privado.</li> <li>• La información del cargo ocupacional se construye a partir de las clasificaciones realizadas por las comisiones sectoriales que determinan los salarios mínimos sectoriales.</li> <li>• La variable del cargo codificado se construye a partir de los cargos establecidos por las comisiones sectoriales, los cuales no son equivalentes a la CIUO. La variable codificada del cargo ocupacional se construye a partir de las descripciones del cargo.</li> <li>• La información es declarada por los empleadores al momento de registrar una nueva afiliación y no está sujeta a un filtro de validación al momento de ingresar la información.</li> </ul>
<b>MDT (Base de Contratos)</b>	Registro administrativo	Mensual	<ul style="list-style-type: none"> <li>• No se encuentran todas las relaciones laborales registradas en el IESS, principalmente en años anteriores a 2015.</li> <li>• Información del cargo ocupacional solamente de los trabajadores del sector privado</li> <li>• La variable codificada del cargo ocupacional se construye a partir de las descripciones del cargo.</li> <li>• La información es declarada por los empleadores y no está sujeta a un filtro de validación al momento de ingresar la información.</li> </ul>
<b>MEF (Distributivo de Personal)</b>	Registro administrativo	Mensual (meses de diciembre entre 2008 y 2017 y todos los meses desde julio de 2018 en adelante)	<ul style="list-style-type: none"> <li>• El universo corresponde a los servidores públicos de la Administración Central.</li> <li>• La información del cargo ocupacional se encuentra originalmente en texto.</li> </ul>
<b>Gobiernos Autónomos Descentralizados (GADs)</b>	Registro de páginas web de transparencia	Mensual (meses de diciembre entre 2014 y 2019)	<ul style="list-style-type: none"> <li>• No hay una publicación sistematizada para gran parte de los GADs parroquiales. Por lo tanto, este grupo, el cual representa el 8% (diciembre 2020) del empleo de todos los GADs, fue descartado del proceso de recuperación de la información.</li> <li>• El cargo ocupacional se obtuvo a través de una correspondencia con los salarios de cada GAD.</li> <li>• La descarga y consolidación de la información tiene un alto componente manual y depende de la información pública disponible.</li> </ul>
<b>Empresas Públicas (EPs)</b>	Registro de páginas web de transparencia	Mensual (meses de diciembre entre 2014 y 2019)	<ul style="list-style-type: none"> <li>• Se priorizó la recuperación de información para las empresas públicas que más empleo concentran.</li> <li>• El cargo ocupacional se obtuvo a través de una correspondencia con los salarios de cada EP.</li> <li>• La descarga y consolidación de la información tiene un alto componente manual y depende de la información pública disponible.</li> </ul>

*Fuente: Elaborado por los autores.*



Estudio Temático  
**Especial Demanda**  
**Laboral**



@ecuadorencifras



@ecuadorencifras



@InecEcuador



t.me/equadorencifras



INEC/Ecuador



INECEcuador



INEC Ecuador