

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/334828677>

Mercado Laboral Chileno en la Cuarta Revolución Industrial

Article · May 2019

CITATION

1

READS

36

3 authors, including:



[Agustín García](#)

Pontificia Universidad Católica de Chile

2 PUBLICATIONS 1 CITATION

[SEE PROFILE](#)



[Hans Schlechter](#)

Pontificia Universidad Católica de Chile

4 PUBLICATIONS 1 CITATION

[SEE PROFILE](#)

Juan Bravo M.
Agustín García
Hans Schlechter

www.clapesuc.cl

Mercado Laboral Chileno para la Cuarta Revolución Industrial

Documento de Trabajo N° 59

Mercado Laboral Chileno para la Cuarta Revolución Industrial

Juan Bravo M.

Agustín García

Hans Schlechter

Clapes UC

Santiago, Chile

Mayo, 2019

Resumen

- Se estima la probabilidad de automatización de las ocupaciones en el mercado laboral chileno utilizando un modelo predictivo de aprendizaje supervisado para clasificación binaria. Se encuentra que la probabilidad de automatización de un empleo promedio en Chile es de 42,2%, mientras que el 17,0% de los ocupados presenta un alto riesgo de automatización.
- Se amplía la base de datos de la Encuesta CASEN 2017 al predecir el uso de habilidades en el ambiente laboral para los ocupados del mercado laboral chileno, utilizando la base de datos de la "Evaluación Internacional de las Competencias para Adultos" (OCDE, 2015), en la que se incluye Chile. El uso de CASEN permite, a diferencia de otros trabajos en el área, contar con información que refleja con mayor riqueza la idiosincrasia del mercado laboral chileno.
- En el estudio se asignan las categorías de automatización (Automatizable/No Automatizable), definidas en Frey y Osborne (2017), a las ocupaciones equivalentes en el mercado laboral chileno. Esto permite identificar 90 ocupaciones en el marco de la Clasificación Internacional Uniforme de Ocupaciones de 1988 (CIUO-88) que serían automatizables o no automatizables en un período indeterminado de tiempo.
- El análisis del empleo de alto riesgo de automatización arroja que los hombres presentan mayor riesgo que las mujeres (18,8% contra 14,7%, respectivamente), así como los jóvenes entre 15 y 24 (19,8%) y los adultos entre 35 y 44 años (20,7%). Además, los quintiles de ingreso intermedios presentan un mayor porcentaje de trabajadores en alto riesgo de automatización que los otros. Las industrias de mayor porcentaje de trabajadores en alto riesgo son Transporte, almacenamiento y comunicaciones (46,9%), Explotación de minas y canteras (27,6%) e Intermediación financiera (23,2%). Según nivel de calificación, los trabajadores de mediana calificación presentan el mayor porcentaje de trabajos en alto riesgo (26,1%).

- Considerando los impactos del avance de la automatización, proponemos una reforma al sistema de capacitación y formación continua para enfrentar los desafíos de una economía 4.0. Dichas propuestas se focalizan principalmente en dos aspectos: realizar una reforma a la arquitectura institucional de sistema de capacitación y formación continua en Chile y en reformar los mecanismos de financiamiento para dar acceso a los trabajadores a capacitaciones con un impacto positivo en empleo y salarios.
- El sistema de capacitación y formación continua requiere una reforma profunda, por lo que es necesario aprovechar la actual oportunidad legislativa que se abre con la discusión del Proyecto de Ley “sobre Modernización de la Franquicia Tributaria y Modificación de Fondos Públicos que Indica”. Si bien el Proyecto contiene elementos positivos que contribuyen a mejorar el sistema, es insuficiente ya que es necesario contar con una arquitectura institucional coordinada entre los distintos actores y que permita alinear los incentivos público-privados de éstos.

Tabla de Contenidos

Introducción	1
Capítulo 1 ¿Cuál es el riesgo de automatización en Chile?	3
I Literatura de la automatización en el empleo.....	5
I.1 Evidencia internacional sobre el riesgo de automatización	5
I.2 Evidencia en la literatura para el mercado laboral chileno	10
II Metodología del estudio para Chile	12
II.1 ¿Qué entendemos por <i>machine learning</i> ?	13
II.2 Modelos predictivos de aprendizaje supervisado	16
II.3 Bases de datos	19
II.4 Fase 1: Predicción de las habilidades en CASEN	22
II.5 Fase 2: Predicción de la probabilidad de automatización en Chile	31
III Riesgo de automatización del mercado laboral chileno	34
III.1 Principales resultados para Chile	34
III.2 Radiografía del segmento de alto riesgo de automatización	42
IV Principales conclusiones sobre la automatización en Chile	48
Capítulo 2 Un sistema de capacitación para un mercado laboral 4.0	50
I Diagnóstico de nuestro sistema de capacitación actual.....	53
I.1 Necesidad de una nueva arquitectura institucional	54
I.2 Capacitación de baja efectividad.....	58
I.3 Deficiencias en el mecanismo de franquicia tributaria	60
I.4 Problemas en el diseño de incentivos de los intermediarios	62
II Propuestas para una reforma profunda a nuestro sistema de capacitación	65
Propuesta 1 Generación de un Marco de Cualificaciones Nacional, integrado y en actualización permanente	69
Propuesta 2 Crear Consejos de Competencias	75
Propuesta 3 Nueva Agencia de Calidad de la Educación para la Formación Técnica Profesional y Capacitación	77

Propuesta 4	Reforma profunda a los programas de capacitación y modificaciones a los OTEC	79
Propuesta 5	Modificaciones al Sistema Nacional de Certificación de Competencias (ChileValora)	81
Propuesta 6	Cambios a la franquicia tributaria (Impulsa Personas)	82
Propuesta 7	Fondo para segmentos no cubiertos por la franquicia tributaria	85
Propuesta 8	Reforma a los OTIC	86
Propuesta 9	Protección del empleo en tiempos de crisis mediante capacitación	87
III	Conclusiones sobre el sistema de capacitación 4.0	89
Referencias		90
Anexos		93
A.	Lista de ocupaciones según su categoría de automatización, según Frey y Osborne (2017)	93
B.	Algoritmos de <i>machine learning</i>	96
B.1	Regresión Logística (LR)	97
B.2	Random Forest (RF)	98
B.3	K-Nearest Neighbors (KNN)	100
C.	Métricas de evaluación de desempeño	103
C.1	Exactitud	104
C.2	F1-Score	104
C.3	ROC-AUC	104
Glosario		106
 Lista de Figuras		
Figura 1.1	– Conjunto de tecnologías de la inteligencia artificial	13
Figura 1.2	– Diagrama de un proceso general para algoritmos predictivos	17
Figura 1.3	– Diagrama de flujo para modelo predictivo de uso de habilidades	22
Figura 1.4	– Comparación de intensidad de uso de habilidades según grupo ocupacional	30
Figura 1.5	– Diagrama de flujo para modelo predictivo de probabilidad de automatización	31

Figura 1.6 – Boxplot de la probabilidad de automatización, según nivel de calificación.....	36
Figura 1.7 – Boxplot de la probabilidad de automatización, según sexo.....	37
Figura 1.8 – Boxplot de la probabilidad de automatización, según rango etario.....	38
Figura 1.9 – Boxplot de la probabilidad de automatización, según quintil de ingreso autónomo per cápita del hogar.....	39
Figura 1.10 – Boxplot de la probabilidad de automatización, según rama de actividad económica.....	41
Figura 1.11 – Descomposición del empleo total y de alto riesgo de automatización, según nivel de calificación.....	43
Figura 1.12 – Descomposición del empleo total y de alto riesgo de automatización, según sexo.	44
Figura 1.13 – Descomposición del empleo total y de alto riesgo de automatización, según rango etario.....	45
Figura 1.14 – Descomposición del empleo total y de alto riesgo de automatización, según quintil de ingresos autónomos per cápita del hogar.....	46
Figura 1.15 – Descomposición del empleo total y de alto riesgo de automatización, según rama de actividad económica.....	47
Figura 2.1 – Distribución de ocupados que ha participado en los últimos 12 meses en capacitación laboral de al menos 8 horas de duración, por nivel de calificación de empleo y riesgo de automatización.	52
Figura 2.2 – Actual arquitectura institucional del Sistema de Capacitación y Formación Continua	54
Figura 2.3 – Marco Nacional de Cualificaciones de Nueva Zelanda.....	55
Figura 2.4 – Mapa de propuestas para un nuevo sistema de capacitación.	65
Figura 2.5 – Principales elementos para un sistema de formación continua 4.0.....	66
Figura 2.5 – Nueva Arquitectura Institucional del Sistema de Capacitación y Formación Continua	68
Figura B.1 – Ejemplo de regresión logística para clasificación binaria.	98

Figura B.2 – Árbol de decisión	99
Figura B.3 – Ejemplo de Random Forest para clasificación binaria.....	99
Figura B.4 – Ejemplo de K-Nearest Neighbors para clasificación.....	101
Figura C.1 – Curva receptor-operador (ROC) y área bajo la curva (AUC).....	105

Lista de Tablas

Tabla 1.1 – Habilidades no automatizables en PIAAC.....	20
Tabla 1.2 – Estadísticas descriptivas de las variables características seleccionadas.....	25
Tabla 1.3 – Estadísticas descriptivas de la intensidad de uso de habilidades en PIAAC....	26
Tabla 1.4 – Resultado de entrenamiento de frecuencia de habilidades, para los algoritmos seleccionados	28
Tabla 1.5 – Estadísticas descriptivas de habilidades predichas en CASEN.....	29
Tabla 1.6 – Resultados entrenamiento de riesgo de automatización	34
Tabla 1.7 – Distribución de riesgo de automatización por nivel de calificación.....	36
Tabla 1.8 – Distribución de riesgo de automatización por rango etario.	38
Tabla 1.9 – Distribución de riesgo de automatización por quintil de ingresos.....	39
Tabla 1.10 – Distribución de riesgo de automatización por rama de actividad económica. 42	
Tabla 2.1 – Cobertura de franquicia tributaria, según tramos de remuneración	60
Tabla 2.2 – Propuesta de cambios al diseño de incentivos y estructura de proveedores e intermediarios del sistema de capacitación y formación continua	67
Tabla 2.3 – Descriptores de niveles seleccionados en marcos nacionales de cualificaciones de Australia	73
Tabla 2.4 – Descriptores de niveles seleccionados en marcos nacionales de cualificaciones de Irlanda	74

Introducción

El avance de la tecnología, en conjunto con la consiguiente transformación del mercado laboral, no es algo único del siglo XXI. Esto ya se ha experimentado en siglos anteriores, y en particular, desde el advenimiento de la Primera Revolución Industrial a fines del siglo XVIII, que revolucionó las formas de producir y los mercados laborales a nivel mundial. Sin embargo, esta nueva época conocida como “Cuarta Revolución Industrial”, tiene un sello característico que es la velocidad del desarrollo y adopción tecnológica.

De acuerdo a lo que señala Schwab (2016), los disruptores de hoy tales como Airbnb, Uber, o Alibaba, entre otros, eran relativamente desconocidos hace apenas algunos años. Hoy, sin embargo, dichas plataformas tienen millones de usuarios. Un ejemplo de ello es Uber que, si bien comenzó en 2009 y con sólo 3 automóviles de prueba en Nueva York, hacia 2017 ya había ganado millones de usuarios y tenía presencia en más de 600 ciudades alrededor del mundo¹. Esta velocidad en el desarrollo y adopción de las innovaciones tecnológicas es lo que está provocando preocupaciones en torno a los impactos en el mercado laboral y principalmente sobre la destrucción de empleo que esto podría generar.

Se suele considerar que la tecnología tiende a reemplazar a aquellos empleos que concentran una gran proporción de tareas rutinarias. Sin embargo, la literatura reciente afirma que los importantes avances en inteligencia artificial han permitido que la tecnología pueda reemplazar un espectro de tareas más amplio y no sólo las rutinarias (Frey y Osborne, 2017). Esto, ha motivado en la literatura la estimación de la probabilidad de automatización de los distintos empleos que hoy existen alrededor del mundo, a fin de evaluar los riesgos de que el trabajo humano sea reemplazado por tecnología y las nuevas oportunidades para el desarrollo laboral.

¹ <http://forbes.es/up-down/40962/uber-las-cifras-del-gigante-del-transporte-alternativo/>

Aunque los avances que se han hecho en este campo han sido importantes, la literatura muestra una notable limitación, ya que gran parte de ella utiliza como base las probabilidades estimadas por Frey y Osborne (2017) para el mercado laboral de Estados Unidos. Esto, como explicaremos más adelante, puede realizar predicciones sesgadas debido a que los mercados laborales pueden diferir entre países.

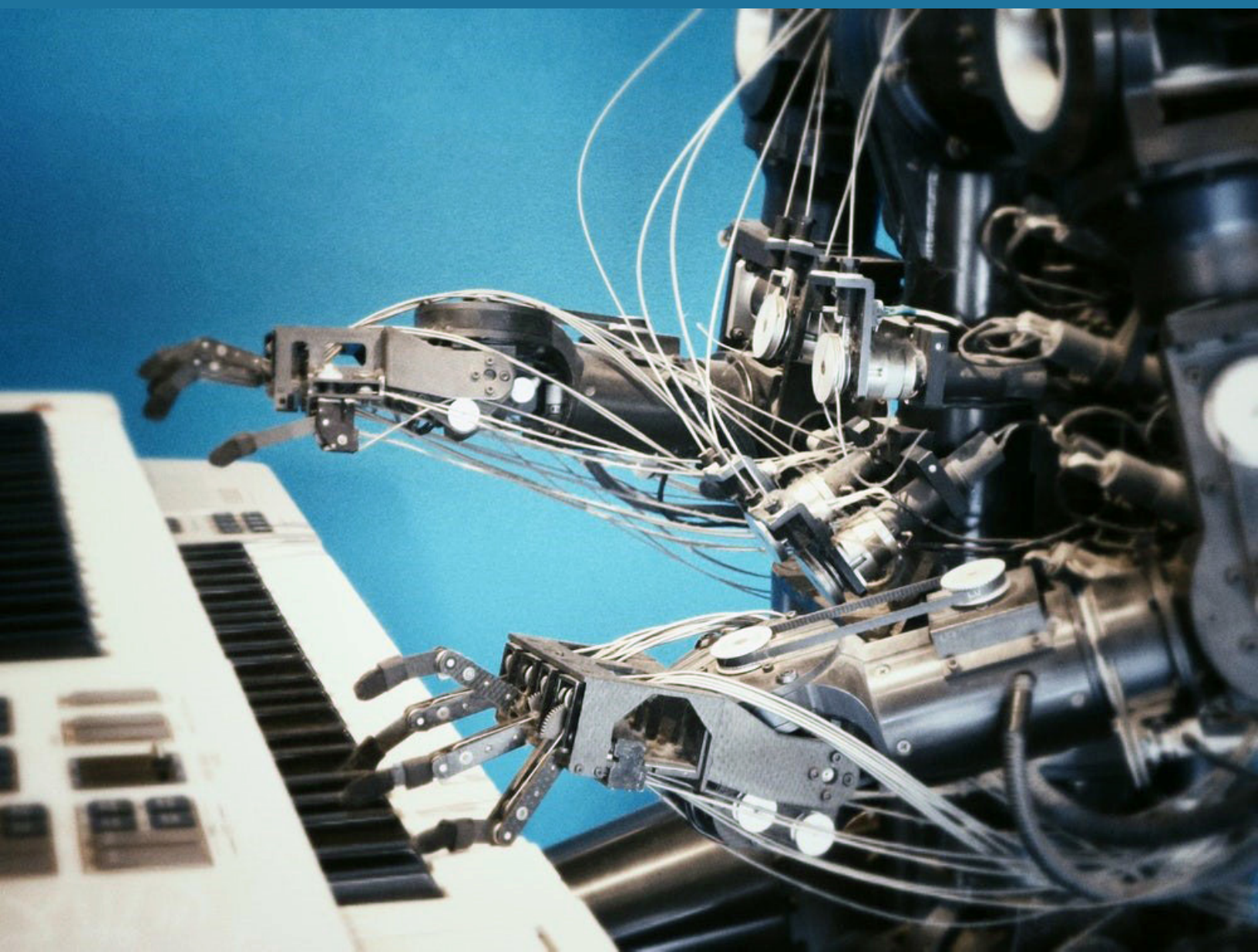
Este documento continúa con la línea de investigación en automatización iniciada en el Documento de Trabajo 50 de Clapes UC (Bravo, García y Schlechter, 2018), en donde se detalló la problemática económica y laboral de este fenómeno y se plantearon lineamientos generales en materia de política pública para abordar los desafíos que impone.

El primer capítulo de este trabajo estima las probabilidades de automatización en el mercado laboral chileno con información que refleja mejor la idiosincrasia de éste. Para ello, se hace uso de técnicas de Aprendizaje Automático (*Machine Learning* en inglés) sobre bases datos que caracterizan el perfil de los trabajadores chilenos tales como CASEN 2017 y la encuesta de la OCDE conocida como “Evaluación Internacional de las Competencias para Adultos” (PIAAC, por sus siglas en inglés), que incluye información para Chile. Esto permite realizar un diagnóstico sobre el posicionamiento del mercado laboral chileno frente al avance de la tecnología.

Dado el contexto de progreso tecnológico permanente en el que vivimos, la capacitación y la formación continua toman un rol crucial para evitar que los trabajadores se queden anclados en ocupaciones altamente reemplazables y sin las capacidades necesarias para poder adaptarse y seguir trabajando. En consecuencia, el segundo capítulo de este documento identifica los principales problemas del actual sistema de capacitación y formación continua, así como también sugiere propuestas para reformarlo de manera profunda, contribuyendo así a reducir el riesgo de automatización, considerando que éste es un mecanismo fundamental para preparar a los trabajadores chilenos a enfrentar los desafíos que impone el acelerado avance tecnológico.

CAPÍTULO 1

¿Cuál es el riesgo de automatización en Chile?



La literatura sobre el impacto de la tecnología en el mercado laboral es extensa, sin embargo, los estudios enfocados en conocer los impactos de las tecnologías 4.0 son más escasos. Este tipo de tecnologías, que se caracteriza por dotar a dispositivos de la capacidad de interactuar y aprender del entorno, han experimentado un importante desarrollo en los últimos años y, por lo mismo, ha llevado a la literatura a cuestionarse en qué grado la inteligencia artificial sería capaz de sustituir tareas realizadas por humanos y cuáles serían los potenciales efectos.

El objetivo del presente trabajo consiste en estimar la probabilidad de automatización del empleo en Chile utilizando información propia del mercado laboral chileno². A diferencia de los estudios internacionales que incluyen a Chile, este trabajo utiliza una encuesta de caracterización socioeconómica nacional (CASEN) en conjunta con una evaluación de habilidades para adultos (PIAAC, realizada por la OCDE y que incluye información para Chile) para efectuar una mejor caracterización de la realidad del mercado laboral chileno. Mediante el uso de algoritmos, los datos de competencias cognitivas y habilidades laborales contenidos en PIAAC son imputados a los datos contenidos en CASEN. Luego, siguiendo procedimientos similares a los de Nedelkoska y Quintini (2018), se obtiene la probabilidad de automatización para las ocupaciones en Chile.

Este capítulo se compone de cuatro secciones: en la primera, se presenta una revisión de los trabajos que han estudiado el impacto de las nuevas tecnologías de la revolución 4.0 en el mercado laboral, poniendo atención en los resultados de estudios que han analizado el mercado laboral chileno bajo esta perspectiva. En la segunda, se presenta el marco metodológico del estudio, el que incluye una descripción conceptual de lo que entendemos por *machine learning* y aprendizaje supervisado, los que serán utilizados para las distintas fases del análisis. Posteriormente, presentamos los resultados obtenidos sobre

² Para esto, se emplearán técnicas de *machine learning*, las que, como se abordarán en la sección metodológica, se refieren a una forma de inteligencia artificial que permite a un sistema o modelo “aprender” desde los datos.

el riesgo de automatización en el mercado laboral chileno, realizando un profundo diagnóstico de la población más susceptible a los cambios tecnológicos. La última sección concluye con los principales resultados y recomendaciones provenientes del análisis.

I Literatura de la automatización en el empleo

I.1 Evidencia internacional sobre el riesgo de automatización

Si bien gran parte de la literatura señala que las ocupaciones más propensas a ser automatizadas son aquellas compuestas por tareas rutinarias, Frey y Osborne (2017) afirman que los progresos logrados en inteligencia artificial han permitido el desarrollo de robots y algoritmos que permiten automatizar un rango de actividades humanas aún más amplio, incluyendo tareas no rutinarias. Años atrás, se señalaba que la conducción de vehículos era una tarea que no era posible automatizar, debido a que involucra un conjunto de factores asociados a la toma de decisiones con información incompleta, que no son capaces de ser sistematizados en reglas y procedimientos definidos (Levy y Murnane, 2005). No obstante, Google anunció en 2010 que había logrado el diseño de automóviles completamente autónomos (Brynjolfsson and McAfee, 2012). Por lo tanto, según lo autores, la automatización de una actividad no dependerá necesariamente de cuán rutinaria sea, sino de que pueda ser definida en términos de reglas codificables y algorítmicos.

El trabajo de Frey y Osborne (2017) presenta una metodología para estimar la susceptibilidad de automatización del empleo en Estados Unidos basada en los algoritmos de *machine learning*. Para ello, un grupo de expertos del Departamento de Ciencias de la Ingeniería de la Universidad de Oxford clasificaron, en base a las estructuras de tareas en Estados Unidos, cuáles eran las ocupaciones automatizables dado el estado de desarrollo de la tecnología en ese momento. En particular, tuvieron que responder la pregunta “¿*pueden*

las tareas de este trabajo ser lo suficientemente especificadas, dada la disponibilidad de Big Data, a fin de que sean llevadas a cabo por equipos de última generación controlados por computadoras?”. De esta manera, las ocupaciones completamente automatizables fueron calificadas con el número 1 (uno), mientras que las no automatizables fueron calificadas con el número 0 (cero). Los autores analizaron la versión del año 2010 de la base O*NET³, que contiene información sobre la composición de tareas de un total de 903 empleos en Estados Unidos—clasificadas bajo la Clasificación Ocupacional Estándar del Departamento de Trabajo de Estados Unidos (SOC). De dicha lista de empleos se seleccionaron 702 ocupaciones, excluyendo aquellas para las cuales la O*NET no poseía datos completos y para reducir riesgos de sesgos subjetivos, fueron clasificadas sólo aquellas ocupaciones de las que los expertos tuviesen absoluta certeza de la categoría a la que los asignaban. En total, sólo 70 ocupaciones fueron identificadas con alguna de las dos categorías mientras que las restantes fueron dejadas a los algoritmos.

La metodología de Frey y Osborne comprendió dos fases. En la primera, los autores identificaron aquellas habilidades en la base O*NET que estuviesen relacionadas con “cuellos de botella tecnológicos”, refiriéndose así a habilidades difíciles de automatizar. Estas habilidades se agrupan en tres categorías: Percepción y Manipulación (“destreza de dedos”, “destreza de manos”, “espacio de trabajo estrecho o posiciones incómodas”), Inteligencia Creativa (“originalidad” y “artes”) e Inteligencia Social (“percepción social”, “negociación”, “persuasión” y “asistencia y cuidado de personas”). En base a esto, se construyó una base de datos que consistía en las variables asociadas a los cuellos de botella junto con las categorías de automatización definidas por los expertos. El segundo paso consistió en el uso de un modelo predictivo de inteligencia artificial para categorizar al resto de las ocupaciones según su probabilidad de automatización, utilizando la base creada

³ U.S. Department of Labor, Employment and Training Administration (2016). Base de datos O*NET 15.0. Disponible en https://www.onetcenter.org/db_releases.html

para elegir el algoritmo que replicase con un menor nivel de error la categorización manual hecha por los expertos. Así, dicho algoritmo permitió identificar la probabilidad de automatización de las 632 ocupaciones restantes.

Frey y Osborne (2017) identificaron que el 47% del total de empleos de Estados Unidos se encontraban en alto riesgo de automatización al momento del estudio, considerando de alto riesgo a aquellos empleos con una probabilidad de automatización superior al 70%. La misma metodología fue ocupada para realizar similares estimaciones en distintos países. El trabajo de Pajarinen y Rouvinen (2014) encuentra que en Finlandia el 35% de los empleos se encuentran en alto riesgo de automatización; Bowles (2014) estima que dicho porcentaje en Europa varía entre 45% y más del 60%, mientras que Brzeski y Burk (2015) señalan que en Alemania es de un 59%.

Si bien el trabajo de Frey y Osborne fue una novedosa contribución a la literatura, posteriores estudios señalan que su enfoque basado en ocupaciones puede sobreestimar el riesgo de automatización, al no considerar la heterogeneidad de tareas que existe dentro de la misma ocupación (Arntz, Gregory y Zierahn, 2016). Tal como lo señala Autor (2014, 2015), las ocupaciones se componen de un conjunto de tareas, en que no todas son fácilmente automatizables.

Dadas las críticas, el estudio de Arntz, Gregory y Zierahn (2016) estima la susceptibilidad de automatización del empleo en 21 países OCDE utilizando una metodología que intenta capturar la heterogeneidad de características y tareas de cada ocupación. A tal fin, la variable de probabilidad de automatización estimada por Frey y Osborne es instrumentada con variables de la Evaluación Internacional de las Competencias para Adultos (PIAAC, por sus siglas en inglés, “*Programme for the International Assessment of Adult Competencies*”). Esta evaluación contiene datos a nivel individual sobre competencias cognitivas, características socioeconómicas, habilidades, tareas laborales, entre otras variables, de una amplia muestra de individuos. De esta

manera, se puede obtener una medición de la heterogeneidad de tareas y habilidades que cada persona usa en su puesto de trabajo y, por lo tanto, ser un mejor reflejo de la probabilidad de automatización en distintos países. Por esta razón, se considera que el enfoque de Arntz, Gregory y Zierahn está basado en tareas, mientras que el de Frey y Osborne está basado en ocupaciones.

Los resultados de Arntz, Gregory y Zierahn muestran que, en promedio, el 9% de los empleos de los 21 países OCDE son automatizables. Dicha cifra, notablemente inferior al 47% de Frey y Osborne, confirma la hipótesis de que un enfoque basado en ocupaciones sobreestima la susceptibilidad de automatización ya que la tecnología no puede sustituir la totalidad de tareas que un puesto de trabajo abarca.

Nedelkoska y Quintini (2018) replican la metodología de Frey y Osborne, pero usando el enfoque centrado en tareas de Arntz, Gregory y Zierahn para expandir el cálculo del riesgo de automatización a una muestra de 32 países OCDE. En primer lugar, fue necesario recodificar las 70 ocupaciones categorizadas por Frey y Osborne de acuerdo al sistema de cuatro dígitos de la Clasificación Internacional Uniforme de Ocupaciones 2008 (CIUO-08). En segundo lugar, se identificaron en la base de datos de la PIAAC aquellas variables que fueran proxies de los tres cuellos de botella tecnológicos (recordando, “Percepción y Manipulación”, “Inteligencia Creativa” e “Inteligencia Social”). Las autoras ocuparon los datos de Canadá como base de datos de entrenamiento para que el algoritmo pudiese predecir la probabilidad de automatización de las ocupaciones en todos los países de la muestra, esto, dado que Canadá cuenta con una muestra amplia en dicha encuesta y las ocupaciones están identificadas a un nivel de cuatro dígitos. Las autoras no consideran esto una desventaja, puesto que postulan que no hay razones para afirmar que la relación entre los cuellos de botella tecnológicos y la susceptibilidad de automatización varía entre países. De hecho consideran que la metodología empleada es un avance respecto del resto de la literatura, puesto que los estudios que ocuparon la encuesta PIAAC lo hicieron

ocupando la clasificación CIUO-08 de 2 dígitos, conduciendo a una identificación menos precisa de las 70 ocupaciones de Frey y Osborne. Asimismo, mientras Frey y Osborne ocupan 70 observaciones para entrenar el algoritmo, el trabajo de Nedelkoska y Quintini ocupa 4.656 observaciones a nivel individual.

En línea con Arntz, Gregory y Zierahn, los resultados de Nedelkoska y Quintini son más acotados que los de Frey y Osborne. En particular, Nedelkoska y Quintini encuentran que el porcentaje promedio de empleos en alto riesgo de automatización es de 14%, existiendo un gran nivel de variabilidad dentro de la muestra. Por ejemplo, en Noruega el 6% de los empleos presenta una probabilidad de automatización superior al 70%, mientras que Eslovaquia dicho porcentaje asciende a 33%.

Otro trabajo destacado que estima el potencial de automatización del empleo según su composición de actividades es el de Manyika et al (2017). Usando datos de O*NET, los autores desagregaron las ocupaciones de Estados Unidos en más de 2.000 actividades. Al mismo tiempo, con la asistencia de expertos y el uso de algoritmos de *machine learning*, se determinaron las competencias y el nivel de rendimiento necesario en cada una de ellas para desempeñar las actividades laborales de manera eficiente⁴. Luego, se analizó el rendimiento que puede tener la tecnología existente para determinar el potencial de automatización. En particular, se estudió si cada competencia puede ser ejecutada por la tecnología a un nivel de rendimiento inferior, igual o por sobre la mediana humana.

Una vez determinado el potencial de automatización de cada actividad, se estimó el tiempo que cada sector de la economía de Estados Unidos ocupa en dichas actividades. Por ejemplo, se encontró de esta manera que los sectores “Servicios de alojamiento y alimentación” e “Industria Manufacturera” tienen un alto potencial de automatización (de 73% y 60%, respectivamente) debido a que destinan mucho tiempo a actividades que la

⁴ Las capacidades son un total de 18 y se agrupan en 5 categorías distintas: “Percepción sensorial”, “Capacidades cognitivas”, “Procesamiento de lenguaje natural”, “Capacidades sociales y emocionales” y “Capacidades físicas”.

tecnología puede llevar a cabo con alto rendimiento. Este análisis permitió encontrar que el potencial de automatización promedio de los empleos de Estados Unidos asciende al 46%. Asimismo, dicho análisis se proyectó a 46 países que representan el 80% de la fuerza laboral mundial y los resultados indicaron que el empleo en los mismos tiene un potencial de automatización cercano al 50%.

I.2 Evidencia en la literatura para el mercado laboral chileno

En el caso de evidencia del riesgo de automatización para el mercado laboral chileno, el estudio de Nedelkoska y Quintini (2018) reporta que la probabilidad promedio⁵ de automatización para Chile es de 52%, la cual está por encima de la OCDE (que asciende a 47%), ubicándose dentro de los países con mayor probabilidad promedio de automatización. De acuerdo a las autoras, las diferencias que se encuentran en la variación de la probabilidad entre países se explican, en gran medida, por la forma en que se organiza el trabajo al interior de cada industria y por las diferencias estructurales de cada economía.

El trabajo de Manyika et al. (2017) estima el potencial de automatización de cada país, es decir, cuantas personas están ocupadas en empleos con alta probabilidad de ser sustituidos por máquinas. Para el caso de Chile, dicho potencial es del 49%, lo que equivale a 3,2 millones de personas. A nivel de ramas industriales, los sectores con mayor susceptibilidad de automatización son “Industria Manufacturera”, “Servicios de alojamiento y alimentación”, “Agricultura, silvicultura, pesca y caza” y “Minería”. De acuerdo a las estimaciones de los autores, tales sectores concentran aproximadamente 1,87

⁵ Es importante no confundir la probabilidad promedio de automatización con el porcentaje promedio de empleos en alto riesgo de automatización. El primero se refiere al promedio de las probabilidades de automatización de cada empleo, mientras que el segundo se refiere a la cantidad de empleos con probabilidad de automatización superior al 70%.

millones de empleos, de los cuales 1,12 millones son empleos potencialmente automatizables.

Otro trabajo relevante es el de la Fundación Chile (2017) que, aunque no usa algoritmos de *machine learning*, estima un “Índice de Intensidad de Rutina” (RII por sus siglas en inglés) para las ocupaciones en Chile, basándose en la metodología de Marcolin et al (2016a, 2016b) y datos de la encuesta PIAAC. La razón detrás de esto es que, partiendo del seminal trabajo de Autor et al. (2003), parte importante de la literatura afirma que los empleos más susceptibles a la automatización son aquellos que concentran una alta proporción de tareas rutinarias. A esta hipótesis se le conoce como Cambio Tecnológico Sesgado en Contra de la Rutina (*Routine Biased Technological Change*, RBTC), la cual al mismo tiempo considera que la tecnología es complementaria del trabajo no rutinario.

El cálculo del RII se realiza con base en variables de la PIAAC que miden qué tan a menudo el trabajador puede modificar la secuencia de las tareas que realiza; el grado de flexibilidad con el que puede seleccionar o cambiar dichas tareas; la facilidad con la que las puede planificarlas; y la posibilidad de organizar sus propios tiempos en el trabajo. El índice, por lo tanto, es el promedio ponderado de cuatro variables: secuencia, flexibilidad, planificación y organización. Su rango de valores es de uno a cinco, en donde uno representa el menor nivel de intensidad de rutina, mientras que cinco el mayor (Fundación Chile, 2017).

Los resultados de Fundación Chile (2017) indican que, entre otros hallazgos, hacia el primer trimestre móvil del año 2017, un total de 1.935.536 trabajadores se encontraban en ocupaciones pertenecientes al cuartil de intensidad de rutina más alto. Esto implica, que un 24% de los trabajadores son susceptibles a la automatización, en comparación al 28% de la OCDE al usar la esta metodología. Si se consideran los trabajadores ubicados en el tercer y en el cuarto cuartil de intensidad de rutina (niveles de rutina medio y alto, respectivamente), estos constituyen un 61% de trabajadores en empleos automatizables

(casi 5 millones en total), tres puntos porcentuales por sobre el promedio de países OCDE. Distinguiendo por sectores, el trabajo de Fundación Chile está alineado con los hallazgos de Manyika et al. (2017) debido a que encuentran que “Minería” y “Agricultura, silvicultura y pesca” están entre los sectores con un mayor potencial de automatización (o mayor índice de intensidad de rutina).

II Metodología del estudio para Chile

El análisis del riesgo de automatización se llevará a cabo en dos fases en las que aplicaremos técnicas de clasificación basadas en *machine learning*, es decir, calcularemos la probabilidad de que una ocupación sea automatizable (o no) utilizando un modelo predictivo de aprendizaje supervisado. En la primera fase, imputaremos a los individuos de la encuesta CASEN 2017 la frecuencia de uso de habilidades no automatizables recogida en la encuesta PIAAC para Chile, para lo cual utilizamos algoritmos de clasificación binaria que permiten predecir qué tan frecuentemente un trabajador utiliza una habilidad específica de acuerdo a las características de su trabajo. En la segunda fase, empleamos las frecuencias predichas en la fase anterior para estimar la probabilidad de automatización de las ocupaciones en el mercado laboral chileno.

Considerando lo anterior, esta sección contiene un marco conceptual relativo a las técnicas que serán utilizadas para el análisis, las bases de datos necesarias, la aplicación de los modelos predictivos para la clasificación del uso de habilidades en la encuesta CASEN, y la obtención de la probabilidad de automatización de las ocupaciones en el mercado laboral chileno.

II.1 ¿Qué entendemos por *machine learning*?

El concepto de *machine learning* se refiere a una forma de inteligencia artificial que permite a un sistema o modelo “aprender” desde los datos, en vez de ser programado de manera explícita. Muchos de los sistemas que se ocupan hoy en día utilizan éstas técnicas como herramientas de análisis y predicción, tanto en los negocios, como en la formulación de políticas públicas o la academia. Un ejemplo cotidiano es el de las interacciones dentro de los sitios de comercio electrónico, como Amazon. Al navegar en este sitio, revisando productos y leyendo sus calificaciones, la página comienza a recomendarnos nuevos productos que podrían ser de nuestro interés, sin necesariamente estar directamente relacionados con la búsqueda que estábamos haciendo. Las sugerencias se originan a partir de un modelo de *machine learning* que recoge la información de nuestro historial de búsquedas, junto con la de otros compradores, así como con la información de compra, para presentar productos de mayor interés, en base a lo que aprende del comportamiento de personas con hábitos de consumo similares.

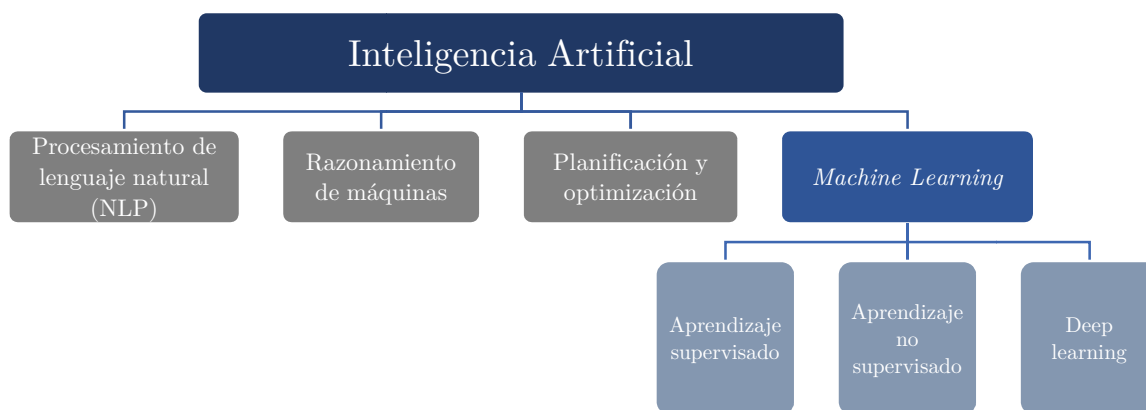


Figura 1.1 – Conjunto de tecnologías de la inteligencia artificial.

Fuente: Elaboración propia.

Para mejorar la comprensión del concepto de *machine learning*, se debe entender bajo el espectro más amplio de tecnologías de Inteligencia Artificial. Comúnmente se utilizan los conceptos de inteligencia artificial, *machine learning*, o *deep learning* de forma intercambiable, sin embargo, existe una diferencia jerárquica entre estas tecnologías. Como se presenta en la Figura 1.1, la Inteligencia Artificial corresponde al espectro más amplio de tecnologías que son capaces de interactuar de manera independiente con su entorno, es decir, capaces de “pensar”. Dentro de este espectro podemos encontrar tecnologías como el procesamiento de lenguaje natural (NPL, por sus siglas en inglés, *Natural Language Processing*), que busca dotar a los computadores con la capacidad de entender el discurso escrito y oral; el razonamiento de máquinas, que permite a los sistemas ejecutar razonamientos lógicos para realizar inferencia sobre datos; o la planificación automática de procesos, en que los sistemas son capaces de manera autónoma de definir la secuencia de procesos para llegar un objetivo final. En el marco de este trabajo, nos enfocaremos únicamente en *machine learning*, el que utiliza una variedad de algoritmos que reconocen la estructura de los datos para mejorar, describir información y realizar predicciones de distintos escenarios. Estos algoritmos son entrenados con base en un conjunto de datos para determinar las relaciones existentes entre ellos y, posteriormente, poder entregar un resultado que depende de nuevos grupos de datos no utilizados en el entrenamiento del algoritmo. Usualmente, estos algoritmos pueden usarse con la finalidad de realizar análisis descriptivos, en que los resultados entregados permiten entender la realidad del contexto en que se encuentra, o con un fin predictivo, en que los algoritmos permiten pronosticar cambios en los datos, según los patrones que logren identificar en los datos con los que se entrenen.

Los algoritmos de *machine learning* pueden clasificarse como supervisados, cuando se le entrega al algoritmo un set de datos definido así como la comprensión de cómo

clasificar los datos; o como no supervisados⁶, en que el set de datos no tiene relaciones definidas, por lo que el algoritmo construye las clasificaciones dependiendo de los grupos (o clústeres) de características que sea capaz de identificar.

Tomemos la clasificación de objetos como un ejemplo aprendizaje supervisado: en la recolección de basura, existen robots capaces de identificar de entre la basura recogida, aquellos objetos que son reciclables. El robot es entrenado con información sobre las características de los objetos, así como con una etiqueta de clasificación (en este caso, si es reciclable o no reciclable), y sobre esta base, el robot es capaz de discriminar los objetos reciclables dentro de la basura.

En cambio, podemos considerar el caso de las redes sociales (Facebook, Twitter o Instagram) como ejemplos de aprendizaje no supervisado: los algoritmos de estas redes se alimentan de las interacciones de los usuarios en la plataforma, para identificar patrones de comportamiento e intereses que permitan agruparlos en clústeres y, por ejemplo, poder realizar recomendaciones de compra o focalizar publicidad. La diferencia entre ambos tipos de entrenamiento radica en que, en el aprendizaje no supervisado, la estructura de los datos no ha sido completamente entendida ni categorizada, mientras que en el aprendizaje supervisado sí.

Las ventajas del uso de algoritmos de *machine learning* radican en (i) una mayor precisión por sobre heurísticas humanas, dado que se basan en la estructura propia de los datos y las relaciones entre ellos, (ii) son capaces de identificar más eficientemente tendencias y patrones en grandes volúmenes de datos, (iii) permite una adaptación instantánea de los sistemas a medida que surge nueva información, sin necesidad de intervención humana, lo que permite contar con predicción actualizadas en tiempo real.

⁶ Por simplicidad, entendida en el marco de este trabajo, no se profundizará en el *deep learning*, sin embargo, esta corresponde a un método específico de *machine learning* que se utiliza para la detección de patrones entre los datos a partir de datos no estructurados. Esta tecnología hace uso de redes neuronales para aprender de los datos de manera iterativa.

Sin embargo, los algoritmos no están exentos de problemas: (i) existe una alta susceptibilidad a los errores (por ejemplo, de especificación del modelo, de ruido en los datos de entrenamiento, muestras desbalanceadas, entre otros), lo que se traduce en predicciones incorrectas; (ii) el manejo de grandes volúmenes de datos es altamente intensivo en recursos computacionales, por lo que la rapidez en la obtención de los resultados dependerá de la capacidad computacional instalada; y (iii) un algoritmo será tan bueno como la base de datos con la que se entrene, por lo que las predicciones realizadas dependerán fundamentalmente de la calidad y cantidad de observaciones que se dispongan para el entrenamiento.

II.2 Modelos predictivos de aprendizaje supervisado

Considerando que el objeto de este estudio es llevar a cabo un análisis predictivo para una clasificación binaria⁷ utilizando algoritmos de aprendizaje supervisado, es que pondremos el foco en la descripción de este tipo de técnicas. En la Figura 1.2 se presenta el esquema general de un proceso de entrenamiento y predicción para un modelo de aprendizaje supervisado. El proceso general se define en dos fases: una primera fase de entrenamiento del algoritmo y una segunda fase de predicción. Para la construcción del modelo predictivo, es necesario recolectar datos relevantes que podrían explicar la variable objetivo que se busca predecir. En el aprendizaje supervisado, estos datos deben corresponder a pares ordenados compuestos por un conjunto de variables características (que identifican la observación en la muestra) y la variable objetivo asociada (es decir, la categoría de la observación). Una vez obtenidos estos datos, se pre-procesan antes de utilizarlos para el entrenamiento del algoritmo—por ejemplo, codificar variables categóricas, tomar decisiones respecto a valores faltantes y outliers, o reescalar variables continuas—con tal de poder minimizar el uso de recursos computacionales.

⁷ Esto es, clasificar los datos de la muestra dentro de dos categorías.

Una vez que se cuenta con la base de datos procesada, ésta se subdivide en tres bases de datos: una base con la que se entrenará el modelo (set de entrenamiento), una base para la calibración de los parámetros del modelo (set de validación) y una base que se utiliza para evaluar el desempeño del modelo entrenado (set de testeo). Las observaciones en la base de entrenamiento sirven como ejemplos para calibrar los parámetros del algoritmo, mientras que las del modelo de validación son ocupadas durante el proceso de entrenamiento para evaluar el modelo seleccionado y optimizar su configuración, con tal de mejorar su desempeño. Una vez entrenado el algoritmo, se utiliza la base de testeo, esto es, observaciones que el modelo no utilizó para entrenar, para evaluar el desempeño del modelo al comparar el verdadero valor de la variable objetivo con los valores predichos por el algoritmo.

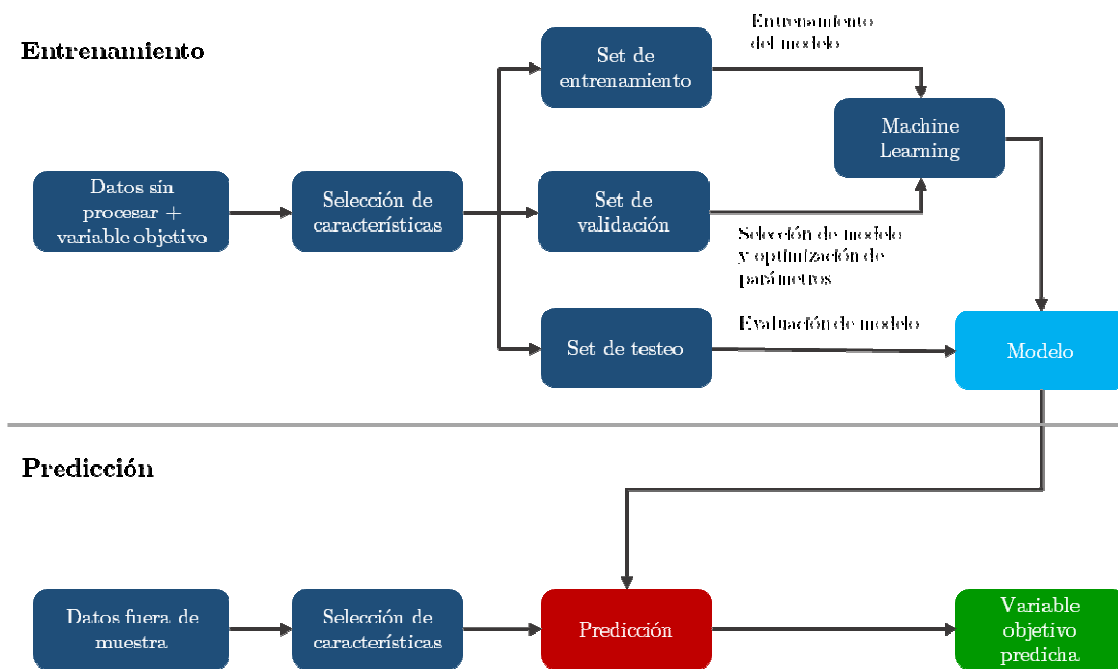


Figura 1.2 – Diagrama de un proceso general para algoritmos predictivos.
 Fuente: Jouganous, Savidan y Bellec (2018)⁸.

⁸ Jouganous, Savidan y Bellec (2018). “A Brief Overview of Automatic Machine Learning Solutions”. Recuperado de: <https://techblog.cdiscout.com/a-brief-overview-of-automatic-machine-learning-solutions-automl/>

Habiendo entrenado y optimizado el algoritmo para obtener el desempeño óptimo, condicional a los datos de entrenamiento usados, se procede a la segunda fase de predicción. Esta fase tiene como objetivo poder asignar una clasificación a observaciones para las cuales conocemos sus variables características, pero desconocemos el verdadero valor de la variable objetivo. Las variables características de estas observaciones deben ser procesadas de la misma manera en que se procesaron las variables del conjunto de entrenamiento, para que puedan ser utilizadas de manera estandarizada por el algoritmo para la predicción. Las observaciones ya procesadas se entregan al modelo para poder obtener una predicción de las variables objetivo asociadas a cada una ellas.

El algoritmo entrenado, considerando el valor de las variables características de las nuevas observaciones, asignará una probabilidad asociada a cada uno de los valores que podría tomar la variable objetivo. Para determinar la categoría asignada a la variable, se define un umbral de tal forma que, si la probabilidad supera el umbral, se asigna dicha categoría a la variable objetivo, mientras que, si no lo supera, se imputa el valor complementario.

Para ejemplificar, tomemos el caso de robots capaces de diferenciar aquellos residuos reciclables de los que no lo son, dentro de la basura⁹. El algoritmo de *machine learning* de aquellos robots está entrenado de tal forma—por ejemplo, chequeando la forma y el material de los objetos, o logos e impresiones en el empaquetado—que asigna una probabilidad asociada a que el objeto sea reciclable y otra a que el objeto sea no reciclable. Si la probabilidad de que sea reciclable es mayor que un umbral definido, el robot reconoce al objeto como un reciclable y es separado del resto de la basura, mientras que, en caso contrario, se reconoce como un elemento no reciclable y continúa a otros procesos de manejo de desechos.

⁹ Jennifer Kite-Powell (2017). “This Recycling Robot Uses Artificial Intelligence To Sort Your Recyclables”. Recuperado de: <https://www.forbes.com/sites/jenniferhicks/2017/04/04/this-recycling-robot-uses-artificial-intelligence-to-sort-your-recyclables/#b7cd3ca2d353>

Si bien el proceso general presentado en esta sección muestra una versión simplificada del funcionamiento de los modelos de *machine learning* predictivos, refleja que estas técnicas están basadas en la recolección y procesamiento de ejemplos que puedan ser utilizados para que los algoritmos se basan en ellos para predecir posibles resultados. Los modelos pueden ser complejizados con tal de mejorar la precisión de las predicciones realizadas, ya sea probando distintos algoritmos, eligiendo distintas métricas para evaluar el desempeño de éstos, o usando diferentes técnicas para pre-procesar los datos, entregando una mejor información al algoritmo sobre las relaciones existentes entre las variables.

II.3 Bases de datos

Para analizar la probabilidad de automatización de las ocupaciones es necesario comprender la caracterización individual de los ocupados, en relación al uso de habilidades dentro del entorno laboral. Durante 2014 y 2015, la Organización para la Cooperación y el Desarrollo Económico (OCDE) implementó una medición para Chile, entre otros países miembros, llamada Evaluación Internacional de las Competencias para Adultos (PIAAC). En el marco de este estudio, se midió el desempeño de los trabajadores en áreas de matemáticas, comprensión de lectura y resolución de problemas, así como también se recolectó información y datos sobre cómo se utilizan las habilidades en un ambiente doméstico y laboral. Para Chile, esta base de datos cuenta con un total de 5.212 observaciones de trabajadores chilenos, de los cuales 3.620 son ocupados, 238 desempleados y 1.332 están fuera del mercado laboral.

Las observaciones que consideraremos serán las de los ocupados, de ambos sexos, entre 16 y 65 años, que tengan respuestas completas a las preguntas de las habilidades no automatizables. Se elige sólo el grupo de ocupados para poder caracterizar la intensidad en el uso de las habilidades según las ocupaciones de cada uno de los individuos en la base. La información principal que será utilizada de esta base de datos es la de la frecuencia de uso

de las habilidades, separándolas en habilidades automatizables y no automatizables, en el marco de los cuellos de botella de ingeniería identificados por Frey y Osborne (2017). Las habilidades no automatizables son aquellas relacionadas con inteligencia creativa, la percepción de espacios físicos y manipulación de objetos, así como inteligencia social, las que corresponden a habilidades que presentan importantes desafíos de ingeniería para su implementación en la tecnología (Tabla 1.1). Cada una se mide en una escala de frecuencia de uso (de 1 a 5), donde 1 corresponde a una habilidad que no se usa nunca, 2 una que se usa menos de una vez al mes, 3 corresponde a una que se usa menos de una vez a la semana pero más de una vez al mes, 4 es una que se usa al menos una vez a la semana pero no todos los días y 5 una que se usa siempre.

Dado que la escala de frecuencia de uso es relativa y no es posible identificar la composición exacta de horas que se dedica a cada una de las habilidades mencionadas, se clasificarán en dos categorías dependiendo de si es una habilidad de uso habitual en el lugar de trabajo (1 si la respuesta PIAAC es 4 o 5), o de uso no habitual (0 si la respuesta PIAAC es 1, 2 o 3).

Tabla 1.1 – Habilidades no automatizables en PIAAC

Cuello de botella tecnológico	Código	Habilidad
Inteligencia creativa	F_Q05a	Resolución de problemas simples
	F_Q05b	Resolución de problemas complejos
	G_Q03h	Uso de matemáticas y estadísticas avanzadas
	G_Q05g	Uso de lenguajes de programación
Manipulación de objetos	F_Q06c	Uso de manos o dedos
Inteligencia social	F_Q01b	Cooperación con otros trabajadores
	F_Q02a	Compartir información relacionada al trabajo
	F_Q02b	Enseñar
	F_Q02d	Vender
	F_Q02e	Aconsejar personas
	F_Q03b	Planificar actividades de otros
	F_Q04a	Influenciar
	F_Q04b	Negociar con personas
	G_Q05h	Discusión en tiempo real en internet

Fuente: Evaluación Internacional de las Competencias para Adultos (PIAAC), 2015.

Esta base de datos contiene información suficiente para caracterizar el perfil de uso de habilidades de los trabajadores, sin embargo, la cantidad de observaciones es insuficiente para realizar una caracterización socioeconómica completa de los riesgos de automatización para el mercado laboral chileno. Por esto, utilizaremos la base de datos de CASEN (Encuesta de Caracterización Socioeconómica Nacional, 2017), dado que contiene una mayor cantidad de observaciones y una mayor riqueza de variables socioeconómicas que no están disponibles en PIAAC. En la muestra de la encuesta CASEN hay 174.338 individuos¹⁰, de los cuales 91.679 están ocupados, 7.803 están desocupados y 74.856 se encuentran inactivos. Dada la estructura de la encuesta CASEN, se cuenta con una representación de las ocupaciones con un mayor grado de granularidad que en la PIAAC, pudiendo identificar la ocupación de los trabajadores con los códigos de ocupaciones de cuatro dígitos (CIUO-88).

Considerando que la CASEN no contiene información sobre el uso de habilidades en el trabajo, uno de los desafíos del presente análisis, previo a la estimación de la probabilidad de automatización, es poder predecir cuál sería la intensidad de uso de habilidades correspondiente a los individuos en la CASEN, en base a las respuestas dadas por los trabajadores en la PIAAC. Aplicaremos un modelo predictivo para generar el valor de estas variables a partir de los resultados de la encuesta PIAAC, suponiendo que las muestras de la CASEN y de la PIAAC provienen de un mismo mercado laboral¹¹, donde el supuesto subyacente es que individuos con características similares (edad, sexo, grupo ocupacional, nivel de estudios, entre otras) deberían responder de manera similar sobre su uso de habilidades.

¹⁰ Se excluyó de la muestra a los individuos pertenecientes a las fuerzas armadas (grupo ocupacional 0), así como también a aquellas observaciones sin dato para el grupo ocupacional.

¹¹ Entendido como muestras provenientes de una misma distribución.

Finalmente, utilizaremos como insumo la categorización de tareas realizada por Frey y Osborne (2017), donde las categoría asignadas a una ocupación corresponden a “Automatizable” (1) o “No automatizable” (0), las cuales están disponibles en el Anexo A.

II.4 Fase 1: Predicción de las habilidades en CASEN

Como se mencionaba, la primera fase consta de todos los procedimientos para poder identificar la intensidad de uso de habilidades de los individuos presentes en la muestra de CASEN 2017. Para ello, la Figura 1.3 esquematiza el flujo de los procesos llevados a cabo para la predicción de la intensidad de uso de habilidades en el trabajo, los que serán precisados de mejor manera en las subsecciones siguientes.

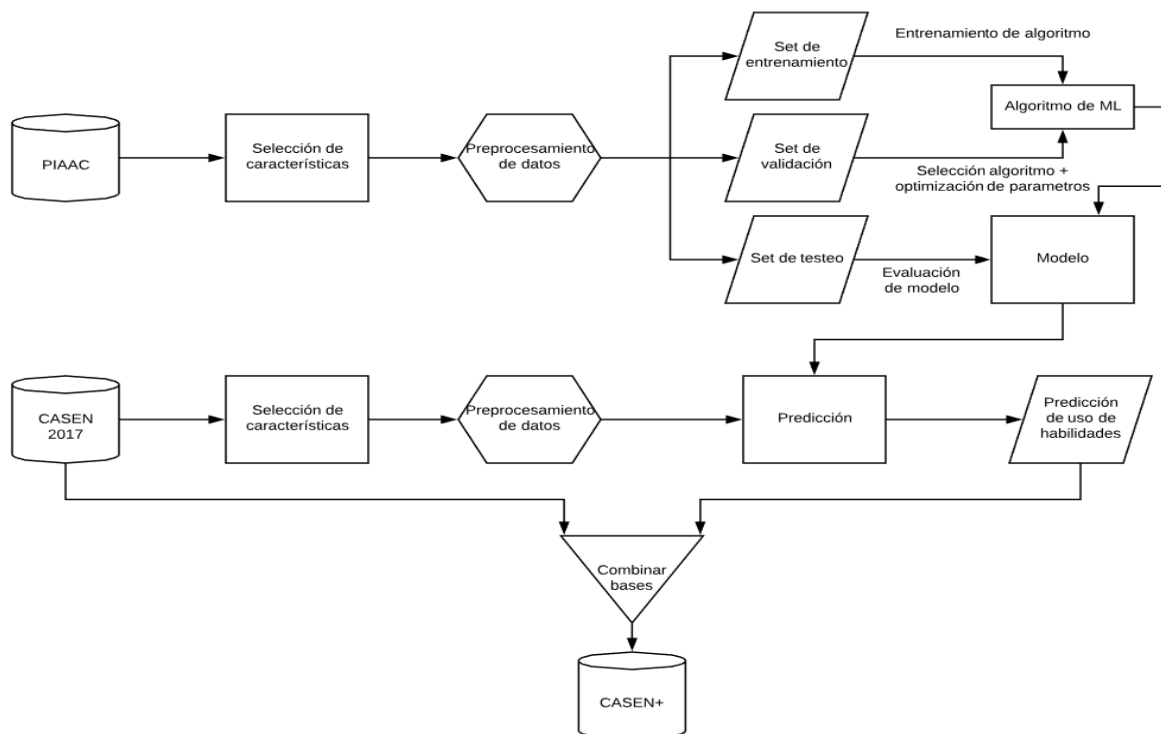


Figura 1.3 – Diagrama de flujo para modelo predictivo de uso de habilidades.

Nota: CASEN+ corresponde a la base de datos CASEN 2017 que incluye la imputación de habilidades en el ambiente laboral.

Fuente: Elaboración propia.

II.4.1. Procesamiento de las bases de datos

De la encuesta PIAAC seleccionamos un conjunto de variables explicativas que sean capaces de recoger la información respecto al uso de habilidades en el trabajo de cada individuo en la muestra. En este caso, el set de entrenamiento corresponderá a la muestra de individuos PIAAC, en que tomaremos como variables características las siguientes: edad, sexo, inmigrante, nivel educacional, nivel de calificación y grupo ocupacional. Las categorías corresponden a las anteriormente descritas en la Tabla 1.1. La base de datos para la predicción corresponderá a la muestra de la CASEN, en que seleccionaremos las mismas variables características elegidas en PIAAC. De estas bases, sólo se seleccionan a individuos ocupados (con grupo ocupacional disponible a cuatro dígitos), que no pertenezcan a las fuerzas armadas (grupo ocupacional 0) y cuyas edades fluctúen entre los 16 y los 65 años.

Para estandarizar las bases se realizan los siguientes procedimientos:

- La variable edad corresponde a una variable entera que varía entre 16 y 65 años para ambas bases. Esta variable se normaliza, llevándola a una escala unitaria (esto es, entre 0 y 1). Para esto se realiza la siguiente transformación:

$$\text{Edad}_{\text{norm}} = \frac{\text{Edad} - 16}{65 - 16}$$

- La variable sexo corresponde a una variable binaria que toma el valor 1 si el individuo es hombre, mientras que toma el valor 0 si es mujer.
- La variable inmigrante corresponde a una variable binaria que toma el valor 1 si el individuo es chileno, mientras que toma el valor 0 en otro caso.
- Para el nivel educacional, primero se agrupa el mayor nivel educacional obtenido por el individuo en tres categorías. La categoría “baja” considera a todos los niveles educacionales menores a la educación primaria incompleta; la categoría media es a los niveles comprendidos entre la educación primaria completa y la educación terciaria de ciclo corto; y finalmente, la categoría “alta” corresponde a los niveles de pregrado y postgrado.

Posteriormente, a partir de esta variable se generan tres variables binarias asociada a cada una de las categorías, por ejemplo, si un individuo posee un nivel educacional alto, la variable binaria “Niv. Educacional alto” debería tomar el valor 1, mientras que las otras deberían tomar el valor 0.¹²

- La variable nivel de calificación se refiere a la agrupación de los trabajadores según su grupo ocupacional. La categoría “alta calificación” comprende a trabajadores en los grupos ocupacionales 1, 2 y 3; la categoría “mediana calificación” a los trabajadores en los grupos del 4 al 8, y la de “baja calificación” a los trabajadores en el grupo 9. Nuevamente, creamos variables binarias para cada una de las categorías del mismo modo que hicimos para el nivel educacional.

- Por último, creamos variables binarias para cada uno de los grupos ocupacionales entre 1 y 9. Estas variables indican con un 1 si el individuo posee una ocupación perteneciente a dicho grupo ocupacional, mientras que toma el valor 0 si no. Tomamos la variable grupo ocupacional de un dígito debido a que PIAAC y CASEN utilizan distintos esquemas de clasificación de ocupaciones (PIAAC usa el CIUO-08, mientras que CASEN el CIUO-88). A nivel de grupo ocupacional de 4 dígitos se generan diferencias entre ambos estándares por la reclasificación de algunas en otras ocupaciones, por lo que la decisión de tomar un dígito se basa en que a nivel agregado, la naturaleza de los grupos no debería presentar cambios entre ambas metodologías.

Esta estandarización se realiza tanto para las variables presentes en la PIAAC, como en la CASEN. En la Tabla 1.2 se presentan estadísticas descriptivas para ambos sets de datos estandarizados.

¹² Este procedimiento recibe el nombre de “One Hot Encoding”, y permite a los algoritmos entender mejor la estructura de los datos en presencia de variables categóricas que no necesariamente tienen un ordenamiento natural.

Tabla 1.2 – Estadísticas descriptivas de las variables características seleccionadas

Variable	Tipo de variable	Valores	Promedio PIAAC	Promedio CASEN
Edad normalizada	Continua	[0,1]	0.46	0.52
Sexo (hombre=1)	Binaria	{0,1}	0.51	0.57
Inmigrante (=1)	Binaria	{0,1}	0.03	0.05
Nivel educacional bajo	Binaria	{0,1}	0.22	0.20
Nivel educacional medio	Binaria	{0,1}	0.63	0.64
Nivel educacional alto	Binaria	{0,1}	0.15	0.16
Alta calificación	Binaria	{0,1}	0.31	0.26
Mediana calificación	Binaria	{0,1}	0.55	0.51
Baja calificación	Binaria	{0,1}	0.14	0.23
Grupo ocupacional 1	Binaria	{0,1}	0.05	0.04
Grupo ocupacional 2	Binaria	{0,1}	0.13	0.12
Grupo ocupacional 3	Binaria	{0,1}	0.13	0.10
Grupo ocupacional 4	Binaria	{0,1}	0.11	0.07
Grupo ocupacional 5	Binaria	{0,1}	0.23	0.16
Grupo ocupacional 6	Binaria	{0,1}	0.03	0.05
Grupo ocupacional 7	Binaria	{0,1}	0.11	0.14
Grupo ocupacional 8	Binaria	{0,1}	0.07	0.09
Grupo ocupacional 9	Binaria	{0,1}	0.14	0.23

Fuente: Elaboración propia, con base a PIAAC (2015) y CASEN (2017).

Este conjunto de entrenamiento será utilizado para “entrenar” al algoritmo, es decir, otorgarle el algoritmo la información necesaria para capturar las relaciones subyacentes entre las variables características y la respuesta sobre cada una de las habilidades consideradas. En la PIAAC, las respuestas para cada pregunta sobre frecuencia de uso de habilidades se asignan a cinco categorías: (1) Nunca se usa, (2) Menos de una vez al mes, (3) Más de una vez al mes, pero menos de una vez a la semana; (4) Más de una vez a la semana, pero no todos los días y (5) Todos los días. Basándonos en estas respuestas, agrupamos las variables en dos categorías para transformar el problema en uno de clasificación binario: “no habitual” (0)/“habitual” (1). La categoría “no habitual” engloba a las frecuencias de uso 1, 2 y 3, considerando que son habilidades que se pueden utilizar en el ambiente laboral, pero no de una manera habitual como para considerarla

parte fundamental de la ocupación. La categoría “habitual” considera aquellas habilidades que se utilizan más frecuentemente y que, por lo tanto, forman parte integral de la tareas que se realizan en la ocupación.

En cuanto a las variables objetivo, las estadísticas descriptivas se presentan en la Tabla 1.3:

Tabla 1.3 – Estadísticas descriptivas de la intensidad de uso de habilidades en PIAAC

	Código	Habilidad	Promedio PIAAC
Inteligencia Creativa	F_Q05a	Resolución de problemas simples	0.69
	F_Q05b	Resolución de problemas complejos	0.39
	G_Q03h	Uso de matemáticas y estadísticas avanzadas	0.05
	G_Q05g	Uso de lenguajes de programación	0.07
Percepción y Manipulación	F_Q06c	Uso de manos o dedos	0.82
Inteligencia Social	F_Q01b	Cooperación con otros trabajadores	0.60
	F_Q02a	Compartir información relacionada al trabajo	0.78
	F_Q02b	Enseñar	0.46
	F_Q02d	Vender	0.34
	F_Q02e	Aconsejar personas	0.46
	F_Q03b	Planificar actividades de otros	0.32
	F_Q04a	Influenciar	0.52
	F_Q04b	Negociar con personas	0.38
	G_Q05h	Discusión en tiempo real en internet	0.08

Fuente: PIAAC (2015).

Al observar la distribución del uso de las habilidades en la muestra, es posible observar que existen habilidades muy concentradas en una sola categoría, mientras que otras están balanceadas. Por ejemplo, las matemáticas y estadísticas avanzadas o los lenguajes de programación (consideradas dentro de las habilidades de inteligencia creativa), son utilizadas de manera habitual por una pequeña proporción de la muestra (5% la primera y 7% la segunda). Si los datos de entrenamiento están muy sesgados hacia una única categoría, los algoritmos tendrán problemas para clasificar correctamente aquellas observaciones menos representadas. Este problema lo manejamos mediante un sobre-muestreo de las observaciones poco representadas (He y García, 2008), esto es,

aleatoriamente repetimos observaciones pertenecientes a la categoría minoritaria al set de entrenamiento (no al de testeo), para que el algoritmo pueda recoger de manera más balanceada las características de la categoría con menor representación.

II.4.2. Entrenamiento de los algoritmos predictivos

Una vez procesadas las bases de datos, tanto de entrenamiento y testeo (PIAAC), como la de predicción (CASEN), procedemos a entrenar tres algoritmos clasificadores¹³ para cada una de las habilidades que buscamos predecir (esto es, un total de 14 habilidades).

En cada algoritmo, la elección de los parámetros que lo definen es una parte relevante en el entrenamiento, ya que distintas combinaciones de parámetros pueden llevar a que el algoritmo se desempeñe mejor o peor, con los mismos datos de entrenamiento. Por ello, la selección de los parámetros de cada uno de los algoritmos se realiza a través de un proceso llamado “búsqueda de grilla aleatoria” (*Random Grid Search*). Esta técnica consiste en seleccionar iterativamente conjuntos de parámetros de manera aleatoria (por ejemplo, el número de árboles y la profundidad de éstos en un Random Forest), entrenar dicho algoritmo con cada conjunto de parámetros y evaluar su desempeño. Este método permite determinar de una manera más eficiente la combinación de parámetros óptima que en una búsqueda de grilla tradicional, ya que una exploración exhaustiva de los parámetros puede resultar en un alto consumo de tiempo y recursos computacionales cuando no todos los parámetros son igualmente relevantes para el análisis.

Otro factor relevante para el análisis es la selección de una métrica apropiada para la evaluación del desempeño de los algoritmos, ya que es la base para seleccionar el modelo que prediga de forma más certera el uso de las habilidades en el trabajo. Hemos considerado tres métricas: la exactitud, el F1-Score, y el área bajo la curva receptor-

¹³ El detalle de los algoritmos clasificadores puede encontrarse en el Anexo B.

operador (ROC-AUC). Estas tres métricas se explican más detalladamente en el Anexo C, pero a grandes rasgos, mientras mayor sea el valor de la medida, mejor es el algoritmo.

Para la selección del mejor algoritmo, utilizamos la métrica ROC-AUC por sobre las otras dos, ya que es una medida de qué tan capaz es el modelo de distinguir entre categorías, que es lo relevante para este caso. Las otras medidas no son necesariamente representativas de un buen desempeño, particularmente cuando estamos frente a muestras desbalanceadas (como algunas de las habilidades ya comentadas), ya que no considera la relevancia de clasificar de manera errónea las categorías minoritarias.

Los resultados del entrenamiento de los tres algoritmos para cada una de las preguntas se presentan en la Tabla 1.4. Cada resultado corresponde al promedio de la medida de desempeño obtenida para el entrenamiento del algoritmo con 100 combinaciones distintas de parámetros.

Tabla 1.4 – Resultado de entrenamiento de frecuencia de habilidades, para los algoritmos seleccionados

Código	Habilidades	ROC-AUC			Algoritmo seleccionado
		LR	RF	KNN	
F_Q05a	Resolución de problemas simples	0.60	0.64	0.65	KNN
F_Q05b	Resolución de problemas complejos	0.66	0.83	0.83	KNN
G_Q03h	Uso de matemáticas y estadísticas avanzadas	0.63	0.64	0.62	RF
G_Q05g	Uso de lenguajes de programación	0.71	0.76	0.76	RF
F_Q06c	Uso de manos o dedos	0.67	0.67	0.66	RF
F_Q01b	Cooperación con otros trabajadores	0.67	0.75	0.76	KNN
F_Q02a	Compartir información relacionada al trabajo	0.69	0.68	0.67	LR
F_Q02b	Enseñar	0.69	0.73	0.73	KNN
F_Q02d	Vender	0.66	0.73	0.75	KNN
F_Q02e	Aconsejar personas	0.63	0.69	0.69	KNN
F_Q03b	Planificar actividades de otros	0.73	0.86	0.86	KNN
F_Q04a	Influenciar	0.78	0.95	0.95	RF
G_Q04	Experiencia usando computadores	0.88	0.89	0.88	RF
F_Q04b	Negociar con personas	0.77	0.94	0.93	RF
G_Q05h	Discusión en tiempo real en internet	0.74	0.93	0.93	RF

Fuente: Elaboración propia.

Considerando los resultados obtenidos, elegimos el clasificador con el mejor desempeño según el ROC-AUC promedio obtenido para cada una de las preguntas. Se puede observar que los mejores desempeños se dan en los casos no paramétricos como los métodos de vecinos más cercanos (KNN) y Random Forest. Con este conjunto de mejores clasificadores, obtenemos una exactitud promedio de las preguntas de un 73%, un valor promedio de F1 de 0.73 y un promedio ROC-AUC de 0.78.

II.4.3. Predicciones obtenidas para CASEN

Con los distintos algoritmos entrenados para cada pregunta, se obtiene la categorización de cada una de las preguntas para los individuos en la muestra CASEN. A continuación, en la Tabla 1.5 se presentan las estadísticas descriptivas de la intensidad de uso de habilidades, de acuerdo a los valores predichos por los algoritmos.

Tabla 1.5 – Estadísticas descriptivas de habilidades predichas en CASEN

Código	Habilidad	Promedio CASEN	Diferencia con prom. PIAAC
F_Q05a	Resolución de problemas simples	0.55	-0.14
F_Q05b	Resolución de problemas complejos	0.42	0.03
G_Q03h	Uso de matemáticas y estadísticas avanzadas	0.14	0.09
G_Q05g	Uso de lenguajes de programación	0.17	0.11
F_Q06c	Uso de manos o dedos	0.68	-0.14
F_Q01b	Cooperación con otros trabajadores	0.59	-0.01
F_Q02a	Compartir información relacionada al trabajo	0.66	-0.12
F_Q02b	Enseñar	0.47	0.00
F_Q02d	Vender	0.35	0.01
F_Q02e	Aconsejar personas	0.46	-0.00
F_Q03b	Planificar actividades de otros	0.41	0.09
F_Q04a	Influenciar	0.51	-0.01
F_Q04b	Negociar con personas	0.47	0.09
G_Q05h	Discusión en tiempo real en internet	0.19	0.12

Fuente: Elaboración propia, en base a PIAAC(2015) y CASEN (2017).

Se observa que existen diferencias entre la intensidad de uso de habilidades entre los resultados predichos para la muestra CASEN y los resultados de la muestra PIAAC.

Por ejemplo, el 55% de los individuos de la CASEN resuelve problemas simples de manera habitual, mientras que en la muestra PIAAC lo hace el 69%, o el 68% de individuos de la CASEN usa las manos o dedos como parte habitual de su trabajo, con un 82% en la PIAAC, lo que significaría que un menor número de individuos utiliza habilidades manuales en su trabajo de manera habitual. Por otra parte, hay diferencias positivas en habilidades como discusiones en tiempo real en internet o planificar actividades de otros, donde hay una mayor cantidad de individuos en CASEN que realizan dichas actividades habitualmente que en la muestra PIAAC.

La Figura 1.4 muestra una comparación entre el uso de habilidades entre ambas bases de datos, según grupo ocupacional, mediante el uso de mapas de calor. Un color verde más profundo significa que dicha habilidad se usa más intensamente (más cercano a 1), mientras que un verde poco saturado implica que dicho grupo utiliza esa habilidad de manera infrecuente.

	Habilidad	CASEN								
		1	2	3	4	5	6	7	8	9
F_Q05a	Resolución de problemas simples									
F_Q05b	Resolución de problemas complejos									
G_Q03h	Uso de matemáticas y estadísticas avanzadas									
G_Q05g	Uso de lenguajes de programación									
F_Q06c	Uso de manos o dedos									
F_Q01b	Cooperación con otros trabajadores									
F_Q02a	Compartir información relacionada al trabajo									
F_Q02b	Enseñar									
F_Q02d	Vender									
F_Q02e	Aconsejar personas									
F_Q03b	Planificar actividades de otros									
F_Q04a	Influenciar									
F_Q04b	Negociar con personas									
G_Q05h	Discusión en tiempo real en internet									

Figura 1.4 – Comparación de intensidad de uso de habilidades según grupo ocupacional.

Fuente: Elaboración propia en base a cálculos de los autores.

II.5 Fase 2: Predicción de la probabilidad de automatización en Chile

En esta fase, utilizando la base de datos con la imputación de la intensidad de uso de habilidades, se aplica un modelo predictivo para la probabilidad de automatización de las ocupaciones en Chile. Para ello, la Figura 1.5 esquematiza el flujo de los procesos llevados a cabo para la predicción del riesgo de automatización, los que se detallan en las subsecciones siguientes.

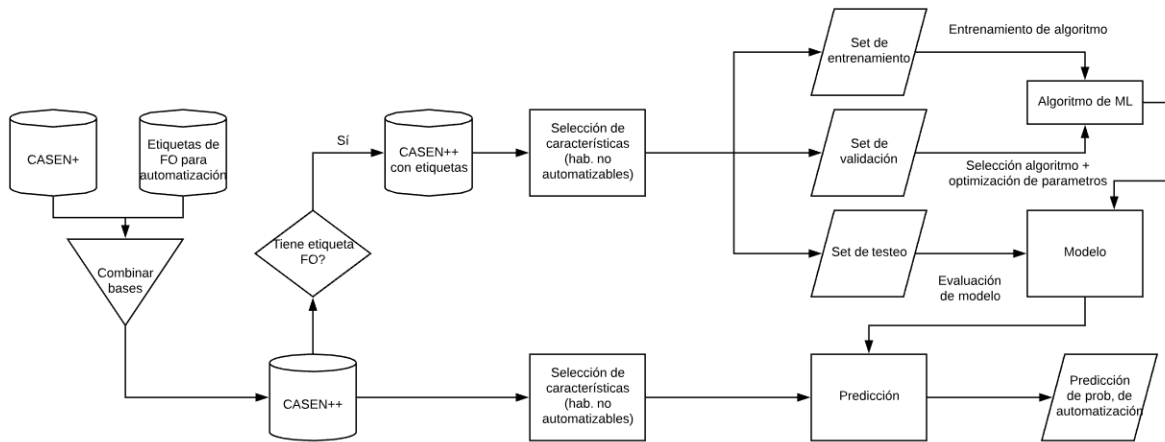


Figura 1.5 – Diagrama de flujo para modelo predictivo de probabilidad de automatización

Nota: CASEN+ corresponde a la base de datos CASEN 2017 que incluye la imputación de habilidades en el ambiente laboral, mientras que CASEN++ corresponde a la base de datos CASEN+ a la cual se le asigna la categoría de "Automatizable/No Automatizable" a cada ocupación, según Frey y Osborne (2017, disponible en Anexo A).

Fuente: Elaboración propia.

II.5.1. Procesamiento de los datos

Con la predicción del uso de habilidades en la CASEN logramos construir las variables características para el entrenamiento de los algoritmos que nos permitan obtener la probabilidad de automatización de las ocupaciones. Sin embargo, se hace necesario construir la variable objetivo, es decir, la categoría de si una ocupación es automatizable o

no. Algunos trabajos en la literatura ocupan directamente la probabilidad obtenida por Frey y Osborne (2017) para realizar los análisis, pero el argumento de este análisis es que dichas probabilidades fueron obtenidas a partir del estudio de las actividades realizadas por las ocupaciones en un mercado laboral que no es representativo de la realidad chilena. Por lo mismo, únicamente tomaremos del trabajo de Frey y Osborne (2017) la clasificación de las ocupaciones en automatizable/no automatizable.

Como se mencionó en la revisión de la literatura, los autores reunieron a un grupo de expertos de la Universidad de Oxford, los que analizaron la composición de tareas de cada ocupación y determinaron si, con el estado del arte de la tecnología al 2010, la ocupación era automatizable, no automatizable, o bien, aún no se podía asegurar el futuro de dicha ocupación. Para esto, tomaron la base de 702 ocupaciones de la O*NET (clasificados bajo el estándar del Departamento de Trabajo de Estados Unidos, SOC), y pudieron determinar con seguridad la categoría de 70 ocupaciones. Es importante mencionar que la categorización hecha por los expertos en 2010 se basaba en la tecnología existente hasta ese año, sin embargo, los desarrollos tecnológicos de los últimos años podrían sugerir que dicha base de datos está rezagada y, por lo tanto, sería necesario actualizar el listado de ocupaciones categorizadas¹⁴. La actualización de dicha base de datos requeriría del esfuerzo conjunto de un gran número de actores, por lo que queda fuera del alcance de este trabajo.

Para poder cruzar esta base de 70 ocupaciones categorizadas con las ocupaciones en la CASEN, es necesario asociar la ocupación de cuatro dígitos basado en el estándar CIUO-88 (utilizada en la CASEN) con la correspondiente ocupación SOC de Frey y Osborne (2017). Para esto, utilizamos la tabla de conversión de SOC-CIUO08 del *Bureau of Labor*

¹⁴ El cual probablemente incluiría ocupaciones de las que no se pudo establecer si sería automatizable, o que ocupaciones que se definieron como no automatizables pasen a integrar la lista de ocupaciones de las cuales no se puede establecer su categoría.

Statistics de Estados Unidos¹⁵ y la tabla de conversión entre CIUO08-CIUO88 de la OIT¹⁶. Realizando las conversiones respectivas, logramos generar una base de 90 ocupaciones bajo el estándar CIUO-88 con sus respectivas categorías. Un mayor detalle de las ocupaciones etiquetadas puede encontrarse en el Anexo A.

Asignamos la etiqueta de estas 90 ocupaciones a la base de datos de la CASEN, obteniendo una base de datos ampliada que considera la intensidad de uso de habilidades predicha y las categorías de automatización de las ocupaciones, en caso de que existan. De esta forma, particionamos la base de datos en una que contendrá todas aquellas observaciones para las que existía una categoría, la cual pasará a ser la base de entrenamiento, y la base para predecir, que serán todas las observaciones para las cuales no existe la variable objetivo. Esta primera base está compuesta de un total de 34.790 observaciones, mientras que la segunda está compuesta por las 56.889 observaciones restantes.

II.5.2. Entrenamiento de los algoritmos predictivos

En este caso, y a diferencia del anterior, el interés no está en clasificar a las observaciones dentro de alguna de las categorías (automatizable/no automatizable), sino que está en la probabilidad de que dicha ocupación sea automatizable, es decir, buscamos el valor de $\Pr(y^* = 1 | \mathbf{x}^*, \mathbf{X}, \mathbf{y})$.

Entrenaremos los mismos tres clasificadores utilizados para la clasificación del uso de habilidades, realizando el procedimiento de “búsqueda aleatoria de grilla” para determinar la combinación óptima de parámetros para los modelos. Al igual que en el caso

¹⁵ Bureau of Labor Statistics (2015). “ISCO 08 to SOC2010 Crosswalk”. Recuperado de: https://www.bls.gov/soc/ISCO_SOC_Crosswalk.xls

¹⁶ Organización Internacional del Trabajo. “Correspondence Table ISCO 88-08”. Recuperado de: <http://www.ilo.org/public/english/bureau/stat/isco/docs/corrtab08-88.xls>

anterior, usamos el valor de ROC-AUC como medida de desempeño del algoritmo, sin embargo, se presentan los resultados de los algoritmos para las tres métricas consideradas.

Los resultados se muestran en la Tabla 1.6.

Tabla 1.6 – Resultados entrenamiento de riesgo de automatización

Algoritmo	Exactitud	F1-Score	ROC-AUC
LR	0.62	0.61	0.68
RF	0.71	0.71	0.80
KNN	0.70	0.70	0.79

Fuente: Elaboración propia.

Dados los resultados obtenidos, seleccionamos el clasificador de Random Forest para la estimación de probabilidad, ya que tuvo el mejor desempeño entre los tres algoritmos para cualquiera de las métricas consideradas.

III Riesgo de automatización del mercado laboral chileno

III.1 Principales resultados para Chile

Con el algoritmo entrenado, se predice la probabilidad de automatización que tendría cada individuo, la que corresponde a la probabilidad de asignarle la categoría “automatizable” a dicha observación. Luego, para determinar el riesgo de automatización agregado se calcula la probabilidad promedio de automatización ponderada (por su factor de expansión) de todos los individuos ocupados en la muestra. Otra medida muy relevante en el análisis es el porcentaje de empleos con alto riesgo de automatización, es decir, aquellos que presentan una probabilidad igual o superior al 70% de ser automatizados, y por ende, de desaparecer.

Utilizando la metodología propuesta se estima que la probabilidad promedio ponderada de automatización en Chile es de 42,2%, a la vez que el 17,0% de los empleos

presentan un alto riesgo de automatización. La probabilidad promedio ponderada es inferior a lo encontrado en otros trabajos que estiman este número para Chile. Si se usa la CASEN 2017 el equivalente en número de empleos para el promedio está en torno a los 3,3 millones, mientras que alrededor de 1,3 millones de ocupados se encuentra en situación de alto riesgo de automatización.

Estas cifras agregadas esconden importantes diferencias entre distintos segmentos de la población. Por ello, es importante hacer un análisis más detallado para entender qué segmentos se encuentran en mayor riesgo de que su empleo sea sustituido debido a la automatización. A continuación, se realizan varios desgloses que permiten caracterizar a los ocupados por riesgo de automatización.

Nivel de calificación

Los resultados muestran que, entre los ocupados de alta calificación, la probabilidad promedio de automatización es de 37,0%, menor que la de los segmentos de mediana (40,1%) y baja calificación (53,1%). Si se consideran los empleos con alto riesgo de automatización, nuevamente la cifra entre los ocupados de alta calificación es menor a las de los otros segmentos, observándose que entre los ocupados de mediana calificación el 26,1% está en esta situación, lo que es considerablemente mayor al segmento de baja calificación (Tabla 1.7). En definitiva, si bien la probabilidad promedio de automatización es mayor en el segmento de trabajadores de baja calificación que en el de mediana, este último tiene una concentración mayor en empleos susceptibles a ser reemplazados por tecnología.

Según se observa en la Figura 1.6, los empleos de alta y de baja calificación presentan una mayor concentración en torno al promedio, mientras que los empleos de mediana calificación tienen una mayor dispersión. Esto demuestra que en los niveles de calificación alta y baja es menos frecuente la existencia de ocupaciones que presenten situaciones extremas en materia de riesgo de automatización. Contrariamente, las

ocupaciones de mediana calificación son más heterogéneas en cuanto al riesgo de automatización, encontrándose tanto ocupaciones con muy bajo riesgo de automatización (como aquellas en el grupo ocupacional 5), como ocupaciones con muy altas probabilidades de ser reemplazadas por tecnología (como aquellas en los grupos ocupacionales 4 y 8).

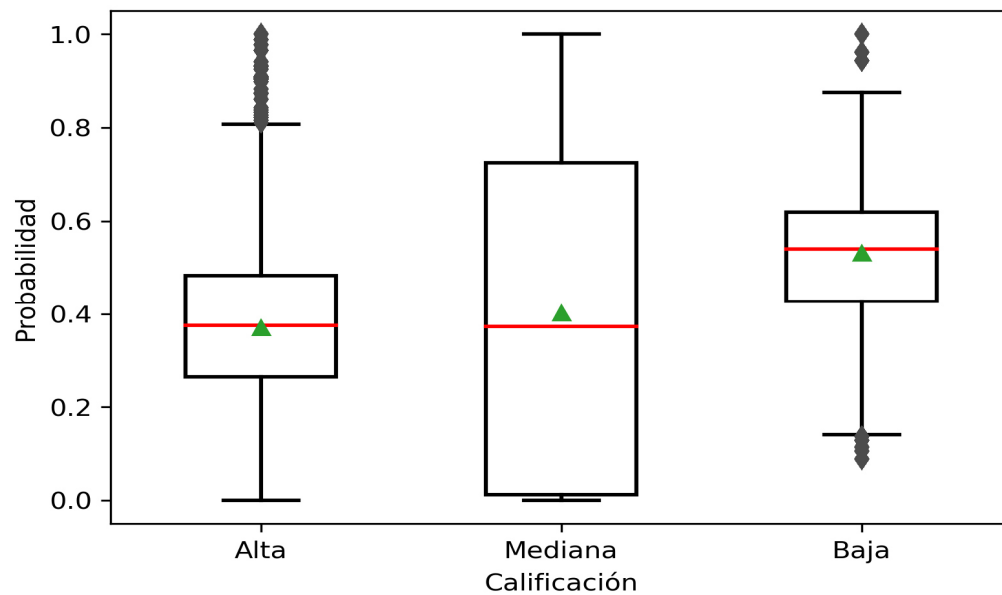


Figura 1.6 – Boxplot de la probabilidad de automatización, según nivel de calificación.

Nota: El triángulo verde corresponde a la probabilidad promedio de automatización, mientras que la línea roja corresponde a la mediana de la distribución de probabilidad. Los rectángulos corresponden al rango intercuartil (entre el cuartil 1 y cuartil 3). Los puntos fuera de boxplot corresponden a observaciones *outliers*.

Fuente: Cálculos propios, con base en CASEN 2017.

Tabla 1.7 – Distribución de riesgo de automatización por nivel de calificación

Nivel de calificación	Probabilidad promedio	% de trabajos alto riesgo
Alta	37.0	5.3
Mediana	40.1	26.1
Baja	53.1	11.0

Fuente: Elaboración en base a cálculos propios y CASEN (2017).

Sexo

Las cifras muestran que la probabilidad promedio de automatización para los ocupados hombres (46,6%) es mayor que la de las mujeres (36,4%), mientras que el porcentaje de trabajos en alto riesgo de automatización es ligeramente mayor para hombres que para mujeres (18,8% contra 14,7%). La distribución de probabilidades se puede observar en la Figura 1.7, en que se observa que los hombres presentan mayor concentración en torno al promedio que las mujeres. Esto se debe en parte a las mujeres se concentran más en grupos ocupacionales de menor riesgo de automatización, por ejemplo, en los grupos 2, 3 y 5, mientras que los hombres se concentran en ocupaciones con mayor susceptibilidad de automatización, como aquellas pertenecientes a los grupos 6 y 8.

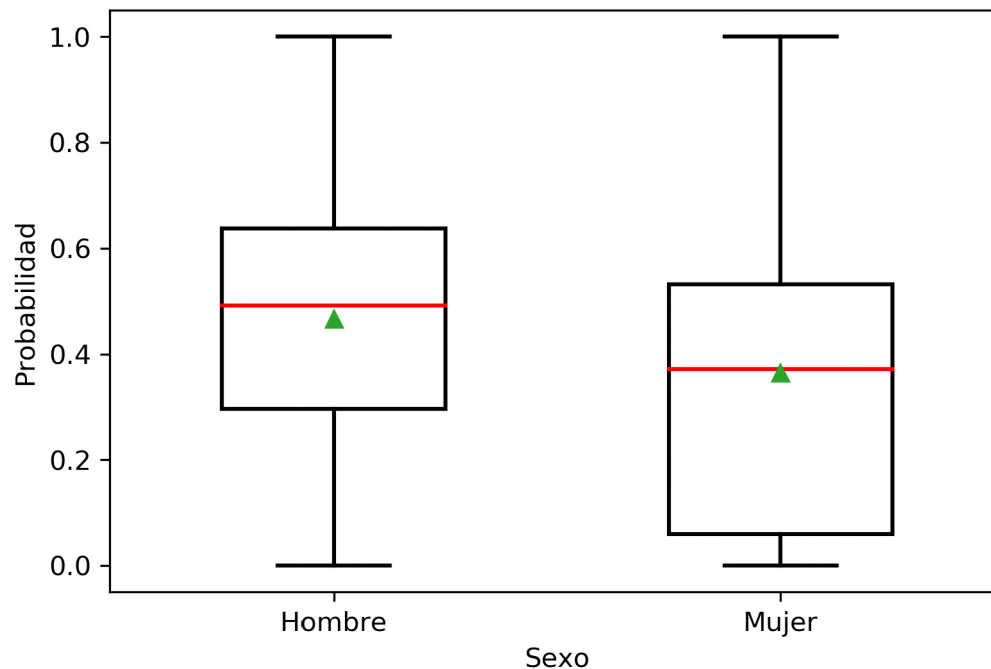


Figura 1.7 – Boxplot de la probabilidad de automatización, según sexo.

Fuente: Cálculos propios, con base en CASEN 2017.

Edad

La descomposición por tramos de edad muestra que el segmento de 25 a 34 años tiene una menor probabilidad promedio que los otros segmentos etarios, al igual que el menor porcentaje de trabajos en alto riesgo de automatización (ver Figura 1.8 y Tabla 1.8). El segmento de trabajadores más jóvenes muestra el mayor porcentaje de empleos con alto riesgo de automatización.

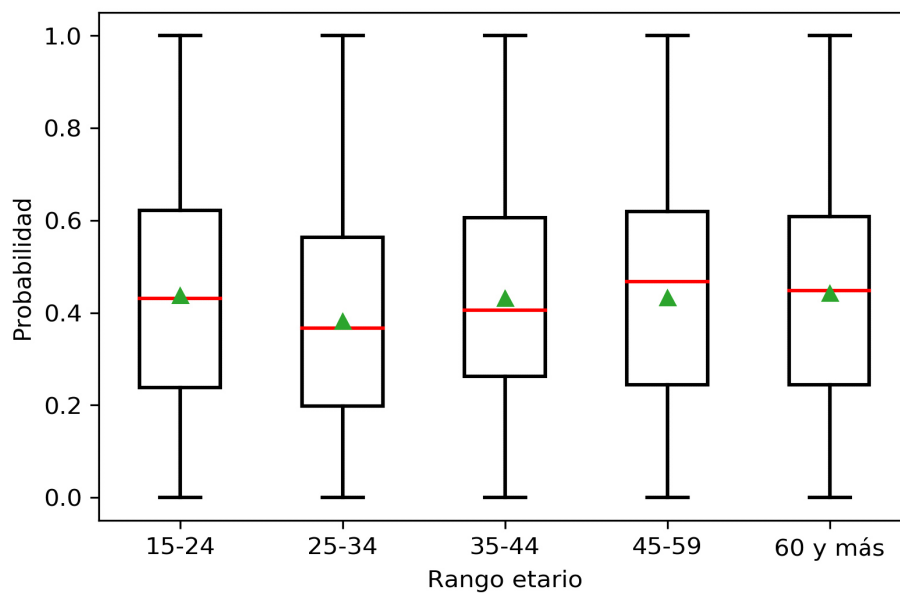


Figura 1.8 – Boxplot de la probabilidad de automatización, según rango etario.
Fuente: Cálculos propios, con base en CASEN 2017.

Tabla 1.8 – Distribución de riesgo de automatización por rango etario.

Rango etario	Probabilidad promedio	% de trabajos alto riesgo
15-24	43.8	19.8
25-34	38.2	12.3
35-44	43.2	20.7
45-59	43.2	17.2
60 y más	44.2	17.4

Fuente: Elaboración en base a cálculos propios y CASEN (2017).

Quintil de ingreso autónomo per cápita del hogar

El análisis por quintil de ingreso autónomo per cápita del hogar muestra que los quintiles 2, 3 y 4 presentan las mayores cifras tanto en probabilidad promedio como en el porcentaje de trabajadores de alto riesgo de automatización, mientras que los quintiles 1 y 5 presentan los menores valores, como se aprecia en la Figura 1.9 y la Tabla 1.9.

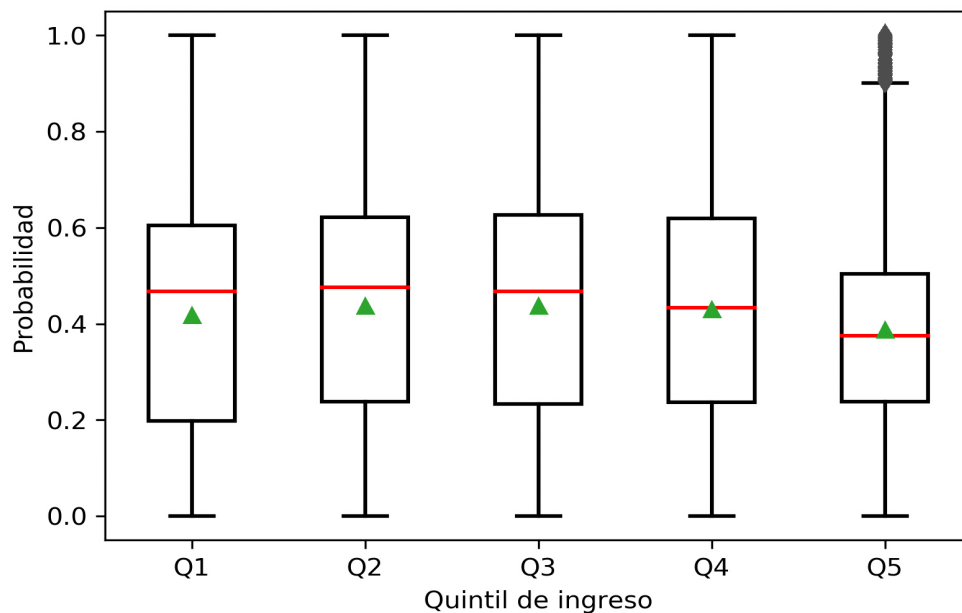


Figura 1.9 – Boxplot de la probabilidad de automatización, según quintil de ingreso autónomo per cápita del hogar.

Fuente: Cálculos propios, con base en CASEN 2017.

Tabla 1.9 – Distribución de riesgo de automatización por quintil de ingresos.

Quintil de ingresos	Probabilidad promedio	% de trabajos alto riesgo
Quintil 1	41.8	15.6
Quintil 2	43.7	18.6
Quintil 3	43.7	19.8
Quintil 4	43.0	19.0
Quintil 5	38.7	11.9

Fuente: Elaboración en base a cálculos propios y CASEN (2017).

Rama de actividad económica

A nivel de rama de actividad económica los resultados revelan un grado de variabilidad importante (ver Figura 1.10 y Tabla 1.10). Mientras en el sector de Transporte, almacenamiento y comunicaciones la probabilidad promedio de automatización es de 64,3%, en el sector de Hoteles y restaurantes es de 22,1%. Si analizamos el porcentaje de empleos en alto riesgo de automatización, la rama de Hogares privados con servicio doméstico presenta el menor porcentaje entre los sectores económicos, mientras que la de Transporte, almacenamiento y comunicaciones, la mayor, con el 46,9%, muy superior al 17,0% a nivel general. Esto se debe a una alta concentración en tareas rutinarias entre los trabajadores de este sector. En efecto, dentro de esta rama, el 54,1% de los trabajadores corresponden al grupo ocupacional 8, lo que los pone en una situación de vulnerabilidad mucho mayor que al resto de los ocupados. Por ende, esto pone una presión importante para ejercer acciones que permitan particularmente a los trabajadores de este sector adaptarse al actual escenario de Cuarta Revolución Industrial y reducir posibles impactos perniciosos.

Cabe destacar que las ramas de “Servicios sociales y de salud” y “Enseñanza” presentan una menor dispersión que las otras ramas de la economía, siendo poco frecuente encontrar ocupaciones de muy alto riesgo de ser automatizadas, lo que puede estar asociado a que corresponden a ocupaciones en donde se usan en forma más intensiva habilidades socio-emocionales.