



**Universidad Nacional de Costa Rica
Facultad de Ciencias Sociales**

Escuela de Economía

***Análisis de los Factores Determinantes del Desempleo en la
Región Central de Costa Rica, 2013-2022.***

Modalidad: Tesis de grado

Sustentantes:

**Jean Carlo Mongrillo Barquero
Mario Alberto Bolaños Gutiérrez**

Campus Omar Dengo, Heredia.

junio, 2024

Hoja con firmas de miembros del Tribunal Examinado.



Facultad de Ciencias Social
Escuela de Economía
Universidad Nacional



Hugo A. Fonseca A.

M.B.A. Hugo Fonseca Argüello
Representante del Decanato
-Quien preside-
Facultad de Ciencias Sociales

Dr. Jorge Andrey Valenciano Salazar
Director
Representante de la Dirección
Escuela de Economía

M.Sc. Maria Leonela Artavia Jiménez
Tutora

M.Sc. Juan Gabriel Alpizar Méndez
Lector

M.Sc. Jorge Solano Ruiz
Lector

Jean Carlo Mongrillo Barquero
Postulante

Mario Alberto Bolaños Gutiérrez
Postulante

Índice de Contenido

1.	Capítulo I: Generalidades de la investigación	3
1.1	Antecedentes	3
1.1.1	Investigaciones Mercado Laboral	3
1.1.1.1	Estudios Logísticos y Probabilísticos	3
1.1.1.2	Estudios con Modelos de Aprendizaje Automático	5
1.1.1.2.1	Modelos de Aprendizaje Automático Supervisado	5
1.1.1.2.2	Modelos de Aprendizaje Automático No Supervisado	6
1.1.2	Investigaciones Complementarias Mercado Laboral Costarricense	7
1.1.3	Políticas Públicas	9
1.2	Justificación y Planteamiento del Problema	10
1.2.1	Problema Investigación	12
1.3.	Objetivos de la investigación	13
1.3.1	Objetivo general	13
1.3.2	Objetivos específicos	13
2.	Capítulo II: Marco Teórico	14
2.1	Desarrollo Económico	14
2.2	Relación Trabajo y Desarrollo Económico	17
2.3	El mercado laboral	18
2.3.1	Tipos de Desempleo	18
2.3.2	Factores Relacionados con el desempleo	19
2.4	Pensamiento Computacional	24
2.5	Modelos de Aprendizaje Automático	26
2.5.1	Modelos de Aprendizaje Automático No Supervisado	26
2.5.2	Modelos de Aprendizaje Automático Supervisado	28
2.5.3	Optimizadores Instrumentos de Aprendizaje Automático Utilizados	29
2.5.3.1	Optimizadores Modelos de Aprendizaje Automático no Supervisado	29
2.5.3.2	Optimizadores Modelos de Aprendizaje Automático Supervisado	30
2.6	Modelos Econométricos	31
2.6.1	Modelos Logísticos	31
2.6.1.1	Validaciones Modelos Logit	32
2.6.1.2	Interpretación Modelos Logit: Odds Ratio	32
3.	Capitulo III: Marco Metodológico	34
3.1	Enfoque de la Investigación	34
3.2	Alcances	35
3.3	Universo de la investigación	35
3.3.1	Sujetos y fuentes de información	35
3.3.2	Muestra	36

3.4 Técnicas e instrumentos de investigación	37
3.5 Softwares Estadísticos Utilizados	40
4. Capítulo IV: Resultados	41
4.1 Evolución de las principales variables relacionadas con la oferta laboral en la Región Central	42
4.1.1 Contexto Región Central de Costa Rica	42
4.1.1.1 Contexto Demográfico	43
4.1.2 Condiciones Laborales en la Región Central	48
4.1.3 Situación Histórica de las Tasas de Desempleo en la Región Central	49
4.1.4 Caracterización de la población desempleada	53
4.2 Determinación de clústeres según las características de las personas desempleadas	54
4.2.1 Resultados de Clusterización por Período de Análisis	55
4.3 Identificación los principales factores demográficos que inciden el desempleo	60
4.3.1 Modelos Logísticos	60
4.3.1.1 Validación Modelos Logísticos Aplicados	60
4.3.1.2 Odds-Ratios de los Modelos Logísticos Aplicados	61
4.3.1.2.1 Modelos total muestra	61
4.3.1.2.2 Modelos separados por sexo	63
4.3.1.2.3 Modelos separados por zona de residencia	64
4.3.2 Modelos de Aprendizaje Automático Supervisado: explicación de la metodología aplicada	65
4.4 Recomendaciones de Acciones de Política	68
5. Conclusiones y Recomendaciones	72
5.1 Conclusiones	72
5.2 Recomendaciones	74
5.2.1 Educación	74
5.2.2 Idiomas	75
5.2.3 Programas de Empleabilidad	75
5.2.4 Aplicación de la Ciencia de Datos como Herramienta en Análisis Económicos	76
6. Anexos	77
7. Referencias	101

Índice de Tablas

Tabla 1. Ilustración Odds Ratios	33
Tabla 2. Secciones Encuesta Continua de Empleo INEC	38
Tabla 3. Estimación Población y Fuerza de Trabajo de Costa Rica por Regiones de Planificación I-2022	43
Tabla 4. Dominio Segundo Idioma (%) por Grupo de Edad Fuerza Laboral Región Central I-2013 y I-2022	46
Tabla 5. Calificación de la oferta de trabajo I-2013 y I-2022	47
Tabla 6. Calificación de la oferta de trabajo según el sexo I-2013 y I-2022	47
Tabla 7. Calificación de la oferta de trabajo según zona de residencia I-2013 y I-2022	48
Tabla 8. Tasa de Desempleo Región Central Costa Rica por grupo de edad I-2013 y I-2022	51
Tabla 9. Brecha de desempleo Región Central Costa Rica según Sexo I-2022	51
Tabla 10. Tasa de Desempleo Región Central Costa Rica según Nivel Educativo I-2013 y I-2022	52
Tabla 11. Tasa de Desempleo Región Central Costa Rica según Dominio Segundo Idioma I-2013 y I-2022	53
Tabla 12. Índices de Desempleabilidad Región Central Costa Rica I-2022	54
Tabla 13. Parámetros Optimizados Aplicación de Clusterización Desempleados Región Central de Costa Rica I-2013 y I-2022	55
Tabla 14. Resultados Clusterización Desempleados Región Central de Costa Rica I-2013	57
Tabla 15. Resultados Clusterización Desempleados Región Central de Costa Rica I-2022	58
Tabla 16: Prueba por Residuos (2 log likelihood - 2LL) a los modelos logísticos	61
Tabla 17: Pruebas Pseudo R-2 (Nagelkerke) a los modelos logísticos	61
Tabla 18. Odds-Ratios Modelos Logísticos Fuerza Laboral Región Central de Costa Rica I-2013 y I-2022	62
Tabla 19. ODDS Ratios Modelos Logísticos por Sexo Fuerza Laboral Región Central de Costa Rica I-2013 y I-2022	64
Tabla 20. ODDS Ratios Modelos Logísticos por Zona Residencia Fuerza Laboral Región Central de Costa Rica I-2013 y I-2022	65
Tabla 21. Resultados de Predicciones de Tasas de Desempleo según Escenarios de Mejoras en Niveles Educativos para Región Central de Costa Rica, I-2022.	66
Tabla 22. Enfoque de Clústers de desempleados. Región Central de Costa Rica	68
Tabla 23: Recomendaciones técnicas para una Política Pública de Empleo en la Región Central	71

Índice De Figuras

Figura 1. Estructura Pensamiento Computacional	25
Figura 2. Clustering Jerárquico de Tipo Disociativo	27
Figura 3. Funcionamiento Redes Neuronales Artificiales	29
Figura 4. Funcionamiento Validación Cruzada K-Fold	30
Figura 5. Regiones de Planificación de Costa Rica	42
Figura 6. Pirámide Poblacional Región Central de Costa Rica I-2022	44
Figura 7. Pirámide Poblacional Oferta de Trabajo Región Central de Costa Rica por sexo I- 2022	45
Figura 8. Nivel Educativo Fuerza de Trabajo Región Central I-2022	47
Figura 9. Tasa de Desempleo Región Central Costa Rica (Trimestres I-2013 al I-2022)	49
Figura 10. Tasa de desempleo Región Central Costa Rica según sexo (Trimestres I-2013 a I-2022)	50
Figura 11. Tasa de desempleo Región Central Costa Rica por zona y sexo I-2013 y I-2022	52
Figura 12: Resultados Gráficos Clusterización Desempleados Región Central I-2013	56
Figura 13: Resultados Gráficos Clusterización Desempleados Región Central I-2022	57
Figura 14: Sugerencia Plan de Acciones de Políticas Educativas Región Central	68

Dedicatoria

Dedico este trabajo a Dios, por darnos la bendición y la sabiduría necesaria para plasmar nuestra visión sobre lo que queríamos exponer en este proyecto. Seguidamente, se lo dedico a mis padres, pilares fundamentales, que siempre me apoyaron e impulsaron a conseguir todos mis objetivos. Y por último, dedicarlo a mi compañero de tesis y a todos y cada uno de los profesores y compañeros de carrera que de una u otra manera contribuyeron a la conclusión de este periodo de vida.

Jean Carlo Mongrillo

Dedico este trabajo en primer lugar a Dios, por darnos la sabiduría para lograr concluirlo. En segundo lugar, a mis papás, hermanas, y abuelos por todo el apoyo durante toda mi etapa universitaria. A mi compañero de tesis y a todos los compañeros y amigos que me acompañaron y apoyaron durante todos estos años en los que tuvo la dicha de estudiar en la Universidad Nacional. Y por último a todos los profesores que, con sus enseñanzas y consejos, serán siempre una pieza fundamental en mi vida profesional y personal.

Mario Bolaños

Agradecimiento

En primer lugar, a Leonela Artavia por todo su apoyo, guía y acompañamiento durante este proceso como tutora. A nuestros lectores, Juan y Jorge por todas sus recomendaciones y observaciones, las cuales nos permitieron concluir dicho estudio de una mejor manera.

A la educación pública y específicamente a la Escuela de Economía de la Universidad Nacional por brindarnos la posibilidad de formarnos como profesionales. A nuestras familias que siempre nos apoyaron en cada etapa de este estudio y por último a Oldemar Rodríguez por su apoyo y orientación técnica.

Introducción

El desempleo estructural existe en las economías por la presencia de desajustes entre las calificaciones de la fuerza de trabajo y los requerimientos de los empleadores; y la existencia continua de este fenómeno resulta ser un factor generador de problemas económicos derivados principalmente de las desigualdades sociales, causadas por la dualización de la sociedad, entre los empleados y los desempleados.

La Región de Planificación Central representa el 21% del territorio costarricense y en ella residen para el año 2022 el 62% de la población y el 66% de la fuerza laboral de Costa Rica. Por lo que teniendo en cuenta además que durante el periodo 2013-2022 esta región muestra tasas de desempleo superiores al 7%, se procede a analizar las principales variables determinantes de la desempleabilidad mediante el estudio de los microdatos resultantes de la Encuesta Continua de Empleo para dichos años.

El análisis se basa en la aplicación del Pensamiento Computacional para la resolución de problemas. Metodología, que presenta cuatro fases: descomposición del problema, reconocimiento de patrones, abstracción de la información y creación de un algoritmo (Plan). Por lo que el estudio parte de un análisis exploratorio descriptivo de la oferta de trabajo en los últimos 10 años, seguidamente se determina la existencia de patrones repetitivos en las características de las personas desempleadas a través de modelos de Aprendizaje Automático No Supervisado y en el que se logran encontrar 6 clústers claves.

Posteriormente, se identifican las variables que impactan en mayor medida la empleabilidad mediante modelos de tipo Logit y se aplica un modelo de Aprendizaje Automático Supervisado que simula casos hipotéticos de políticas públicas mediante la predicción de las posibilidades de empleabilidad de un individuo con características dadas. Resultados cuantitativos que se utilizan finalmente para evaluar las políticas públicas aplicadas en los últimos años, concluyendo así con una serie de recomendaciones de políticas dirigidas a obtener un mayor bienestar social.

1. Capítulo I: Generalidades de la investigación

1.1 Antecedentes

Como antecedentes a la investigación se presentan investigaciones previas realizadas al mercado laboral principalmente relacionados con aplicaciones de modelos econométricos y de Aprendizaje Automático. Así mismo se presenta el contexto de la oferta laboral actual de la región en análisis basado en los resultados de la Encuesta Continua de Empleo. A su vez, como complemento adicional se plantean las principales políticas públicas aplicadas en los últimos años dirigidas a reducir el nivel de desempleo en el país analizado.

1.1.1 Investigaciones Mercado Laboral

1.1.1.1 Estudios Logísticos y Probabilísticos

Entre los principales estudios aplicados al análisis de del desempleo en las economías, se destaca a Dhiman, Saraogi y Devatha (2019), quienes realizan un modelo de regresión logística para determinar las causas del desempleo entre los jóvenes de la India. Sus resultados indican como bajo el supuesto “ceteris paribus”, para los jóvenes (15–35 años) con cada año que pasa, la probabilidad de estar desempleado disminuye en un 4%. Además, que con cada año de educación se disminuye la probabilidad de estar desempleado en un 12%.

Para el caso de España, Martínez (2015), determina la influencia de variables demográficas sobre el desempleo. Su estudio se basó en datos de la encuesta trimestral de la población económicamente activa, la utilización de las variables: Edad, Nacionalidad, Nivel de estudios y sexo, con la influencia del sexo, la nacionalidad y la educación; y la aplicación de un modelo Logit. Sus resultados indicaron que todas las variables anteriormente mencionadas resultaron significativas. Siendo la educación superior y la educación secundaria las más influyente con una reducción del 17% y 6% respectivamente, en las probabilidades de estar desempleado.

En Latinoamérica, se destacan los estudios logísticos de Salamanca (2015), quien identifica en Medellín al género, ser jefe de hogar, estar comprometido, la cantidad de años de educación y la edad son variables influyentes en la condición laboral de la región, a través de los datos de la Gran Encuesta Integrada de Hogares (GEIH).

Para Ecuador, Martínez y Solórzano (2018) concluyen mediante un modelo logístico y la utilización de los datos de Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU), que la edad, la zona de residencia, el sexo, la experiencia al cuadrado, jefatura de hogar y educación son factores significativos en alterar la probabilidad de estar desempleado.

Por su parte, Márquez (2015) identifica para el caso mexicano con base en los microdatos de la Encuesta Nacional de Ocupación y Empleo y el uso de un modelo logístico multinomial que analiza el desempleo abierto y el desempleo desalentado, que los jóvenes, las personas con bajos niveles de escolaridad, el tener pocas responsabilidades en el hogar y las personas con calificaciones en actividades poco especializadas, son quienes tienen mayores probabilidades de ser desempleados.

Para el caso de Costa Rica, Pacheco (2013) estima las determinantes del desempleo juvenil en Costa Rica, con datos de la Encuesta Nacional de Hogares 2012 y la utilización de modelos logísticos. Obteniendo como las principales variables que probabilísticamente favorecen el estar desempleado: la zona de residencia siendo la zona rural la más propensa, el nivel de educación con mejoras en oportunidades a partir de secundaria completa y el sexo de las personas, teniendo a su vez la mujer estadísticamente el doble de chance de estar desempleada respecto a un hombre.

Morales y Segura (2018) y Fernández (2022), determinan mediante un modelo logístico enfocada a la población económicamente activa femenina y con la utilización de los datos de la Encuesta Nacional de Hogares que los factores que inciden mayoritariamente en que una mujer se logre integrar a la fuerza de trabajo son el puesto que desempeñen el hogar, sus niveles de educación, zona de residencia, cantidad de niños en el hogar, sumado a las de

barreras de entrada existentes causadas principalmente por las responsabilidades domésticas adquiridas en los hogares por las femeninas que son cabezas de hogar.

1.1.1.2 Estudios con Modelos de Aprendizaje Automático

La aplicación de algoritmos de Aprendizaje Automático en investigaciones económicas empíricas ha sido promovida en los últimos años por economistas como Athey (2018) y (2019), Varian (2014). Quienes proponen su utilización principalmente para la modelación de comportamientos complejos, debido a su alta capacidad predictiva y análisis de relaciones no lineales. Por lo que dichas herramientas se plantean como un gran instrumento de apoyo en el diseño y análisis de políticas económicas y además como un complemento importante con los habituales análisis econométricos.

1.1.1.2.1 Modelos de Aprendizaje Automático Supervisado

Respecto a investigaciones específicas de Aprendizaje Automático Supervisado aplicado al mercado laboral, Martínez (2018) con datos del Instituto Nacional de Empleo y Desarrollo Profesional de Portugal y con algoritmos de tipo XG-Boost, predice las posibilidades o riesgo de quedar desempleadas, con la intención de fortalecer a las personas con mayores probabilidades mediante el desarrollo y mejora de sus habilidades, previniendo así que en el futuro se conviertan en desempleados.

Morales (2018) realiza una modelación predictiva de las probabilidades de inserción laboral mediante la utilización del algoritmo Bosques Aleatorios, enfocado hacia el análisis particular de las mujeres. Sus resultados obtuvieron un 97% de precisión y establecieron principalmente que para el caso de las mujeres concluir la secundaria aumenta sus probabilidades de inserción en 7 puntos porcentuales y en la que tener 40 años reduce 10 puntos porcentuales la probabilidad en comparación con tener 30 años.

Dhiman, Saraogi y Devatha (2019) utilizan seis algoritmos de aprendizaje automático supervisado para clasificar a los jóvenes de la India según sus factores demográficos como desempleados y empleados. Modelación en la que, para el caso de los estimadores por Redes Neuronales y Árboles de Decisión, se obtuvieron precisiones superiores al 98%, evidenciando así su alto poder predictivo.

Kutuk (2019) realiza una predicción de las probabilidades de transición de desempleo a empleo para Turquía. Su análisis consistió en realizar una comparación entre modelos de Aprendizaje Automático Supervisado con la utilización de los algoritmos Bosques Aleatorios y XG Boosting y por otro lado un enfoque econométrico mediante la aplicación de una Regresión Logística. Estudio que obtuvo como resultado que el modelo logístico obtuvo índices de precisión menores en comparación con el modelo de AAS, evidenciado de dicha manera el alto poder predictivo de las nuevas tecnologías informáticas.

1.1.1.2.2 Modelos de Aprendizaje Automático No Supervisado

En lo que respecta modelos de Aprendizaje Automático No supervisado se destaca el análisis de Núñez, F. Rodríguez, A y Usabiaga, C (2014), en el que se identifica mediante métodos de AANS la existencia de patrones repetitivos en los desocupados de Andalucía, con la presencia de 5 distintos grupos de trabajadores sin empleo con características intragrupalas muy similares, siendo estas principalmente su región, edad y especialidad académica.

Varian (2014) ejemplifica la aplicación de las tecnologías de aprendizaje automático a través de la identificación de cuales variables resultan ser importantes predictoras del crecimiento económico para 72 países distintos y 42 variables distintas con datos de Sala-i-Martin (1997). Investigación que permitió definir a la expectativa de vida, apertura de la economía, grado de capitalismo, inversión en equipamiento, el PIB de 1960 y al Imperio de la Ley como significativas.

Dhiman, Saraogi y Devatha (2019) realizan un análisis para las principales ciudades de la India, a través de la aplicación de un Análisis de Componentes Principales para reducir la dimensionalidad de los datos, ya que los datos sin procesar tienen 335 atributos. Seguidamente realizan una agrupación de las ciudades mediante el algoritmo de K-medias, lo que permitió posteriormente estimar un Modelo de Bosques Aleatorios, el cual a su vez contribuyó finalmente para la predicción del nivel de desempleo por Estado a través de una regresión lineal múltiple.

1.1.2 Investigaciones Complementarias Mercado Laboral Costarricense

Villegas y Ulate (2013) realizan un análisis dinámico del mercado laboral para el periodo 2010-2011. Investigación que analiza los cambios en las condiciones laborales de la fuerza de trabajo en el corto plazo y en la que se establecen como variables altamente influyentes en las condiciones laborales de los costarricenses: la edad, el nivel educativo, el hablar otro idioma, el género, el estado conyugal, la zona de residencia, el encontrarse en subempleo, la experiencia laboral y la rama de actividad.

Sánchez (2014) mediante la estimación de la curva de Beveridge, determina la existencia de un desempleo estructural en Costa Rica debido a las divergencias existentes entre los perfiles de los desocupados y los demandados por las empresas, sumado una oferta académica que no es consecuente con los requerimientos del sector privado. Además, como la imposibilidad de que algunos trabajadores migren de sectores de menor crecimiento (construcción, agro, entre otros) a otros de mayor crecimiento (servicios financieros, a empresas etc.); así como problemas de información entre oferentes y demandantes.

Por otro lado, Loria (2016) describe las tendencias de la población costarricense, según todos sus componentes y los distintos indicadores presentados por el INEC. Identificando, así como los principales grupos afectados por el desempleo para dicho año, son las mujeres, seguido por las personas de centros urbanos, la población joven, los que poseen bajo nivel de estudio y por último los residentes en zonas fronterizas y costeras.

Morales (2017) se refiere a las causantes de los problemas estructurales del mercado laboral costarricense estableciendo como el desalineamiento existente entre la oferta y la demanda laboral, el desenganche entre el crecimiento económico y el crecimiento del empleo, como consecuencia, entre otras razones, del cambio en la estructura productiva del país, y del mayor dinamismo en actividades intensivas en capital y no en mano de obra; la concentración de la producción y las oportunidades laborales en pocas zonas y regiones y los avances tecnológicos que vienen desplazando a la mano de obra.

La Organización para la Cooperación y el Desarrollo Económico, en su estudio del año 2016 para la economía costarricense, presenta la existencia de problemas en el mercado laboral relacionados principalmente por la baja productividad, ya que a pesar de que es uno de los países donde más horas en promedio se trabaja en la Región, su productividad es tres veces inferior al promedio de los países miembros de la organización, señalando incluso (OCDE,2016):

De no corregir esta situación, Costa Rica “corre el riesgo de quedarse atrapada en un "círculo vicioso", donde las personas con pocas habilidades y acceso limitado a oportunidades quedarán limitadas a desempeñar trabajos de baja productividad con salarios bajos, reduciendo la productividad agregada y empeorando la desigualdad” (p.62).

Para el año 2019, Morales plantea la existencia de un desacople entre la oferta y la demanda de trabajo en cuanto a nivel de cualificación. A su vez la apertura comercial también afecta la generación de empleo en tanto las actividades productivas desarrolladas internamente no sean capaces de competir en el mercado internacional o contra las importaciones. Adicionalmente el cambio tecnológico y la robotización han venido desplazando a la fuerza de trabajo, tanto a nivel nacional como internacional, debido a que muchas actividades se han venido mecanizando y con ello desapareciendo puestos de trabajo (Morales, Comunicación personal, 01 de abril de 2019)

Por último, evidenciando el problema de desempleo existente para el caso de Costa Rica, Pacheco plantea como es posible atribuirle aproximadamente en un 70% a causas estructurales. Revelando de dicha manera la alta complejidad de dicha problemática en el diseño de políticas que contribuyan en su reducción. Pudiéndose considerar inclusive unos de los principales problemas económicos costarricenses para el año 2019. (Pacheco, entrevista personal, 07 de abril del 2019).

1.1.3 Políticas Públicas

Respecto a las políticas públicas aplicadas durante los últimos años dirigidas a reducir el desempleo de Costa Rica, se destaca en el año 2014 el diseño de una Estrategia Nacional de Empleo y Producción, creada con la intención de colocar la política de empleo como objetivo estratégico institucional del Estado costarricense y establecida además como una medida para combatir directamente la pobreza y la desigualdad (MTSS,2014). Su principal objetivo consistió bajo el marco de una Economía Social Solidaria reducir la tasa de desempleo abierto al 7% para el año 2018, mediante la creación de al menos 217.000 nuevos empleos.

De las herramientas propuestas durante dicha estrategia se destaca la promoción de la inversión privada, apoyo a las PYMES y a las mujeres, sumado a inversiones públicas en la ampliación de la infraestructura, las comunicaciones y la energía, así como la inversión en capital humano especializado. Además, se reconoce el interés hacia el sector agrícola, por lo que se propone favorecer su acceso al crédito, su generación de valor agregado y su comercialización mediante la formación de mercados de precio justo y aumentando la asistencia técnica hacia dichas actividades (MTSS,2014).

A la Estrategia Nacional de Empleo y Producción, se le añaden como políticas dirigidas al mercado de trabajo, el programa de Red de Cuido y Desarrollo Infantil, dirigido a los hijos de las madres que desean laborar, permitiéndoles así acceder a un empleo con jornada completa; el programa virtual Empléate que intenta apoyar mediante orientación, formación e intermediación laboral a los jóvenes y discapacitados y el Programa Nacional de Empleo (PRONAE) que favorece a las personas pobres desempleadas (MTSS,2014).

Por último, en el Plan Nacional de Desarrollo e Inversión Pública 2019-2022 (PNDIP) presentado por el gobierno de Alvarado, se presenta como meta nacional la reducción del desempleo. Siendo las principales políticas aplicadas de carácter educativo, como son el aumento en un 50% las personas graduadas del INA y el desarrollo de programas de Bilingüismo y de educación técnica dual para la población (MIDEPLAN,2018, p91).

El PNDIP anteriormente mencionado se propone además el aumento en la cantidad de emprendimientos financiados por el Sistema de Banca para el Desarrollo y favorecimiento de la inserción de las personas con discapacidad, proponiéndose lograr según estimaciones una tasa de desempleo abierto en un rango de 8,4% y 7,3% para el año 2022, es decir una reducción entre 0.8 y 1.9 puntos porcentuales, con respecto a la línea base (MIDEPLAN,2018, p93).

1.2 Justificación y Planteamiento del Problema

Con la transformación del modelo productivo costarricense hacia una mayor presencia del sector terciario y una reducción del sector manufacturero y primario (INEC,2022) y a la llamada revolución industrial 4.0, que incorpora al modelo productivo tecnologías como la robótica, la analítica, la inteligencia artificial, las tecnologías cognitivas, la nanotecnología y el Internet de las cosas. Se afectan no solo los procesos de fabricación de los productos, sino que sufren transformaciones los demás agentes económicos incluyendo a la fuerza de trabajo.

Revolución que sin duda exige el desarrollo de estrategias de modernización de las características de la fuerza de trabajo, con la intención de propiciar un adecuado nivel de bienestar futuro en la sociedad. En las investigaciones planteadas y en los datos descriptivos obtenidos con base en las encuestas realizadas por el INEC, se han observado factores que han dificultado el equilibrio entre ofertantes y demandantes de trabajo en Costa Rica como son: la edad de los trabajadores, su calificación educativa incluyendo la carencia del dominio de segundo idioma y su nivel de experiencia.

Pese a ello, en la actualidad no se conocen con claridad la existencia de tipos de desocupados con rasgos similares o las variables que determinan estadísticamente en mayor medida la alta tasa de paro aplicado a nivel específico a la Región de Planificación Central en concreto.

Sumado a ello, la economía costarricense se encuentra en una etapa del ciclo económico tenso, con tasas de crecimiento en su producto interno bruto moderadas, finanzas públicas con resultados deficitarios y una deuda pública superior al 50% respecto al PIB (Hacienda,2019).

Condiciones que evidencian la necesidad de mejorar la situación actual del mercado laboral, con la intención de suavizar los efectos de las dificultades económicas en las personas, específicamente a las que, debido a su desocupación, su nivel de bienestar se reduce cada día más. Sumado a que la existencia de altos niveles de desempleo afecta directamente otras variables relevantes como son la pobreza y el nivel de delincuencia, causando finalmente una afectación en el desarrollo económico del país.

Durante los últimos años se han realizado esfuerzos hacia la atracción de inversión extranjera directa, garantizar el cumplimiento de las legislaciones laborales en temas como salarios mínimos, seguridad social y reglamentos empleador-empleado. Así mismo, se han aplicado programas educativos a cargo del Ministerio de Educación Pública (MEP) y en el Instituto Nacional de Aprendizaje (INA) junto a programas que pretenden facilitar la coordinación entre las ofertas y demandas de empleos como es el caso del programa Empléate.

Políticas que deberían permitir acceder a un empleo de calidad a la población, poniendo especial énfasis en los grupos más vulnerables del país. Sin embargo, hasta la fecha dichas políticas resultan ser insuficientes o inadecuadas, debido a que la tasa de desempleo se mantiene superior al 7% durante el periodo 2013-2022.

Contexto que justifica la importancia de realizar un análisis profundo de la oferta laboral de la región en estudio con la intención de diseñar políticas públicas más concretas dirigidas a grupos específicos de personas o sectores, que intenten subsanar las deficiencias existentes entre ofertantes y demandantes y que por tanto conduzcan a disminuciones en las tasas de desocupados, trayendo consigo aumentos en el desarrollo económico nacional y por tanto mejoras en el bienestar de la población en general, teniendo en cuenta además del

impacto negativo observado asociado por la pandemia del Covid-19, en la cual el nivel de desempleo tuvo un aumento significativo en algunos sectores.

1.2.1 Problema Investigación

La Región de Planificación Central tiene el mayor porcentaje relativo de habitantes, pese a que solo cubre el 21% del territorio costarricense, y es donde reside el 66% de la fuerza laboral, debido en gran parte a que la mayoría de las compañías generadoras de empleo durante los últimos años han tendido a desarrollar sus actividades económicas en la zona.

El mercado laboral tal y como lo plantea Morales (2017) tiene un problema carácter estructural, ya que la población presenta una inconsistencia en su nivel de cualificación, respecto a las cualificaciones demandadas, condición que ha contribuido a que la tasa de desempleo en la región sea superior al 7,5% en el periodo 2013-2022.

Debido a los problemas estructurales existentes, la población desempleada requiere mejorar o cambiar sus calificaciones para lograr insertarse al mercado laboral. Consciente de lo anterior el gobierno ha intentado con diversas políticas mejorar el escenario, pero en vista que la tasa de desempleo no se reduce, es posible afirmar que estas estrategias han sido inadecuadas o quizás insuficientes.

Por lo que, el problema al que se refiere la investigación es a entender ¿Cómo las condiciones de la oferta laboral han llevado a la existencia de desempleo estructural en la Región de Planificación Central de Costa Rica durante los años 2013-2022?

1.3. Objetivos de la investigación

1.3.1 Objetivo general

Analizar los factores determinantes del desempleo en la Región de planificación central para el periodo 2013-2022.

1.3.2 Objetivos específicos

- Describir la evolución y comportamiento de las principales variables relacionadas con la oferta laboral de la región central durante el periodo en análisis.
- Determinar la existencia de grupos de personas desempleadas con características similares mediante la aplicación de algoritmos de Aprendizaje Automático No Supervisado.
- Identificar los principales factores demográficos que inciden en el desempleo mediante la aplicación de modelos Logit y modelos predictivos basados en algoritmos de Aprendizaje Automático Supervisado.
- Recomendar acciones de políticas que contribuyan a reducir el desempleo estructural de la región.

2. Capítulo II: Marco Teórico

En el presente capítulo se presentan los fundamentos teóricos relacionados con el tema de esta investigación, introduciendo así al desarrollo económico y su relación con el mercado laboral con el fin de entender su estrecha relación en las economías. Partiendo, de un desarrollo en las economías visto desde una perspectiva enfocada en la sociedad y sus libertades, así como estas se relacionan directamente con la posibilidad de acceder a un empleo digno.

Seguidamente se analizan los componentes del mercado laboral, en el que confluyen la oferta y demanda de trabajo. Sumado a una sección específica enfocada en presentar en detalle el desempleo, sus tipos y los factores que desde el punto de vista teórico afectan positiva o negativamente la empleabilidad de las personas; lo anterior con la intención de poder tener un sólido punto de partida para el análisis de este fenómeno.

Finalmente, se presentan los aspectos relacionados con la metodología aplicada en la investigación, iniciando con el Pensamiento Computacional y sus componentes, el cual es la base angular de la estructura de la investigación. Por último, se plantean los detalles de los instrumentos econométricos e informáticos que fueron aplicados en la investigación para realizar la analítica deseada, así como las herramientas utilizadas para optimizar dichas estimaciones.

2.1 Desarrollo Económico

En la literatura económica existen diversas corrientes que se inclinan por definir el desarrollo económico como un proceso de crecimiento económico, relacionado con la acumulación de rentas, a través del estudio de indicadores cuantitativos como el crecimiento del Producto Interno Bruto o el incremento de los ingresos per cápita.

Sin embargo, la presente investigación toma como base el concepto de desarrollo económico planteado por Sen (2000), donde se plantea a este desde un punto de vista centrado en el ser humano y en las libertades reales que este tiene para lograr los objetivos en su vida, es decir no consiste en analizar la cantidad de ingresos de las personas sino en el tipo de vida que estas tienen en la práctica y su nivel de bienestar.

Por lo que el desarrollo económico en dicha investigación se basa en el proceso de expansión de las libertades humanas reales que disfrutaban las personas y en el que si bien se reconoce la importancia del crecimiento económico en las economías y se considera como una parte importante del desarrollo ya que es uno de los medios para lograr esta expansión, en esta se identifican la existencia de otros factores que afectan estas libertades, tales como los derechos civiles y políticos, así como la presencia de instituciones sociales y económicas y el acceso a oportunidades de empleo, educación y atención médica.

Por tanto, teniendo claro lo anterior, el concepto de libertad humana dentro de esta teoría se entiende según Sepúlveda (2013) como el fin mismo u objetivo final del desarrollo económico, ya que las acciones de las personas en principio se dirigen a alcanzar su libertad, de tal manera que esta les permita alcanzar una vida dichosa o por lo menos justa.

Dentro del concepto de libertad como eje del desarrollo económico planteado por Sen (2000), se plantean dos distintas vertientes en las que se divide esta misma. La primera que presenta a la libertad como un conjunto de procesos, es decir a la capacidad de decidir de las personas por ellas mismas. Y en segundo lugar se concibe la libertad desde un punto de vista de oportunidades, es decir de la oportunidad que las personas tienen de lograr sus objetivos. Entendida además como la capacidad real de conseguir o alcanzar cosas, en el que se destacan las oportunidades por ejemplo de tener una vida larga y digna, así como tener acceso a un empleo justo.

La investigación se enfoca en la segunda vertiente del desarrollo económico, por lo que a continuación se detallan las cinco libertades intermedias que según Sen (2000) le pertenecen a esta sección y que según el autor son instrumentos o medios claves para el desarrollo de una economía:

- Libertades políticas: Hace referencia a la libertad de conciencia, expresión y oposición políticas, así como al pluralismo político y los derechos de voto, es decir se refiere a la oportunidad de decidir de las personas acerca de quien las gobierna y sobre qué principios.

• Libertades y Derechos Económicos: En esta se incluye la libertad de funcionamiento de los mercados y los derechos de las personas de disfrutar sus recursos económicos y prosperar, es decir la oportunidad de realizar intercambios, producir o consumir recursos, así como la existencia y acceso a recursos financieros por parte de todos los agentes económicos y además a la oportunidad de acceder a un empleo justo y un mercado laboral abierto.

• Oportunidades sociales: Hace referencia a la existencia de recursos sociales como son los servicios públicos como la educación y sanidad de la sociedad, los cuales les permitan a las personas acceso a educación y tener una vida larga y digna con salud, acompañado además de la presencia de libertades como la equidad de género, bienestar de las mujeres y cuidado de los niños.

• Garantías de transparencia: Parte de la garantía de un “limpio” funcionamiento de todos los demás sistemas sociales, políticos y económicos, es decir la existencia de confianza en las relaciones sociales y mecanismos para buscar justicia, de tal manera que se prevengan acciones como la corrupción, la irresponsabilidad financiera y de tratos poco limpios (SEN, 2000, p. 59).

• Seguridad Protectora: Esta libertad parte del hecho de que en la sociedad por lo general existen personas vulnerables a sufrir privaciones personales por lo que se propone la existencia de una red de protección social que impida que esta parte de la población caiga en miserias e incluso la muerte. Por tanto, hace referencia a la existencia de instituciones fijadas por la ley que contribuyan a minimizar efectos del desempleo, indigencia y pobreza.

2.2 Relación Trabajo y Desarrollo Económico

Para la presente investigación, se presenta al trabajo como una actividad que tiene una serie de efectos cotidianos en el bienestar de las personas y por consiguiente en sus vidas. Ya que según Grabois (2014), el trabajo debe concebirse como un concepto que posee un valor que sobrepasa lo económico ya que se convierte en una condición integradora socialmente de las personas y que contribuye al desarrollo de su dignidad, autoestima, fraternidad y les permite organizar sus tiempos en torno a sus proyectos de vida.

El concepto de trabajo, se relaciona directamente con la teoría del desarrollo como libertad planteada anteriormente, específicamente en la libertad intermedia llamada “Libertades y Derechos Económicos”, ya que en ella se contempla como la capacidad de acceder a un empleo justo es un medio o instrumento clave para el desarrollo, debido a que la garantía de un trabajo justo para todas las personas es además, una garantía de justicia social, ya que el trabajo, es la fuente de riqueza, o de subsistencia en la mayoría de los casos del ser humano.

Por lo que desde el punto de vista económico, la garantía de un trabajo decente para las personas también no solo contribuye a una mejora en las libertades económicas de las personas, sino que también según la OIT (2015), el permitir el acceso a un empleo a las personas interviene de otras maneras al desarrollo en las economías, como por ejemplo en las reducciones de los niveles de pobreza y del hambre, disminución de las desigualdades, hacer el planeta más limpio, más seguro y sustentable, así como impulsar mejoras en la salud, educación, justicia y prosperidad de las personas.

2.3 El mercado laboral

El mercado laboral según Kiziryán (2018) se define como aquel donde confluyen la oferta de trabajo que está formada por el conjunto de trabajadores que están dispuestos a trabajar y la demanda de trabajo, la cual es entendida como el conjunto de empresas o empleadores que se disponen a contratar personas. Así mismo, es importante destacar que en este mercado además intervienen otros componentes como son el gobierno a través de las normas jurídicas, las instituciones sociales, las relaciones colectivas y las valoraciones personales.

Para el caso en estudio, el concepto de oferta laboral llamada técnicamente fuerza de trabajo se entiende según INEC (2012) como el conjunto de personas de 15 años o más con las capacidades y deseos de participar en la producción de bienes y servicios de una economía, incluyendo así tanto los que se encuentran ocupados como los que se encuentran buscando activamente un empleo dentro del mercado, los cuales se plantean como desempleados. Por tanto, a nivel conceptual la tasa de desempleo se establece como el porcentaje de la población desempleada respecto a la fuerza de trabajo (FT).

$$Tasa\ de\ desempleo = \frac{Población\ desempleada}{Fuerza\ de\ trabajo(FT)}$$

2.3.1 Tipos de Desempleo

Vicente (2015) plantea la existencia de 4 distintos tipos de desempleo, inicialmente el desempleo friccional que es aquel resultante del tiempo necesario para emparejar a los trabajadores con los puestos de trabajo, ya que lo lógico, en estos casos, es que los individuos se tomen un tiempo en buscar un trabajo que se ajuste a sus cualificaciones y con un salario acorde, así como que las firmas tarden en obtener el capital humano deseado. Y según el autor este es causado principalmente por la presencia de información imperfecta entre ofertantes y demandantes, así como por el proceso de destrucción y creación de empleo y por el aumento de la oferta de fuerza laboral.

Por otro lado, el desempleo estacional, que es aquel dependiente de los factores climáticos y de los cambios en la demanda de trabajo debido a las distintas estaciones del año y el desempleo cíclico, conocido también como desempleo coyuntural, que está presentes principalmente en recesiones económicas de manera contra cíclica y es definido como el que aparece cuando la demanda agregada es insuficiente para alcanzar la producción de pleno empleo o potencial.

Por último, el desempleo estructural que surge por la existencia de desajustes entre la oferta y la demanda de trabajo; ya sea por excesos de oferta de trabajo en una actividad económica o por la no correspondencia entre las calificaciones requeridas y las ofrecidas en el mercado laboral. Y este inclusive en ocasiones se puede presentar por la presencia de políticas de sociales en pro de la protección hacia los trabajadores, que desincentiva la demanda de trabajo o la búsqueda de trabajo.

El desempleo estructural, según Blanchard y Johnson (2013), puede darse debido a factores como cambios en la estructura económica de un país, como por ejemplo el cambio de una economía basada en el sector primario hacia una más dirigida a sectores secundarios o terciarios. O además cuando se dan reubicaciones de grandes empresas, que dejan a sus anteriores trabajadores con dificultades para encontrar empleos que calcen con las habilidades desarrolladas anteriormente.

En segundo lugar, Acemoglu y Restrepo (2020) definen como una posible causante del desempleo estructural, el surgimiento de avances tecnológicos como la automatización de trabajos que convierten en obsoletas las calificaciones de cierta parte de la oferta laboral, convirtiendo a esta parte en no concordante con las habilidades demandadas.

2.3.2 Factores Relacionados con el desempleo

La investigación se enfoca en identificar relaciones entre el desempleo y diversas variables demográficas, por tanto, a continuación, se plantean referencias teóricas que explican dichas relaciones

Edad

La edad resulta ser un factor altamente relacionado con la empleabilidad. Pero pese a que existen numerosas investigaciones, teóricamente resulta complejo definir una relación clara entre dichas variables. Debido a que existen corrientes que plantean una relación inversa debido a las mejoras personales y técnicas que obtienen los individuos conforme pasan los años. Sin embargo, otros estudios presentan una relación directa, ya que conforme las personas envejecen sufren deterioros en algunas de sus capacidades, por lo que pueden resultar menos atractivas para los empleadores.

Con respecto a las investigaciones que plantean un impacto positivo de la edad en la empleabilidad, Feldman (2013) plantea que los empleadores a menudo valoran positivamente la cantidad de años de experiencia que posea un candidato en un campo específico, ya que demuestra que dichas personas han adquirido y aplicado sus habilidades y conocimientos en un ambiente laboral similar al que se requiere contratar. Por lo que, debido a que conforme las personas envejecen, generalmente acumulan más experiencia laboral en sus áreas de trabajo, es posible plantear el impacto positivo de la edad en las posibilidades de emplearse, relacionado con la acumulación de experiencia.

Existen otros factores que se enfocan los beneficios de empleabilidad conforme se envejece, tales como lo que plantea Burt (2000) en relación con la tendencia de las personas de ir construyendo redes profesionales más sólidas a través de los años. Condición que puede ayudar positivamente a encontrar empleos, debido a que en el mercado laboral suelen conseguirse empleos mediante contactos o referencias.

Por otro lado, en lo referente a los planteamientos teóricos relacionados con una afectación negativa de la edad en la empleabilidad. Serneels (2001), plantea que esta relación se hace negativa conforme se envejece, debido al deterioro de las habilidades motrices y cognitivas que sucede con el tiempo. Además de dichos deterioros, el autor plantea que conforme las personas envejecen, estas son más proclives a desarrollar enfermedades y además su edad de jubilación resulta ser más pronta en comparación con los más jóvenes.

Sumado a lo planteado anteriormente, existen investigaciones como la de Burn y Button (2019) que plantean un factor negativo de la edad en relación con la empleabilidad, relacionado no con las características físicas y cognitivas de las personas, sino con los prejuicios y estereotipos que existen en las sociedades en relación de las personas de mayor edad en comparación con los más jóvenes. Específicamente con el pensamiento negativo de sus capacidades de aprender nuevas habilidades, de adaptarse a los cambios tecnológicos o las nuevas tendencias del mercado laboral, así como para mantener altos niveles de energía y productividad.

Sexo

A pesar de que existen estrategias que intentan establecer una igualdad laboral entre hombres y mujeres, autores como Reich, Gordon y Edwards (1973) con su teoría de la dualidad del mercado laboral, establecen la existencia de segmentaciones en el mercado laboral. En el que ciertos empleadores en diversas actividades económicas se inclinan más por emplear personas de un sexo en específico debido a sus características distintivas respecto al sexo opuesto. Situación que causa directamente dificultades para la inserción laboral de hombres o mujeres según sea las actividades que se desarrollan en la economía en estudio.

Sumado a lo anterior, Heilman (2012) plantea la existencia de estereotipos y prejuicios arraigados en la sociedad que provocan la existencia de discriminaciones de género, es decir a tratos desiguales a las personas debido a su género. Según el autor, dicha situación afecta en procesos laborales como durante la selección de candidatos, en los tratos en el lugar de trabajo y en las oportunidades de ascenso. Condiciones que según sea el caso pueden afectar positiva o negativamente algún género en específico en sus probabilidades de empleabilidad.

Con referente específico a las mujeres, teóricamente se plantea un impacto negativo en comparación con los hombres que está relacionado con la maternidad. Dicha relación según Glauber (2007) sucede en la selección de candidatos, ya que los empleadores pueden tener preocupaciones de la disponibilidad de las mujeres que son o serán madres a futuro, por temas como sus necesidades de cuidado infantil o por la existencia de responsabilidades familiares que las obliguen a reducir sus horas de trabajo. Causando así una reducción en las probabilidades de empleabilidad para ese tipo de mujeres, debido a dichas discriminaciones.

Nivel Educativo y dominio de segundo idioma

En cuanto a las variables de educación y dominio de un segundo idioma, la teoría del capital humano de Becker (1975) plantea el concepto de capital humano, el cual se considera como el conjunto de conocimiento, habilidades y competencias adquiridas por una persona, ya sea mediante la educación y/o experiencia laboral. Y dicho capital, es un factor que aumenta el nivel de productividad de cada individuo y por tanto incrementa sus probabilidades de empleabilidad y de aspirar a salarios más altos.

Cabe además destacar que Becker (1993) plantea que, sumado al impacto positivo de las mejoras en el capital humano individual de cada persona, sucede una externalidad positiva para toda la sociedad de dichas mejoras. Ya que un mercado laboral con trabajadores mejor calificados, que puedan aspirar a trabajos mejores remunerados, contribuye positivamente en el desarrollo y bienestar en las economías.

Sin embargo, cabe aclarar que el autor plantea que esta relación posee una tasa marginal decreciente, es decir, cada año de mayor capacitación aporta menos porcentaje de no estar desempleado y además es importante indicar que en dicha teoría se plantea la existencia de otros factores relevantes que complementan esta variable como es el nivel de experiencia o la formación en habilidades que sean correspondientes por las demandadas en el mercado laboral y que sin duda impactan las probabilidades de conseguir empleo.

Como complemento, se considera relevante presentar para el caso del dominio de idiomas, Reich (1991) plantea la Teoría de la Competencia Global, bajo un contexto de economías y de empresas altamente globalizadas, es decir interconectadas con otras alrededor del mundo, los empleadores buscan personas que puedan laborar en contextos globales.

Por tanto, el dominar un segundo idioma que sea hablado en otras partes del mundo, les permite a las personas poder comunicarse, trabajar con equipos internacionales y tomar negociaciones con socios o clientes internacionales. Lo cual, eleva considerablemente su productividad generando una ventaja competitiva y por tanto pudiendo afectar positivamente sus posibilidades de empleabilidad, ya que el poseer dicha habilidad les permite acceder a trabajos de empresas globales.

Zona de residencia

La zona de residencia urbana o rural de las personas está relacionada con sus probabilidades de empleabilidad debido a las diferencias que suelen existir tanto a nivel educativo como socioeconómico en cada zona. Así como por la diversidad que existe en sus mercados laborales, debido al tipo de sectores o industrias que desarrollan sus actividades económicas en cada una, basado principalmente en lo que estas ofrecen y poseen según sus características geográficas.

Winters (2011) plantea las diferencias en acceso a la educación que suelen existir entre la zona rural y la urbana, principalmente por la escasez que existe en las zonas rurales de centros educativos superiores o de formación especializada. Condición que dificulta el acceso a empleos que demanden altas calificaciones para las personas residentes en dicha zona, debido a las dificultades que enfrentan de lograr obtener los estudios requeridos.

Sumado a las diferencias educativas, Van Winkle (2018) plantea que las áreas urbanas tienden a concentrar una mayor cantidad de empresas y de una mayor variedad de sectores e industrias, por lo que en dichas zonas suelen existir más y de mayor diversidad de demandas de empleo en comparación con la zona rural, la cual suelen tener mercados laborales más limitados y que acostumbran a basarse en una serie de sectores específicos, tales como agricultura, ganadería o alguna industria local.

2.4 Pensamiento Computacional

Existen diversas circunstancias o fenómenos que generan obstáculos para el desarrollo de las economías y que son entendidos como problemas económicos por resolver. De manera conceptual, la resolución de problemas se define como “la capacidad de participar en un proceso cognitivo para entender y resolver problemas donde no hay un método de solución inmediatamente obvio” (OECD ,2014).

Por lo que para lograr resolver problemas a nivel científico existen diversos tipos de pensamientos, tales como el Pensamiento Crítico o el Pensamiento Matemático que plantean una serie de procesos a tomar para llegar a una posible solución. Por tanto, dicha investigación con la intención de tener una estructura científicamente clara toma los procesos presentados por el Pensamiento Computacional para la resolución de problemas.

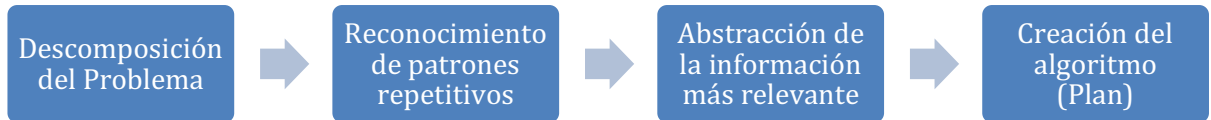
El Pensamiento Computacional fue desarrollado por Wing (2006), quien inspirada en el conjunto de competencias y habilidades que un profesional utiliza cuando crea una aplicación computacional, plantea una metodología de resolución de problemas aplicable a cualquier rama científica.

Específicamente según Parra (2019) el Pensamiento Computacional se enfoca en la aplicación de modelos mentales diseñados para entender problemas a través de los conceptos fundamentales de la informática, por lo que para llevarse a cabo implica la existencia de 4 fases:

- i. **Descomposición:** Consiste en la ruptura de un sistema o problema complejo en partes más pequeñas para que así sean más fáciles de solucionar.
- ii. **Reconocimiento de patrones repetitivos:** Una vez que se ha descompuesto el problema complejo en varios más pequeños, la metodología busca estándares de características comunes.
- iii. **Abstracción de información:** se refiere a centrarse en la información importante, dejando de lado aquellas características irrelevantes e innecesarias. Luego de contar con estas características generales, se debe crear un “modelo” del problema, el cual es la idea general del problema que se intenta resolver.

- iv. **Creación del algoritmo:** un algoritmo se puede entender básicamente como un plan, por lo que el método termina con la existencia de un conjunto de instrucciones que lleven paso a paso a la resolución del problema.

Figura 1. Estructura Pensamiento Computacional



Fuente: Parra (2019)

2.5 Modelos de Aprendizaje Automático

El aprendizaje se define según Sancho (2017) como el proceso a través del cual se adquieren o modifican habilidades, destrezas, conocimientos, conductas o valores como resultado del estudio, la experiencia, la instrucción, el razonamiento y la observación.

Por lo que para este caso el Aprendizaje Automático, según Athey (2018) se refiere a un campo de la inteligencia artificial que les permite a los computadores aprender, es decir adquirir conocimientos a través del análisis de información. Y sus bases parten de la utilización de técnicas estadísticas y optimización matemática mediante la aplicación de algoritmos capaces de generalizar comportamientos y reconocer patrones a partir de los datos suministrados. Por tanto, según su objetivo dichos algoritmos se pueden subdividir en dos categorías: Supervisado y No Supervisado, por lo que a continuación se detallan ambas.

2.5.1 Modelos de Aprendizaje Automático No Supervisado

El Aprendizaje Automático No Supervisado es definido por Athey (2018) como los algoritmos que se enfocan en encontrar grupos de observaciones que son similares en términos de sus covariables (X) llamados típicamente de agrupamiento y en los que para el caso de la investigación se utilizará la siguiente técnica:

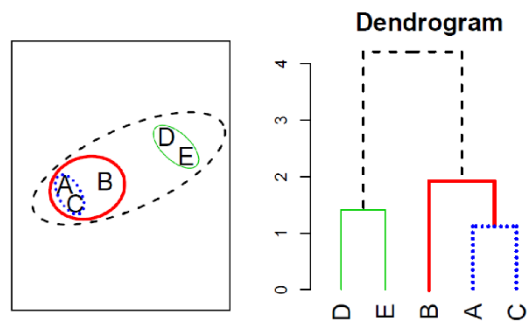
Clustering Jerárquico

Según Gallardo (2020) el clustering jerárquico tiene por objetivo agrupar clústers para formar uno nuevo o bien separar alguno ya existente para dar origen a otros dos, de tal manera que, si sucesivamente se va efectúa este proceso de aglomeración o división, se minimice alguna distancia o bien se maximice alguna medida de similitud. Estos métodos se subdividen en aglomerativos y disociativos, los cuales además tienen un gran conjunto de variantes.

1. Los métodos aglomerativos: se conocen como ascendentes y su metodología consiste en que se comienza el análisis con tantos grupos como individuos haya. A partir de estas unidades iniciales se van formando grupos, de forma ascendente, hasta que al final del proceso todos los casos tratados están englobados en un mismo conglomerado.

2. Los métodos disociativos: se conocen como descendentes y su metodología es en sentido inversa al método anterior. Es decir que se comienza con un conglomerado que engloba a todos los casos tratados y, a partir de este grupo inicial, a través de sucesivas divisiones, se van formando grupos cada vez más pequeños. Al final del proceso se tienen tantas agrupaciones como casos han sido tratados.

Figura 2. Clustering Jerárquico de Tipo Disociativo



Fuente: Rodríguez (2019)

Método de aglomeración

Para el presente estudio se utilizará un método aglomerativo llamado Average que consiste inicialmente calcular la distancia existente entre todos los posibles pares formados por una observación del clúster A y una del clúster B. Posteriormente, tal y como lo presenta Xu, Finkelman, Dai y Peng (2021) se procede a definir el promedio general de dichas distancias entre cada par de clúster y se procede a fusionar los dos clústeres que tengan la menor distancia (D) promedio entre sí, por lo que durante cada iteración la cantidad de grupos se reduce a n-1.

$$D_{(r,s)} = \frac{T_{rs}}{(N_r * N_s)}$$

Donde:

- Suma de las distancias entre el clúster “r” y el clúster “s”.
- Tamaño del clúster “r”.
- Tamaño del clúster “s”.

2.5.2 Modelos de Aprendizaje Automático Supervisado

En segundo lugar, se encuentra el Aprendizaje Automático Supervisado que implica el uso de un conjunto de covariables (X) para predecir un resultado (Y), llamado predicción si la variable Y es continua y clasificación si la variable dependiente es discreta (Athey,2018).

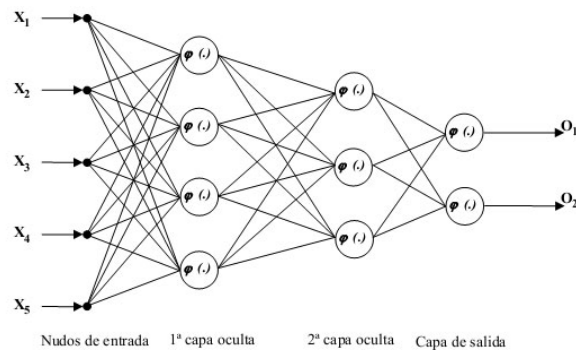
Redes Neuronales Artificiales (RNA)

Según Enzyme (2019) las redes neuronales artificiales (RNA) son modelos computacionales que procesan información mediante la imitación del funcionamiento de las neuronas biológicas. Por tal razón, son redes compuestas por “neuronas” llamadas en este caso nodos, los cuales reciben, transmiten y envían información.

El funcionamiento de estas se basa en la existencia de 3 tipos de nodos: los nodos de entrada que reciben información del exterior de la red, nodos de salida que envían información al exterior y los nodos ocultos que se ubican entre los de salida y entrada y su función es transmitir información entre todos los nodos de la red.

Por lo que las RNA están formadas por múltiples capas de nodos ocultos, llamadas “capas de aprendizaje”, que les permite tener un funcionamiento que consiste en que los nodos de entrada reciben una serie de datos del exterior, los cuales son enviados al interior de la red hacia los nodos ocultos quienes procesan, modifican y transfieren información de una capa a otra; finalizando así con una serie de nodos de salida que brindan un resultado de la predicción.

Figura 3. Funcionamiento Redes Neuronales Artificiales



Fuente: Villada, Muñoz y García 2012

2.5.3 Optimizadores Instrumentos de Aprendizaje Automático Utilizados

2.5.3.1 Optimizadores Modelos de Aprendizaje Automático no Supervisado

Cuando se aplican este tipo de métodos predictivos se hace necesario aplicar índices que estimen la calidad de las estimaciones, como vía para determinar los mejores parámetros posibles entre todos los distintos algoritmos. Por lo que para el presente estudio se utiliza el indicador llamado Silhouette, el cual busca estimar la disimilitud promedio entre todas las observaciones de los grupos creados, según Shutaywi y Kachouie (2021) a través de la siguiente fórmula:

$$S(x_i) = \frac{b(x_i) - a(x_i)}{\max \{b(x_i), a(x_i)\}}$$

Sea:

- $S(x_i)$: El indicador Silhouette para la observación x_i perteneciente al clúster “r”

- $a(x_i)$: Distancia de la observación x_i respecto a las demás observaciones del clúster “r”
- $b(x_i)$: Distancia mínima de la observación x_i respecto a otro clúster diferente a “r”

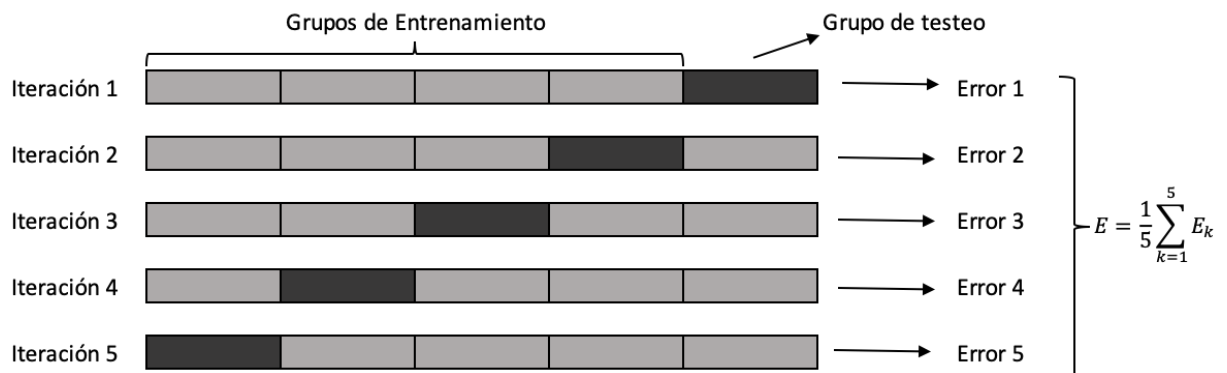
2.5.3.2 Optimizadores Modelos de Aprendizaje Automático Supervisado

Validación Cruzada K-Fold

Según Amat (2016), la validación cruzada K-Fold es un proceso utilizado para la validación de la calidad de los modelos predictivos, en el que a través de un proceso iterativo en el que se divide la muestra de forma aleatoria en k grupos, k-1 grupos se emplean para entrenar el modelo y el sobrante grupo se emplea como validación. Por lo que, este proceso se repite “k” veces utilizando un grupo distinto como validación en cada iteración.

El proceso genera durante cada iteración un error, por lo que al final el error general de cada método de predicción se calcula como el promedio de los errores obtenidos durante las “k” estimaciones. Lo que permite, por tanto, obtener indicadores para dichos modelos altamente estables.

Figura 4. Funcionamiento Validación Cruzada K-Fold



Fuente: Elaboración Propia

2.6 Modelos Econométricos

Ceballos (2019) plantea como una de las ramas más activas de la ciencia económica, aquella que se ocupa de la aplicación de métodos cuantitativos para el análisis del comportamiento de las variables y las relaciones que observamos en el acontecer diario.

Dentro de esta rama económica, se destaca la econometría, definida por Ceballos (2019) como el campo de la estadística matemática y la inferencia estadística aplicada para probar las relaciones empíricas sugeridas por la teoría económica. Área de estudio que es clasificada por su nivel de aplicación, en la microeconometría que se encarga de simular o estudiar situaciones económicas resultado de las decisiones de personas o empresas teniendo en cuenta diferentes variables y la macroeconometría que analiza el comportamiento de datos o variables macroeconómicas.

2.6.1 Modelos Logísticos

Dentro de la microeconometría, se destacan los modelos logísticos en los que sea “Y” una variable dependiente binaria con posibles valores finitos y sean un conjunto “k” de variables independientes, (X, X_2, X_i) , observadas con el fin de predecir/explicar el valor de Y. La herramienta usada más comúnmente para estimar modelos Logit y Probit es el método de máxima verosimilitud y dichos modelos son expresados genéricamente bajo la siguiente ecuación. (Marín, 2014):

$$p_i = F(Z)$$

En el que p es la probabilidad de que ocurra Y, por lo que está entre cero y uno. $F(Z)$ es la función de probabilidad y tiene como características que es continua y lineal en $0 < p_i < 1$.

Para el caso del modelo Logit, su expresión $F(Z)$ es la inversa de la función de distribución logística y se representa de la siguiente manera (Coa,2019):

$$p_i = \frac{1}{1 + e^{-(z)}}$$

Sea:

- p_i = Variable dependiente (valores entre cero y uno)
- $Z = \beta_0 + \beta_i X_i$ = Scoring logístico
 - β_0 = constante
 - $\beta_i X_i$ = parámetros de las i variables independientes

Por último, los modelos Probit utilizan la función inversa de la distribución normal estándar como función $F(X)$ (Coa,2019):

$$p_i = \Phi(x) = (2\pi)^{-1} \int_{-\infty}^x e^{-\frac{z^2}{2}} dz$$

2.6.1.1 Validaciones Modelos Logit

Para el presente estudio, en lo que respecta a las validaciones a los modelos Logit, se aplicará inicialmente la prueba de 2 log likelihood, la cual es también conocida como Test de Razón de verosimilitud y según Hosmer y Lemeshow (2004), su objetivo es comparar los valores observados de la variable de respuesta para predecir los valores obtenidos de los modelos con y sin la variable en cuestión. Y se calcula, por el análisis de la desviación de los errores del modelo nulo y la desviación de los errores con el modelo aplicado.

En segundo lugar, se analiza el indicador de Pseudo R-2 propuesto por Nagelkerke (1991), en el que se propone un ajuste al R-2 de Cox y Snell para setear su valor entre 0 y 1. El cual por su parte se basa en el comparar el logaritmo de la verosimilitud del modelo en con el logaritmo de la verosimilitud de un modelo de línea base, como vía para estimar la cobertura de la predicción a la variable a predecir.

2.6.1.2 Interpretación Modelos Logit: Odds Ratio

Para entender el concepto de Odds Ratio inicialmente hay que definir una Odds, la cual según Benavides (2020) se entiende como la frecuencia con la que sucede un evento y su cálculo se basa en el cociente del número de observaciones que presentan una característica entre el número de observaciones que no la presentan.

Por lo que el Odds Ratio hace referencia al cociente entre dos Odds y este se interpreta como cuantas veces es más frecuente un evento entre los expuestos a una condición específica sobre los que no están expuestos. Siendo un ejemplo ilustrativo el siguiente:

Tabla 1. Ilustración Odds Ratios

Exposición	Empleado	No empleado
Dominio Segundo Idioma (Expuestos)	a	b
Sin Dominio Segundo Idioma (No Expuestos)	c	d

Fuente: Elaboración Propia basado en Benavides (2020)

Sea, por tanto:

$$Odds_{Empleado/Dominio\ Segundo\ Idioma} = \frac{a}{b}$$

$$Odds_{No\ empleado/Sin\ Dominio\ Segundo\ Idioma} = \frac{c}{d}$$

$$Odds\ Ratio = \frac{Odds_{Empleado/Dominio\ Segundo\ Idioma}}{Odds_{No\ empleado/Sin\ Dominio\ Segundo\ Idioma}}$$

La odds ratio al ser un cociente sus valores oscilan entre 0 y el infinito, por lo que su interpretación muestra tres escenarios: valores entre 0 y 1 que indican una relación negativa entre la exposición de la variable analizada y la ocurrencia del evento. En segundo lugar, valores iguales a 1 que hacen referencia que la exposición no tiene ninguna afectación en la ocurrencia del evento, ya que la frecuencia de eventos es igual entre expuestos y no expuestos. Por último, se encuentra el escenario de valores mayores a 1, los cuales indican una relación positiva entre la exposición y la ocurrencia del evento, ya que la frecuencia de suceso de los expuestos es mayor que los no expuestos.

3. Capítulo III: Marco Metodológico

En el presente capítulo, se presentan los detalles metodológicos utilizados en la investigación. Por lo que se presenta su enfoque, sus alcances y la estructura analítica que se seguirá para lograr abordar los objetivos planteados anteriormente. Asimismo, se describe los principales aspectos de la Encuesta Continua de Empleo de Costa Rica, la cual es la fuente de *datos* secundaria utilizada en el estudio.

3.1 Enfoque de la Investigación

Un proyecto de investigación según Hernández (2014) “es un conjunto de procesos sistemáticos, críticos y empíricos que se aplican al estudio de un fenómeno o problema y los procesos metodológicos a implementar varían según el tipo de enfoque que se utilice” (p.4).

Según Hernández (2014), las investigaciones cuantitativas parten de la utilización de datos para probar hipótesis con base en la medición numérica y el análisis estadístico, con el fin de establecer pautas de comportamiento y probar teorías. Mientras que el enfoque cualitativo utiliza la recolección y análisis de datos para afinar las preguntas de investigación o revelar nuevas interrogantes en el proceso de interpretación. Y, por último, el de tipo mixto que se caracteriza por ser una combinación de los dos enfoques mencionados anteriormente.

El enfoque de investigación que se ejecutará en el presente estudio es el de tipo mixto ya que el análisis contempla análisis cuantitativos estadísticos mediante la utilización de las bases de datos del INEC y en añadido se estudian elementos cualitativos relacionados con las políticas públicas dirigidas al mercado laboral.

3.2 Alcances

El alcance de la investigación es correlacional, ya que según Hernández (2014) dicho tipo se emplea para asociar variables mediante un patrón predecible para un grupo o población. Es decir que su utilidad permite determinar cómo se puede comportar un concepto o una variable al conocer el comportamiento de otras variables vinculadas, definición que concuerda con los objetivos de este estudio, donde se busca identificar la relación de diversas variables demográficas con la condición de actividad de la fuerza de trabajo de la región.

3.3 Universo de la investigación

3.3.1 Sujetos y fuentes de información

El universo de estudio de la investigación está conformado por las personas mayores de 15 años residentes habituales de la región de planificación central de Costa Rica, que forman parte de hogares particulares, que habitan en viviendas individuales y que forman parte de la muestra trimestral de la encuesta tomada como fuente secundaria, la cual debido a su tamaño y validez estadística, permite realizar estimaciones generales para toda las personas de dicha región.

Los datos cuantitativos serán recolectados, específicamente de las bases de datos resultantes de la aplicación trimestral de la Encuesta Continua de Empleo (ECE), estudio que recolecta información periódica sobre la situación laboral de las personas y las características del empleo en Costa Rica.

Los datos cualitativos provendrán de fuentes secundarias, específicamente de los reportes realizados por las instituciones públicas que atienden dicho mercado, así como los informes realizados por el Ministerio de Planificación Nacional y Política Económica.

3.3.2 Muestra

La muestra utilizada en el estudio es la definida y aplicada por el Instituto Nacional de Estadísticas y Censos para aplicar la Encuesta Continua de Empleo, la cual al ser de carácter pública, se toman dichos datos posterior a su publicación. Su diseño es de tipo probabilístico, estratificado, bietápico de conglomerados y replicado. Su marco de muestreo inicial se construyó con la base cartográfica e información del Censo Nacional de Población y Vivienda del año 2011, que se denominó Marco Muestral de Viviendas 2011 (MMV-2011)

- **Probabilístico:** a los hogares seleccionados se le asignan una probabilidad conocida de selección, lo cual permite hacer inferencias a la población con base en la teoría de probabilidades.
- **Estratificado:** previamente a la selección, las unidades del marco se clasificaron en estratos o grupos relativamente homogéneos de acuerdo con características geográficas y socioeconómicas.
- **Bietápico:** en una primera etapa se seleccionan Unidades Primarias de Selección (UPM) y en una segunda etapa se seleccionan viviendas dentro de las UPM seleccionadas en la primera etapa.
- **Replicado:** cada trimestre se realiza la rotación del 25% de las viviendas de la muestra, es decir, se recolecta información en el 75% de las viviendas del trimestre previo, y se sustituye el 25% restante.

Dado que el diseño muestral de la ECE está planeado para visitar las mismas viviendas, se les da seguimiento a las viviendas y no a los hogares. Por lo que se trata de un semipanel de viviendas que facilita obtener mejores indicadores sobre los cambios observados en la composición de la fuerza de trabajo.

La muestra seleccionada del Marco Muestral de Viviendas MMV-2011, contiene 794 Unidades Primarias de Muestreo (UPM) y 9528 viviendas y la recolección de dicha muestra se distribuye a lo largo de 12 semanas en el trimestre.

Ponderación

Los factores de expansión se calculan respecto al Marco Muestral de Viviendas y proyecciones de población referidas al Censo Nacional de Población y Vivienda del 2011 (MMV-2011). Además, los factores de expansión de las ECE son trimestrales, es decir que se calculan tanto para los trimestres calendario (cuatro trimestres naturales del año), como para los móviles (unión de tres meses consecutivos, aparte de los trimestres calendario)

El ajuste por ponderación se realiza en tres tipos:

- **Ajuste por no respuesta:** corrige las diferencias en las probabilidades de selección por la no respuesta de las viviendas seleccionadas. El ajuste consiste en darle un mayor peso a las viviendas que respondieron para compensar las viviendas que no respondieron.
- **Ajuste por desactualización:** corrige de manera parcial los problemas de desactualización del marco muestral de viviendas debido a que se efectúa únicamente en las UPM seleccionadas. El ajuste consiste en actualizar las probabilidades de selección de las viviendas entrevistadas según los cambios poblacionales observados dentro de cada UPM de la muestra.
- **Ajuste por proyecciones:** se hace a nivel de región de planificación por sexo y dos grupos de edad: menores de 15 años, y de 15 años y más.

3.4 Técnicas e instrumentos de investigación

Para llevar a cabo la investigación se toma como base el cuestionario aplicado trimestralmente por el Instituto Nacional de Estadísticas y Censos, llamado Encuesta Continua de Empleo (ECE) y que consta de 10 secciones tal y como lo muestra la tabla 2 (INEC,2019b):

Tabla 2. Secciones Encuesta Continua de Empleo INEC

- Identificación del Hogar	- Características Sociodemográficas
- Condición de Actividad	- Personas Ocupadas
- Personas Ocupadas Independientes	- Personas Ocupadas Asalariadas
- Personas con Empleo Secundario	- Personas con Empleo con Insuficiencia de horas y empleo inadecuado
- Personas Desempleadas	- Producción Primaria para Autoconsumo del Hogar

Fuente: Elaboración Propia con datos de INEC 2019b

Metodología de la investigación

La secuencia metodológica de la investigación se basa en el método científico de resolución de problemas llamado Método Computacional, aplicado inicialmente en la informática por Wing (2006) y que presenta cuatro fases: La descomposición, el reconocimiento de patrones repetitivos, abstracción de la información y la creación del algoritmo. Por lo que para el caso en estudio se adapta cada etapa, para el análisis económico de la mano con instrumentos econométricos y de Aprendizaje Automático.

Etapas 1: Evolución de las principales variables relacionadas con la oferta laboral en la Región Central 2013-2022 (Descomposición)

En la etapa inicial de la investigación la intención es mediante la utilización los softwares estadísticos y los microdatos publicados entre el I trimestre del año 2013 y el I trimestre del año 2022, realizar análisis estadísticos exploratorios de las principales variables del mercado laboral y las características demográficas de la oferta de trabajo de la región relacionadas principalmente con el desempleo.

Etapas 2: Determinación de clústeres según las características de las personas desempleadas (Reconocimiento de Patrones)

El propósito de la segunda etapa es aplicar algoritmos de Aprendizaje Automático No Supervisados ejecutados en el software R con el objetivo de realizar agrupaciones de individuos desempleados con características similares según sus covariables.

Etapas 3: Identificación los principales factores demográficos que inciden el desempleo (Abstracción de la información y creación del algoritmo)

En la tercera etapa de la investigación se pretende realizar una estimación de un modelo logístico, con variable dependiente codificada: 0= Desempleado y 1= Empleado y con los microdatos de la Encuesta Continua de Empleo. Siendo el objetivo de la estimación econométrica identificar las variables con mayor influencia estadística en el desempleo en la región.

Con base en el modelo econométrico estimado anteriormente y un modelo basado en algoritmos de Aprendizaje Automático Supervisado de clasificación ejecutado en el software R. Se plantea analizar casos hipotéticos simulados resultantes de la aplicación de políticas públicas, convirtiendo así la herramienta en un laboratorio para medir la posible efectividad de estas, ya que se obtendrían las probabilidades de que una persona con características específicas se inserte al mercado de trabajo, así como el impacto general en el desempleo de una estrategia.

Etapas 4: Recomendaciones de acciones de política (Aplicación del Algoritmo al problema en análisis)

Con la obtención previa de los principales factores determinantes del desempleo en la fuerza de trabajo de la región, la identificación de conglomerados de individuos desempleados con características similares y la simulación de políticas públicas en los modelos predictivos se pretende diseñar una serie de políticas basadas en dichos resultados.

El objetivo del análisis consiste en identificar posibles diferencias entre ambos ejes y de dicha manera formular nuevas estrategias o proponer mejoras en las existentes, que contribuyan en intentar reducir la insuficiencia o ineffectividad de los programas laborales actuales.

3.5 Softwares Estadísticos Utilizados

- **PSPP (Versión 2.0.0):** Programa estadístico de tipo software libre, basado en el lenguaje de programación C y que utilizando la biblioteca científica GNU permite realizar una gran serie de análisis tales como leer fuentes de datos, consolidarlas, calcular frecuencias y tablas cruzadas (GNU, 202). Su uso fue para la lectura y consolidación de los microdatos trimestrales de la Encuesta Continua de Empleo, así como para el análisis descriptivo del contexto histórico laboral de la región durante dicho periodo.
- **R (Versión 4.2.2) y RStudio (Versión 2022.07.02):** Programa y entorno especializado en la computación estadística gratuito de tipo software IDE (Entorno de Desarrollo Integrado) en el que, con base en el uso de librerías y códigos específicos, es posible realizar análisis estadísticos. Su uso para el presente estudio fue para la realización de la Clusterización y la estimación de los modelos logísticos y de aprendizaje automático
 - Librerías descargadas del CRAN (Repositorio de R) utilizadas: Car, Caret, Cluster, Dplyr, Factoextra, FactoMineR, Fmsb, Foreign, Gtools, NbClust, Pscl, Readxl, Rjava, Sparklyr, Splitstackshape, TidyR, TraineR, epiDisplay y Xlsx.

4. Capítulo IV: Resultados

En el presente capítulo se muestran los resultados obtenidos, tras aplicar la metodología de pensamiento computacional para identificar los principales factores determinantes del desempleo en la región. Por lo que en la primera sección se abarca el primer objetivo de describir la evolución histórica de la oferta laboral de la región en el periodo 2013-2022, mediante un análisis estadístico descriptivo de sus principales variables encontradas en la Encuesta Continua de Empleo.

En la siguiente sección llamada “determinación de clústeres según las características de las personas desempleadas”, se exponen los resultados obtenidos relacionados con el segundo objetivo, en el que se analizaron las principales relaciones en el grupo de desempleados en la región. Lo anterior a través de la identificación de 6 clústers de personas con características similares.

Seguidamente, en el apartado titulado “identificación los principales factores demográficos que inciden el desempleo”, se amplían los hallazgos buscados en el tercer objetivo que busca identificar qué factores tienen mayor incidencia en la empleabilidad. Esto con dos estimaciones, inicialmente con modelos logísticos que permiten aproximar el impacto de la tenencia o no de una condición en sus posibilidades de empleabilidad, y en segundo lugar, con un modelo de Aprendizaje Automático que predice las posibilidades de empleabilidad de una persona con base en unas características demográficas dadas. El cual, se utiliza como herramienta para medir el posible de impacto en la empleabilidad de mejoras educativas en grupos específicos de la región.

Por último, en la sección “Recomendaciones de Acciones de Política” se abarca el cuarto objetivo en el que se presenta un plan con una serie de recomendaciones basadas en el análisis anterior, con las cuales se pretende mejorar el desarrollo económico en la región a través de aumentos en la empleabilidad mediante un enfoque de mejora en las calificaciones educativas en los clústers encontrados.

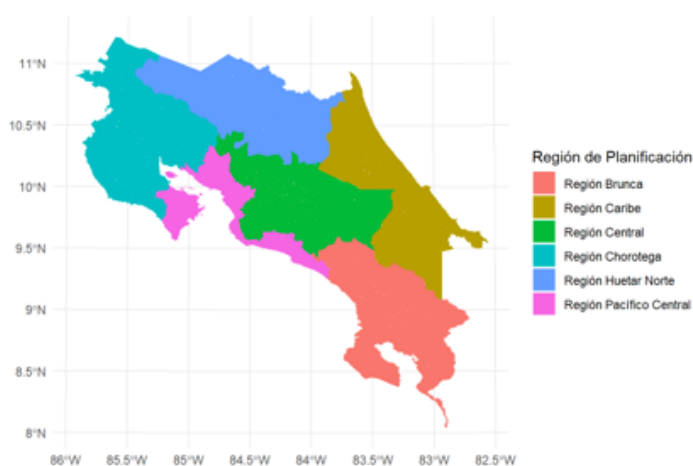
4.1 Evolución de las principales variables relacionadas con la oferta laboral en la Región Central

A continuación, se aborda el desarrollo del primer objetivo del estudio, mediante la presentación de un diagnóstico estadístico descriptivo de la evolución histórica de la oferta laboral en la Región Central de Costa Rica durante el periodo 2013-2022, teniendo en cuenta sus condiciones demográficas, sus principales tendencias de empleabilidad, así como detalle de las tasas de desempleo segregadas por variables como edad, sexo, niveles educativos y zonas de residencia. Con la finalidad de que sirva como un punto de partida en el análisis de dicha problemática laboral en la región y permita tener un panorama más claro para las siguientes secciones del estudio.

4.1.1 Contexto Región Central de Costa Rica

Costa Rica posee una extensión geográfica aproximada de 51 000 metros cuadrados y está dividida geográficamente según el Instituto de Estadísticas y Censos (INEC) en 6 regiones de planificación, tal y como lo muestra la Figura 5. Siendo la Región Central la de mayor área con un 21% de dicho territorio y con una ubicación en la zona central del país, incluyendo así las provincias de Cartago, Heredia (Sin incluir Sarapiquí), San José (Sin incluir Pérez Zeledón) y Alajuela (9 cantones).

Figura 5. Regiones de Planificación de Costa Rica



Fuente: Elaboración propia con datos del SNIT

4.1.1.1 Contexto Demográfico

A continuación, se presentan las principales características demográficas de la oferta laboral de la Región Central, tanto para el año 2022, así como su evolución durante el periodo 2013-2022. Lo anterior con la intención de obtener un primer detalle de las principales características de esta, como vía para determinar los principales factores que estén influyendo en el desempleo.

La Región de Planificación Central pese a que su extensión es menor al 25% del área total, en esta reside más del 60% del total de personas del país. Lo que la convierte en la región con la mayor densidad poblacional con alrededor de 300 hab/km², valor que además es considerablemente superior al promedio nacional de Costa Rica que es de 101 y al de otras regiones como por ejemplo la Pacífico Central, que es la segunda más densa con un total de 111.

El comportamiento de alta acumulación de personas se mantiene en lo que respecta a la fuerza de trabajo, ya que tal y como se muestra en la tabla 3 en la Región Central reside el 66% de la fuerza laboral de Costa Rica para el año 2022. Siendo, por tanto, también la que posee el mayor indicador de trabajadores por km², con un valor de 151, cifra que además es considerablemente superior al promedio nacional que es de 48.

Tabla 3. Estimación Población y Fuerza de Trabajo de Costa Rica por Regiones de Planificación I-2022

Región	Población		Fuerza de Trabajo	
	#	%	#	%
Central	3.205.335	62%	1.606.804	66%
Brunca	468.007	9%	166.982	7%
Huetar Caribe	402.332	8%	138.738	6%
Huetar Norte	309.744	6%	141.172	6%
Pacífico Centra	433.884	8%	191.234	8%
Chorotega	373.241	7%	187.373	8%
Total	5.192.543	100%	2.432.303	100%

Fuente: Elaboración propia con datos del INEC 2022

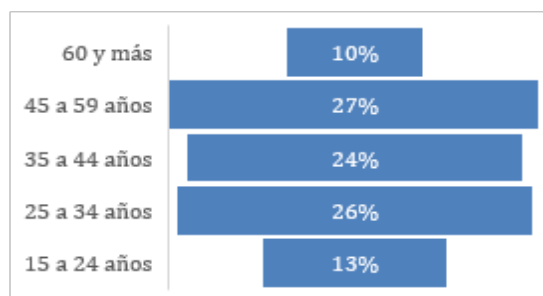
Sexo

En lo relacionado con el sexo, la región presenta a lo largo de los 10 años en análisis una oferta laboral dividida aproximadamente por un 58% hombres y un 42% de mujeres, destacando además un leve crecimiento relativo en las mujeres a lo largo de este periodo de tiempo, ya que para el año 2013 la proporción de estas era cercana al 41%.

Edad

En lo que respecta a la edad, la oferta laboral muestra una tendencia durante el periodo en análisis a envejecerse conforme pasan los años, ya que para el año 2013, aproximadamente el 56% era mayor de 35 años, y para el año 2022, tal y como se observa en la figura 6 esta proporción representa el 61%; situación que arriesga la sostenibilidad productiva en el largo plazo debido al aumento cercano de una alta cantidad de jubilados y la imposibilidad de mantener el sistema de pensiones por parte de los ocupados que contribuirán en el futuro, teniendo en cuenta además que la cantidad relativa de personas jóvenes muestra un comportamiento a la baja.

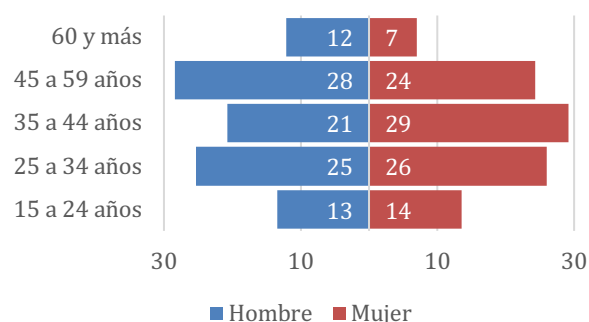
Figura 6. Pirámide Poblacional Región Central de Costa Rica I-2022



Fuente: Elaboración propia con datos del INEC 2022

Por otro lado, en lo que respecta a la situación etaria de la oferta de trabajo de la región por sexo, la figura 7 presentan una situación levemente diferente entre ambos sexos. Y en la que se destaca la mayor proporción relativa de personas mayores de 45 años para el caso de los hombres (41%) en comparación con las mujeres (31%), sumado a una acumulación importante de mujeres entre los 35 y 44 años (29%).

Figura 7. Pirámide Poblacional Oferta de Trabajo Región Central de Costa Rica por sexo I-2022



Fuente: Elaboración Propia con datos de INEC 2022

A nivel histórico por sexo en el periodo 2013-2022, se destaca el descenso relativo del grupo de personas jóvenes (15 a 24 años) en ambos sexos, ya que para el caso de los hombres decayó 4p.p y en mujeres 3p.p. Por otro lado, resulta distintivo el crecimiento relativo de la cantidad de mujeres de 35 a 44 años dispuestas a trabajar, ya que esta proporción aumentó 8 p.p en el periodo analizado pasando del 21% del total de mujeres al 29%.

Zona de Residencia

La oferta de trabajo de la región central de Costa Rica está principalmente concentrada según su residencia en la zona urbana, con un 87% del total. A nivel de según sexo, se destaca que las mujeres tienen una menor proporción en la zona rural (11%) que en la urbana (15%). Y a nivel histórico durante el periodo 2013-2022 estas proporciones han permanecido con valores similares ubicados con variaciones alrededor de +- 2p.p.

Dominio de Segundo Idioma

Con respecto al dominio de un segundo idioma, a nivel general la región muestra una proporción muy reducida con estas habilidades lingüísticas y esta además muestra una tendencia general sin progreso, ya que tanto para el año 2013 y 2022, solamente el 16% de estos reportaban tenerla. A nivel sexo, las mujeres presentan para el 2022 un mayor dominio (17,2%) que los hombres (15,6%) y a nivel histórico por su parte se destaca la mejora en la

proporción para el caso femenino (+2,3 p.p) y por otro lado un deterioro para los masculinos (-1.2 p.p).

Por zona de residencia, la carencia de dominio de un segundo idioma se intensifica en la zona rural con un 93,4%, mientras que la urbana por su parte el 82.2% no posee dicha capacidad. A nivel histórico, la situación para ambas zonas se ha deteriorado en el periodo analizado, siendo la rural la más afectada ya que se ha reducido en 1 p.p y la urbana por su parte se ha disminuido en un 0.6 p.p.

Por grupo de edad tal y como se observa en la tabla 4, la situación muestra un mayor dominio en las personas de los 25 a los 44 años en comparación de los de edades más avanzadas. A nivel histórico, se destaca la mejora para el grupo entre 34 y 44 años, quienes han aumentado en 5 p.p su proporción de dominio y por otro lado los extremos (<25 años y >60 años) quienes al contrario se han deteriorado.

Tabla 4. Dominio Segundo Idioma (%) por Grupo de Edad Fuerza Laboral Región Central I-2013 y I-2022

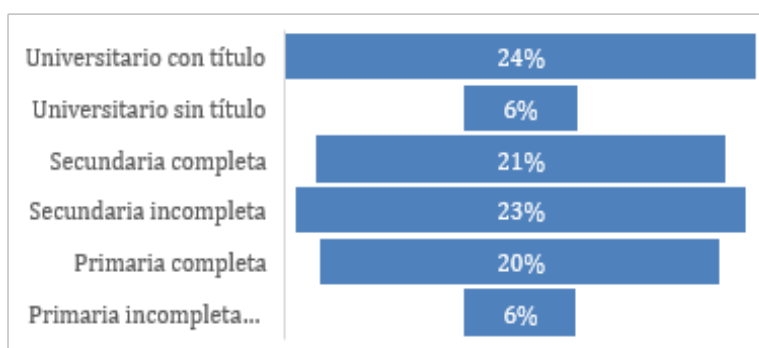
	2013	2022
15 a 24 años	17%	16%
25 a 34 años	23%	23%
35 a 44 años	14%	18%
45 a 59 años	11%	11%
60 años y más	13%	9%

Fuente: Elaboración propia con datos del INEC 2022

Nivel de Educación

En lo referente a nivel educativo, tal como se observa en la figura 8 la región muestra tres grandes grupos, inicialmente un alto porcentaje de personas con secundaria incompleta o menos (49%), seguidamente una importante cantidad con un título secundaria completa o universitario sin título (27%) y por último un menor grupo de oferta más calificada con un título universitario (24%).

Figura 8. Nivel Educativo Fuerza de Trabajo Región Central I-2022



Fuente: Elaboración propia con datos del INEC 2022

A nivel histórico tal y como se observa en la tabla 5 la calificación de la oferta de trabajo de la región ha mejorado su calificación dentro del periodo 2013-2022, ya que tanto la oferta con título universitario y el grupo de personas con secundaria completada han aumentado su proporción.

Tabla 5. Calificación de la oferta de trabajo I-2013 y I-2022

	2013	2022	Var p.p
Universitario con título	21%	24%	+3
Secundaria Completa o Universitario sin título	24%	27%	+3
Secundaria Incompleta o menos	55%	50%	-5

Fuente: Elaboración propia con datos del INEC 2022

Por sexo, tal y como se observa en la tabla 6 la situación resulta ser considerablemente distinta entre ambos sexos ya que las mujeres de la región en estudio poseen un mayor nivel educativo a nivel general. A nivel histórico, además pese a que la proporción universitaria y secundaria para ambos ha aumentado, los hombres son quienes reportan un menor crecimiento.

Tabla 6. Calificación de la oferta de trabajo según el sexo I-2013 y I-2022

	Hombres			Mujeres		
	2013	2022	Var p.p	2013	2022	Var p.p
Universitario con título	17%	19%	+2	26%	31%	+5
Secundaria Completa o Universitario sin título	24%	26%	+2	25%	27%	+2
Secundaria Incompleta o menos	59%	55%	-4	49%	42%	-7

Fuente: Elaboración propia con datos del INEC 2022

El nivel de educación según zona de residencia, tal y como se observa en la tabla 7 existe una importante brecha entre ambos lugares, siendo la rural la menos calificada con menos del 15% de personas con títulos universitarios y más del 60% sin título de secundaria.

A nivel histórico, durante el periodo en análisis ambas zonas han mejorado su calificación. Destacándose, la importante mejora en el nivel universitario en ambas zonas y el incremento en nivel secundaria en la rural. Sin embargo, por otro lado, se observa un muy leve crecimiento relativo de personas con secundaria completa en la zona urbana, situación que evidencia el posible deterioro en dicho nivel educativo en los lugares urbanos.

Tabla 7. Calificación de la oferta de trabajo según zona de residencia I-2013 y I-2022

	Urbana			Rural		
	2013	2022	Var p.p	2013	2022	Var p.p
Universitario con título	22%	26%	+4	11%	14%	+3
Secundaria Completa o Universitario sin título	26%	27%	+1	18%	24%	+6
Secundaria Incompleta o menos	52%	47%	-5	71%	62%	-9

Fuente: Elaboración propia con datos del INEC 2022

4.1.2 Condiciones Laborales en la Región Central

Con respecto a las características laborales de la fuerza de trabajo residente en la región de planificación central de Costa Rica, para el año 2022, el 14% de estas se encuentran desempleadas y dentro de las que tienen empleo un 40% laboran en condiciones de informalidad y un 10% tienen problemas de subempleo (INEC,2022).

Por otro lado, el principal empleador en dicha región para el año 2022 es el sector terciario, abarcando aproximadamente el 74% del total de los trabajadores, las actividades secundarias emplean el 21% y el restante 5% es contratado por el sector primario. Por otro lado, las ramas económicas con mayor cantidad de ocupados son el comercio con el 18% y la industria manufacturera con un 13%.

Durante los últimos 10 años, en lo que respecta al sector estas proporciones se han mantenido similares y en lo relacionado a nivel de rama de actividad empleadora se destaca el descenso relativo del comercio, la enseñanza y las de suministro de electricidad, gas, vapor y aire acondicionado las cuales han decaído más de un 13%. Por otro lado, las actividades de

información y comunicación, profesionales, científicas y técnicas e inmobiliarias son las que más han crecido relativamente con aumentos superiores al 35%. (INEC,2022).

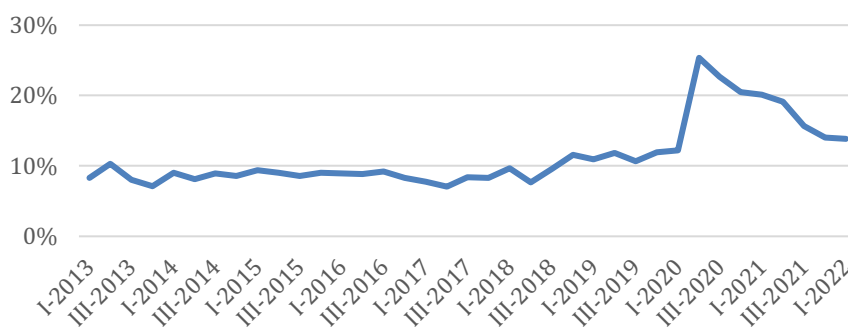
Respecto al sector institucional para el año 2022 donde laboran los ocupados, el 12% lo realiza en el sector público, proporción que ha venido en descenso teniendo en cuenta que en el año 2013 este representaba alrededor del 15%. Así mismo, en lo que respecta a quien emplea a esas personas el 86% pertenecen al gobierno general, el 7% a sociedades financieras y el restante 7% en otro tipo de instituciones. Los trabajadores del sector privado por su parte son contratados por sociedades no financieras (64%), hogares (34%) y sociedades financieras (2%).

En lo relacionado a la provincia donde laboran estas personas corresponde a un 55% en San José, un 19% en Alajuela, en Cartago un 14%, Heredia con un 11% y el restante 1% en otro lugar. A nivel histórico, en el periodo del 2013 al 2022 se destacan los crecimientos relativos de Alajuela y Cartago, las cuales han crecido un 31% y 12% en su proporción relativa respectivamente. Y el descenso de Heredia y San José, provincias que han caído un 20% y 4% respectivamente en valores relativos.

4.1.3 Situación Histórica de las Tasas de Desempleo en la Región Central

La tasa de desempleo en la Región Central de Costa Rica, tal y como lo muestra la Figura 9, se ha ubicado en valores superiores al 7% durante el periodo analizado reflejando así para el I trimestre del año 2022 la carencia de empleo para aproximadamente 220.000 personas para la zona en estudio, lo que equivale a una tasa del 14% (INEC,2022b).

Figura 9. Tasa de Desempleo Región Central Costa Rica (Trimestres I-2013 al I-2022)

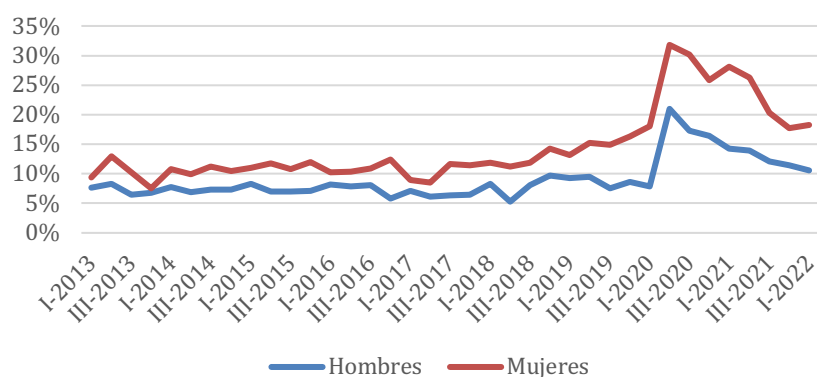


Fuente: Elaboración propia con datos del INEC 2022

Sexo

Para el caso, del desempleo por sexo durante el periodo en estudio se observa la existencia de una brecha entre ambos sexos, siendo las mujeres las que poseen una mayor tasa de desempleo y las cuales se destaca además que durante el periodo anterior a la pandemia por Covid-19 esta brecha era en promedio de 4 p.p y por otro lado para el periodo registrado durante este suceso hasta el año 2022 se identifica un aumento en dicha diferencia con un promedio de 10 p.p, alcanzando valores inclusive de 13p.p y 14p.p durante el año 2020 y 2021. Reflejando, así como el impacto relacionado con dicha crisis sanitaria fue mayor para las mujeres.

Figura 10. Tasa de desempleo Región Central Costa Rica según sexo (Trimestres I-2013 a I-2022)



Fuente: Elaboración propia con datos del INEC 2022

Edad

El desempleo por grupo de edad, tal y como se observa en la tabla 8 la fuerza laboral de menor edad es la más afectada, con valores considerablemente superiores al valor general de la región (14%) e inclusive la situación resulta ser tan profunda para los más jóvenes que inclusive duplican este valor. A nivel histórico, se presenta un deterioro general en todos los grupos etarios y en el que a mayor sea la edad del grupo, menos ha sido su impacto en el periodo analizado.

Tabla 8. Tasa de Desempleo Región Central Costa Rica por grupo de edad I-2013 y I-2022

	2013	2022	Var p.p
15 a 24 años	22,7%	36,0%	+13,3
25 a 34 años	5,8%	15,3%	+9,5
35 a 44 años	5,5%	9,9%	+4,4
45 a 59 años	5,2%	8,1%	+2,9
60 años y más	4,3%	4,9%	+0,6

Fuente: Elaboración propia con datos del INEC 2022

Continuando con el análisis por grupo de edad, la situación para el caso según sexo se presenta con valores superiores para las mujeres en todos los grupos, siendo las edades entre los 35 y 59 años las que presentan una mayor brecha entre ambos sexos.

Tabla 9. Brecha de desempleo Región Central Costa Rica según Sexo I-2022

	Hombres	Mujeres	Brecha p.p
15 a 24 años	33,80%	38,90%	5,1
25 a 34 años	13,40%	17,90%	4,5
35 a 44 años	2,70%	17,00%	14,3
45 a 59 años	5,00%	13,00%	8,0
60 años y más	5,80%	2,60%	3,2

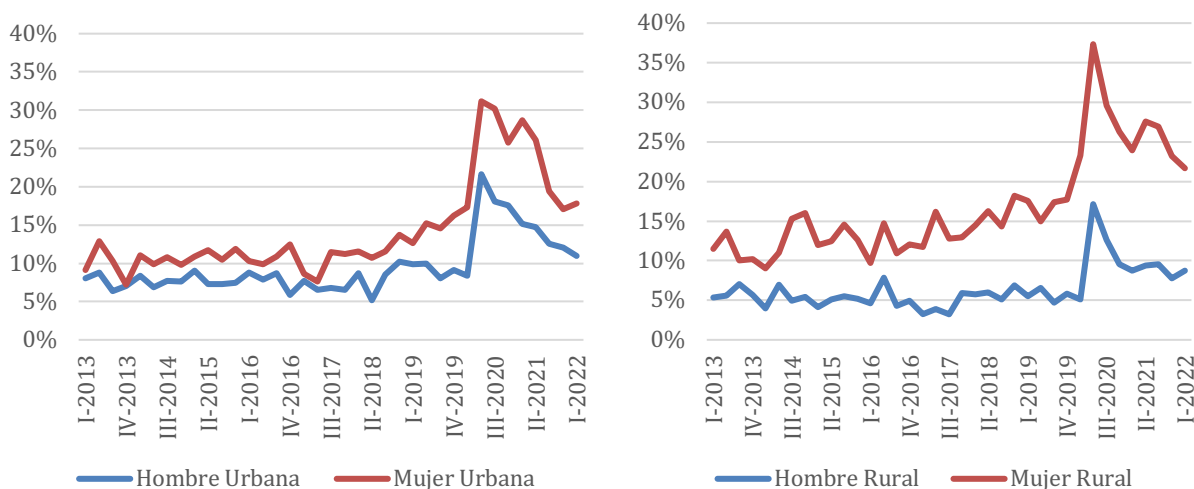
Fuente: Elaboración propia con datos del INEC 2022

Zona

La situación por zona presenta para el año 2022 a la urbana como el lugar con mayor nivel de desempleo con un 13.9% y la rural por su parte muestra una tasa del 13.3%. A nivel histórico, se destaca que pese a que para el año 2013 la urbana continuaba siendo la más afectada con un 1p.p superior, la zona rural es quien más se ha deteriorado en valores relativos durante este periodo ya que incrementó 5.8 p.p. ocasionando así que esta brecha se haya reducido para el 2022 a 0.6 p.p.

En lo relacionado por sexo, tal y como se observa en la figura 11, se mantiene el comportamiento de un mayor nivel de desempleo para las mujeres en ambas zonas. Siendo destacado además la existencia de una mayor brecha entre ambos sexos en la zona rural, ya que en ciertos periodos analizados llegó a ser hasta de 20p.p.

Figura 11. Tasa de desempleo Región Central Costa Rica por zona y sexo I-2013 y I-2022



Fuente: Elaboración propia con datos del INEC 2022

Nivel Educativo

Tal y como se observa en la tabla 10, la situación de desempleabilidad según nivel educativo muestra dos comportamientos destacables, inicialmente como los de altamente educados son quienes son menos afectados en desocupación y en segundo lugar como los de educación media y baja son quienes más se han deteriorado en el periodo analizado, ya que su tasa se ha aumentado considerablemente.

Tabla 10. Tasa de Desempleo Región Central Costa Rica según Nivel Educativo I-2013 y I-2022

	2013	2022	Var p.p
Universitario con título	6%	7%	+1p.p
Secundaria completa o Universitario sin título	9%	19%	+10 p.p
Secundaria incompleta o menos	9%	14%	+5p.p

Fuente: Elaboración propia con datos del INEC 2022

Dominio de Idioma

Tal y como se observa en la tabla 11, las personas sin dominio de un segundo idioma son las más afectadas a nivel de desempleabilidad en comparación con las que sí lo dominan. Siendo además destacable, como este grupo ha sido el más deteriorado durante los 10 años de análisis, teniendo en cuenta que su tasa de desempleo aumentó 5.7 p.p y por otro lado para los que sí poseen dicha capacidad la cantidad de desocupados aumentó en 4.5 p.p.

Tabla 11. Tasa de Desempleo Región Central Costa Rica según Dominio Segundo Idioma I-2013 y I-2022

	2013	2022
Dominio Segundo Idioma	7%	11%
Sin Dominio Segundo idioma	9%	14%

Fuente: Elaboración propia con datos del INEC 2022

4.1.4 Caracterización de la población desempleada

A continuación, se describen las características demográficas de los desempleados para el año 2022, como vía para determinar la participación de según la cantidad de personas que cada uno de estos grupos representa del total de personas sin empleo. Y como vía para ponderar el impacto que posee cada grupo, a nivel regional según sea su nivel de desempleo y su tamaño relativo.

Índice de Desempleabilidad = Tasa de Desempleo (%) * Tamaño relativo Grupo (%)

Tal y como se observa en la tabla 12, con base en el índice estimado la situación de desempleabilidad para ambos sexos es muy diferente, ya que para algunos grupos de variables la desempleabilidad se agrava para un sexo en comparación con el otro, resultados que concuerdan con la teoría en la que se plantea las diferencias existentes según esta variable.

Por ejemplo, para los hombres son más afectados los más jóvenes y los que dominan un segundo idioma. Por otro lado, las mujeres se muestran más afectadas, independiente de la zona del país, en las personas mayores a 24 años, las de niveles de educación medio-bajo

y entre la fuerza laboral que no domina un segundo idioma. Los resultados específicos y los datos utilizados para el cálculo se encuentran en los anexos 1, 2 y 3.

Tabla 12. Índices de Desempleabilidad Región Central Costa Rica I-2022

	General	Hombres	Mujeres
Zona Urbana	<u>12%</u>	<u>10%</u>	<u>16%</u>
Zona Rural	2%	1%	3%
15 a 24 años	<u>13%</u>	<u>14%</u>	<u>11%</u>
25 a 34 años	4%	4%	<u>5%</u>
35 a 44 años	2%	0%	<u>5%</u>
45 a 59 años	1%	1%	2%
60 años y más	0%	0%	0%
Universitario con título	1%	1%	1%
Secundaria Completo o Universitario sin título	7%	7%	<u>8%</u>
Secundaria Incompleta o menos	7%	4%	<u>13%</u>
Dominio Segundo Idioma	1%	2%	1%
Sin Dominio Segundo Idioma	<u>12%</u>	<u>9%</u>	<u>18%</u>

Fuente: Elaboración propia con datos del INEC 2022

Se destaca, además las diferencias existentes por zona en la que, por su gran tamaño, en términos de ponderación la zona urbana resulta ser la de mayores índices. Además, se destaca las diferencias que existen por nivel educativo en la que se muestra el comportamiento sugerido teóricamente en el que a mayor educación o con mayores habilidades lingüística los niveles de desempleabilidad tienden a reducirse.

4.2 Determinación de clústeres según las características de las personas desempleadas

A continuación, se aborda el segundo objetivo del estudio en el que se presentan los grupos de personas desempleadas con características semejantes llamados “Clústers”, los cuales fueron identificados a través de la aplicación de algoritmos de aprendizaje no supervisado y los reportes del I trimestre del año 2013 y I trimestre del año 2022 de la ECE. Con la finalidad de encontrar patrones históricos que permitan orientar las principales causas en el desempleo según el tamaño de estos grupos y a que su vez orienten a crear acciones claras hacia la mejora de la empleabilidad de estos clústers.

Así mismo, mediante la estimación de modelo logit y sus respectivos Odds-Ratios, se presenta el primer hallazgo del tercer objetivo de la investigación que busca encontrar que factores inciden en mayor medida en la empleabilidad de la región, y demostrar el impacto o no de las condiciones educativas, sociales y demográficas en las probabilidades de estar empleado.

Cabe destacar que se probaron ocho distintos métodos de clusterización disponibles en el software R para calcular la n cantidad de grupos (Entre 1 y 8 grupos posibles) óptima y en el que se estima la calidad de la estimación a través del indicador Silhouette y en la que se escoge la predicción que muestre un mayor valor de esta medida.

Tabla 13. Parámetros Optimizados Aplicación de Clusterización Desempleados Región Central de Costa Rica I-2013 y I-2022

Periodo	Método de Aglomeración	Cantidad de Clústers
I-2013	Average	6
I-2022	Average	6

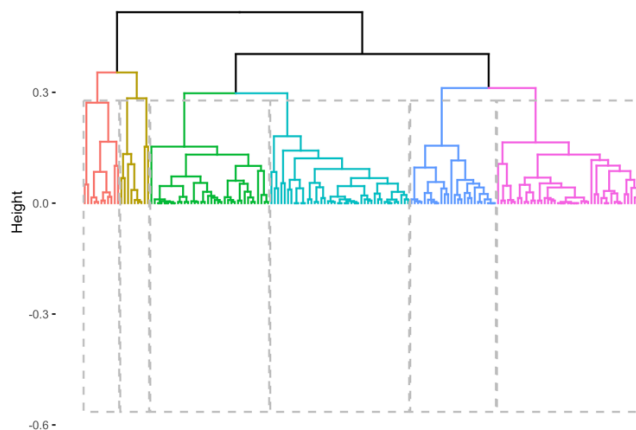
Fuente: Elaboración propia

4.2.1 Resultados de Clusterización por Período de Análisis

Periodo I trimestre 2013

Tal y como se observa, en la tabla 14 y en la figura 12 se detectaron 2 grandes clústeres que entre ellos forman el 76% del total de desempleados para la región en dicho periodo, los cuales según dicha metodología en su momento eran los más destacados. Siendo estos, el primer grupo de hombres y el segundo mujeres, y ambos casos son urbanos, con un nivel educativo reducido, dominio de segundo idioma nulo y de edad principalmente joven o media.

Figura 12: Resultados Gráficos Clusterización Desempleados Región Central I-2013



Fuente: Elaboración Propia

Además, se identifican otros dos clústeres de personas con altos niveles educativos y dominios de segundo idioma, siendo el primero conformado por hombres (7%) los cuales dentro de sus principales especialidades se destacan: Arquitectura, Artes Gráficas y Administración. En segundo lugar, un grupo de mujeres (5%) con especialidades de Ciencias Sociales, Mercadeo, Secretariado, Ciencias de Alimentos y Medicina, los cuales es relevante poner especial atención el entender por qué si teóricamente sus posibilidades de empleabilidad son altas, investigar más a detalle a para identificar si existen otros factores como por ejemplo que su especialización puede ser de áreas con baja demanda laboral en la región o si en algunos casos es requerido para ingresar al mercado el tener un título universitario concluido, ya que muchos de estos se reportan como aún en proceso de obtenerlo.

Por último, se destacan dos clústeres conformados por hombres en el primero y mujeres en el segundo que únicamente son de la zona rural, con bajos niveles educativos, sin dominio de un segundo idioma y de edades variadas. Grupos, los cuales merecen una especial atención en temas de acciones de políticas.

Tabla 14. Resultados Clusterización Desempleados Región Central de Costa Rica I-2013

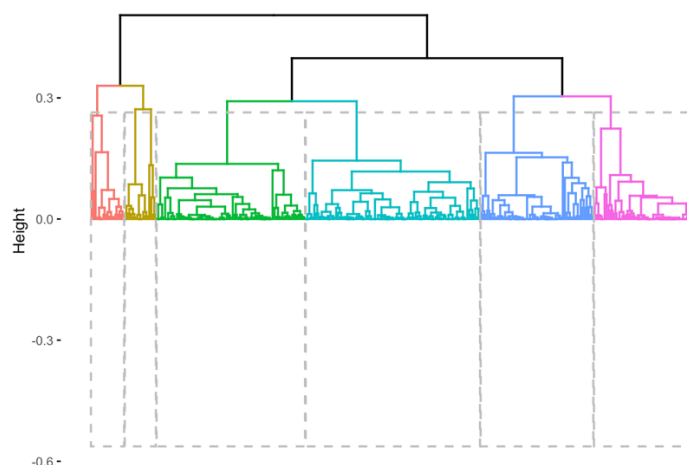
Clúster	%	Sexo	Zona	Dominio Segundo Idioma	Nivel de Educación	Edad (Años)
1	41%	Hombres	Urbana	No	Bajo/Medio	45%: 15 a 24, 19%: 35 a 44 y 18%: 45 a 59
2	35%	Mujeres	Urbana	No	Bajo/Medio	57%: 15 a 24 y 16%: 25 a 34
3	7%	Hombres	Urbana	Si	Medio/Alto	50%: 15 a 24 y 32%: 25 a 34
4	6%	Mujeres	Rural	No	Bajo/Medio	40%: 15 a 24, 24%: 45 a 59 y 22%: 15 a 24
5	5%	Mujeres	Urbana	Si	Medio/Alto	44%: 25 a 34, 28%: 45 a 59 y 20%: 15 a 24
6	5%	Hombres	Rural	No	Bajo	28%: 25 a 34, 25%: 15 a 24 y 25%: 45 a 59

Fuente: Elaboración propia

Periodo I trimestre 2022

Tal y como se observa, en la tabla 15 y en la figura 13 el algoritmo detecta para el I trimestre del 2022 de igual manera seis clústeres. Y en el que destacan en similitud con el anterior periodo la existencia de 2 grandes grupos, que para este caso abarcan el 75% del total de desempleados. Siendo mujeres y hombres de lugares urbanos con niveles educativos bajos, sin dominio de un segundo idioma y edades jóvenes o media principalmente.

Figura 13: Resultados Gráficos Clusterización Desempleados Región Central I-2022



Fuente: Elaboración Propia

Además, se identifican al igual que el periodo anterior dos grupos de personas con altos niveles educativos y dominios de segundo idioma, siendo la principal diferencia un cambio en las especializaciones de estudio. En el que para el caso de los hombres (7%), se mantienen Ciencias Sociales y se añaden Administración, Educación, Humanidades y Artes

Aplicadas . Asimismo, para el caso del grupo conformado por mujeres (5%), se mantienen Ciencias Sociales y se añaden Educación, Humanidades y Artes Aplicadas, Arquitectura y Hotelería.

Por último, continuando con la similitud, se destacan dos clústers conformados por mujeres en el primero y hombres en el segundo que son únicamente de la zona rural, con bajos niveles educativos, sin dominio de un segundo idioma y de edades variadas.

Tabla 15. Resultados Clusterización Desempleados Región Central de Costa Rica I-2022

Clúster	%	Sexo	Zona	Dominio Segundo Idioma	Nivel de Educación	Edad (Años)
1	43%	Mujeres	Urbana	No	Bajo/Medio	28%: 15 a 24, 28%: 35 a 44 y 23%: 25 a 34
2	32%	Hombres	Urbana	No	Bajo/Medio	43%: 15 a 24, 26%: 25 a 34 y 19%: 45 a 59
3	8%	Hombres	Urbana	Si	Medio/Alto	64%: 25 a 34 y 29%: 15 a 24
4	7%	Mujeres	Urbana	No	Bajo/Medio	40%: 15 a 24, 28%: 25 a 34 y 20%: 35 a 44
5	5%	Mujeres	Urbana	Si	Medio/Alto	40%: 25 a 34, 33%: 35 a 44 y 23%: 15 a 24
6	5%	Hombres	Rural	No	Bajo/Medio	58%: 15 a 24, 19%: 25 a 34 y 12%: 60 y más

Fuente: Elaboración propia

Evolución histórica: Clústers 2013 con respecto a Clústers 2022

Los 6 clústeres detectados por el algoritmo de Aprendizaje Automático No Supervisado aplicado en el I-2013 se mantienen para el I-2022, siendo los principales cambios detectados en las edades de estos y una variación en las primeras posiciones, ya que las mujeres urbanas con bajos niveles de educación han aumentado su proporción relativa, tomando el primer lugar en el I-2022 gracias al incremento de 8 p.p. y los hombres urbanos con baja educación por su parte cayeron al segundo puesto debido a su reducción de 9 p.p.

Se destaca para las mujeres urbanas con baja educación además algunos cambios como en su edad, ya que pasaron de ser principalmente muy jóvenes (15 a 24 años) a ser más distribuidas en edades medias (24 a 44 años), lo que refleja que este mismo grupo mantiene su condición solamente que sus edades han aumentado y a este se le suman también las nuevas jóvenes que introducen al mercado con las mismas deficiencias.

Situación que destaca la importancia de atender con políticas de empleabilidad a dicha generación de mujeres residentes en zonas urbanas, de edades jóvenes/medias, con baja educación y sin dominio de idioma antes de que envejecan aún más y resulta más compleja mejorar sus calificaciones y así facilitar su inserción laboral.

Para los hombres urbanos con baja educación, si bien estos mostraron una reducción en su proporción relativa, con un descenso de 9 p.p., en ellos se destaca el crecimiento de personas entre 25 a 34 años en este grupo, teniendo en cuenta que para el año 2013 aparecían como minoritarios y para el 2022 representaron el 26% de este grupo. Lo que muestra además su posible envejecimiento, sumando a que además empieza a tomar relevancia en dicho clúster personas de edad avanzada, lo cual revela la importancia al igual que el grupo de mujeres poner especial interés en dicha generación con políticas que logren mejorar su condición de empleabilidad.

En los grupos minoritarios sus proporciones se mantuvieron considerablemente estables. Destacándose para el caso de los hombres urbanos con alta nivel de educación un aumento relativo de los de edades medias (25 a 34) ya que crecieron 32 p.p. y una caída en la proporción de los más jóvenes, los cuales cayeron 21 p.p. Las mujeres rurales con baja educación destacan por su incremento en la cantidad de personas entre los 35 a 44 años pertenecientes a este grupo, situación que al igual que los grupos mayoritarios concuerdan con el aumento en la edad de la oferta laboral.

Por último, los hombres rurales de baja educación muestran un aumento considerable de personas muy jóvenes (15 a 24 años) en ese grupo ya que crecieron 23 p.p. y además la aparición de un grupo que permanecía minoritario en el 2013 el cual son las de 60 años y más, las cuales representan para el 2022 el 12%, situación a tener en consideración para el diseño de políticas, y brindar alternativas para que puedan ingresar o permanecer en el mercado laboral, teniendo en cuenta que son los extremos en edades y que posiblemente por su falta de experiencia o vejez, se les dificulta su inserción laboral.

En síntesis, la modelación realizada mediante la clusterización, en comparativa con los años 2013 y 2022, muestra resultados muy similares que se excepcionan con los aumentos de edades que esta población presenta y cambios en las especializaciones universitarias con más incidencia de desempleo. Es decir, los residentes en dicha región presentan las mismas deficiencias durante todo el periodo analizado en los niveles de escolaridad, en el aprendizaje cultural de un segundo idioma para algunos grupos y en otro lado un grupo que a pesar de estar alto-medianamente calificados no logran emplearse.

Evidenciando un rezago en la atención de acciones de política y/o soluciones para esta población.

4.3 Identificación los principales factores demográficos que inciden el desempleo

4.3.1 Modelos Logísticos

A continuación, se presentan los resultados de identificar los principales factores demográficos que inciden en la condición de actividad “ocupado” en la región estudiada. Lo anterior mediante modelos logísticos para el primer trimestre de los años 2013 y 2022, con base en los microdatos de la Encuesta Continua de Empleo, filtrando solamente la región central y personas en la fuerza de trabajo (Empleadas o Desempleadas).

Siendo las siguientes variables utilizadas, las cuales se recodificaron para convertirlas en dicotómicas, donde se indica 1 cuando cumple la condición y 0 cuando no la cumple:

- P (E) = Probabilidad de Empleabilidad (1 Empleado, 0 Desempleado).
- Edad > 25 años = 1 (Edad mayor a 25 años), 0 (Edad menor a 25 años).
- Sexo = 1 Hombre, 0 Mujer.
- Zona = 1 Rural, 0 Urbana.
- Idioma = 1 Dominio segundo idioma, 0 Sin dominio segundo idioma.
- Calificación = Nivel educativo (Variable ajustada según la estimación específica).
 - Calificación 1 = Primaria completa o superior (1) e inferior (0).
 - Calificación 2 = Secundaria completa o superior (1) e inferior (0).
 - Calificación 3 = Universitario con título (1) e inferior (0).

Siendo el modelo logístico aplicado a través del siguiente comando en el software R:

```
modelo_logit = glm(Condicion_Actividad ~ "Variable Analizadas", data  
= ECE, family = "binomial")
```

4.3.1.1 Validación Modelos Logísticos Aplicados

Los resultados generales de la estimación pueden ser revisados en el anexo 5, a continuación, en la tabla 16 y 17 se presentan los resultados de validación de las pruebas: Prueba por residuos (2 log likelihood-2LL) y Pseudo R2.

Tabla 16: Prueba por Residuos (2 log likelihood - 2LL) a los modelos logísticos

	2013		2022	
	Desviación nula del Modelo	Desviación del Modelo	Desviación nula del Modelo	Desviación del Modelo
Muestra Completa	831 102	768 045	1 290 732	1 172 687
Hombres	454 358	429 472	629 579	565 173
Mujeres	375 299	330 887	642 379	589 898
Zona Urbana	725 973	663 766	1 123 546	1 026 146
Zona Rural	104 919	99 857	167 134	140 547

Fuente: Elaboración Propia con datos del ECE-INEC (2013-2022)

Tal y como se observa en la tabla 17, todos los indicadores resultan ser favorables, ya que se identifica una reducción entre la desviación nula y la desviación del modelo.

Tabla 17: Pruebas Pseudo R-2 (Nagelkerke) a los modelos logísticos

	2013	2022
Muestra Completa	0,10	0,13
Hombres	0,07	0,14
Mujeres	0,15	0,12
Zona Urbana	0,11	0,12
Zona Rural	0,06	0,22

Signif. Codes: $p > 0.05^*$

Fuente: Elaboración Propia con datos del ECE-INEC (2013-2022)

4.3.1.2 Odds-Ratios de los Modelos Logísticos Aplicados

4.3.1.2.1 Modelos total muestra

Tal y como se observa en la tabla 18, los jóvenes muestran una clara desventaja en comparación con los de mayor edad, teniendo en cuenta que el ser mayor de 25 años incrementa las posibilidades de obtener un empleo más de 4 veces. Siendo destacable una ligera reducción en dicha ratio en el periodo analizado, lo que evidencia un cambio específicamente para este caso en una afectación para los de mayor edad, ya que las tasas de desempleo para personas mayores de 25 años en algunos grupos aumentaron.

La variable Sexo, cumple a su vez con la teoría planteada y con lo observado en el análisis descriptivo, teniendo en cuenta que los hombres tienen más probabilidad de emplearse que las mujeres según los modelos estimados. Situación que además resulta agravarse para las mujeres entre el año 2013 y 2022, ya que dicha probabilidad pasó de 1.27 a 2.18 veces.

Por otro lado, el tema de la zona de residencia muestra una leve ventaja para la zona rural en ambos años, con una mayor diferencia observada en 2013. El dominio de segundo idioma muestra una ventaja importante con respecto a la obtención de empleo para el año 2013. Sin embargo, dicha ventaja no se observa tan claro en el 2022.

La educación por su parte muestra la mayor diferencia en la calificación universitaria y con comportamiento hacia el alza positivamente, ya que para el año 2022 el que posee un título universitario tiene más del doble de probabilidades (2.43 veces) de emplearse que los que no lo tienen y además se destaca la influencia observada en dicho año para la tenencia de primaria que impacta en 1.61 veces con respecto a quien no la poseen.

Tabla 18. Odds-Ratios Modelos Logísticos Fuerza Laboral Región Central de Costa Rica I-2013 y I-2022

Variable	Odds Ratios (Veces)	
	2013	2022
Edad>25 años	4,67	4.18
Sexo (Hombre)	1.27	2,18
Zona (Rural)	1.27	1,16
Segundo Idioma (Dominio)	1,31	1,02
Calificación (1 =Primaria completa o superior)	1,05	1,61
Calificación (1 =Secundaria completa o superior)	1,16	-
Calificación (1 =Universitaria completa)	1,14	2,43

Nota: valores = "-" significa que los resultados no fueron estadísticamente significativos.

Fuente: Elaboración propia con datos de ECE INEC 2013 y 2022.

4.3.1.2.2 Modelos separados por sexo

En segundo lugar, se estimaron modelos logísticos según la variable Sexo como vía para identificar diferencias entre dichos grupos. Los resultados, tal y como se muestran en la tabla 19, por zona los hombres muestran una mayor ventaja en la zona rural que en la urbana y las mujeres la situación resulta completamente contraria, con mayores posibilidades para las residentes en zona urbana.

En lo que respecta a la edad ambos casos muestran problemas para los más jóvenes, destacándose el deterioro histórico en los hombres en los que los menores de 25 años poseen menos posibilidades durante el año 2022 en comparación del 2013. Las mujeres jóvenes por su parte muestran una mejora durante el periodo analizado, teniendo en cuentas que las posibilidades para estas en comparación con los de mayor edad aumentaron.

Con respecto al dominio de un segundo idioma, el cambio se da en mayor medida a las mujeres y muestra un comportamiento al alza para ellas. Los hombres, por su parte, en el 2013 mostraban una influencia positiva, sin embargo, atípicamente en comparación con lo que dicta la teoría para el año 2022 dicha habilidad no muestra influencia en su empleabilidad. Situación que calza con el incremento en nivel de desempleo de dicho grupo de hombres con dicha habilidad para dicho periodo.

En lo relacionado al nivel de calificación, el impacto de mayor educación resulta ser mayor en las mujeres en todos los niveles. Siendo en el que hay mayor distancia en el universitario, donde para las mujeres el poseer este grado les duplica sus posibilidades de empleabilidad en comparación con los hombres con dicha calificación.

Tabla 19. ODDS Ratios Modelos Logísticos por Sexo Fuerza Laboral Región Central de Costa Rica I-2013 y I-2022

Variables	Odds Ratios (Veces)			
	Hombres		Mujeres	
	2013	2022	2013	2022
Edad>25 años	3,76	5,89	6.36	3.26
Zona (Rural)	1.64	1.41	0.92	0.94
Idioma (Dominio)	1,51	0,76	1,14	1,42
Calificación (1 =Primaria completa o superior)	0,56	1,92	1,69	1,45
Calificación (1 =Secundaria completa o superior)	1,05	0,72	1,37	1,32
Calificación (1 =Universitaria completa)	0,80	1,62	1,55	2,85

Nota: valores = “-” significa que los resultados no fueron estadísticamente significativos.

Fuente: Elaboración propia con datos de ECE INEC 2013 y 2022.

4.3.1.2.3 Modelos separados por zona de residencia

En tercer lugar, se realizaron estimaciones logísticas dividiendo la muestra por zona de residencia. Los resultados, tal y como se observa en la tabla 20 la situación por sexo en ambas zonas las mujeres tienen menos posibilidades de emplearse, destacándose además el deterioro en ambas zonas y la diferencia superior que existe para el caso rural. Situación que cumpla teóricamente con los supuestos de que existen diferencias en la empleabilidad tanto por sexo como por zona.

El dominio de un segundo idioma beneficia positivamente ambas zonas, siendo la rural la que la tenencia de esta habilidad presenta un mejoramiento levemente superior en la posibilidad de empleo en comparación con la urbana. En lo que respecta a la edad, en ambos casos los jóvenes se encuentran en desventaja, siendo destacable un comportamiento de mejora en la urbana y un deterioro en la rural para estos grupos etarios.

La educación por su parte para la zona urbana afecta positivamente en todos los niveles. Sin embargo, se destaca la leve baja en la tenencia de secundaria y el comportamiento y el alza del nivel universitario. Por otro lado, la zona rural, muestra un

deterioro en el periodo analizado en dicho factor, en el que inclusive los niveles de secundaria e inferiores muestran ratios inferiores a 1. El nivel universitario para dicha zona, por su parte si bien se ha deteriorado, aún se mantiene como un factor positivo relevante.

Tabla 20. ODDS Ratios Modelos Logísticos por Zona Residencia Fuerza Laboral Región Central de Costa Rica I-2013 y I-2022

Variables	Odds Ratios (Veces)			
	Urbana		Rural	
	2013	2022	2013	2022
Edad>25 años	5.26	4.06	2.01	4,83
Sexo (Hombre)	1.16	2,06	2.45	3,22
Idioma (Dominio)	1,35	0,93	-	1,71
Calificación (1 =Primaria completa o superior)	-	1,78	1,24	1,09
Calificación (1 =Secundaria completa o superior)	1,19	1,13	0,90	0,46
Calificación (1 =Universitaria completa)	1.04	2,35	3,34	3,07

Nota: valores = "-" significa que los resultados no fueron estadísticamente significativos.

Fuente: Elaboración propia con datos de ECE INEC 2013 y 2022.

4.3.2 Modelos de Aprendizaje Automático Supervisado: explicación de la metodología aplicada

Etapa 1: Calibración de algoritmo

La primera etapa consiste en identificar el algoritmo que genera mejores predicciones. Dicha calibración, se basa inicialmente en optimizar los parámetros de cada modelo a través de iteraciones en las que se corren decenas de modelos con distintas configuraciones y en los que al final se determina la combinación modelo/parámetros que brinda el mínimo error.

Posteriormente, dichos modelos se someten a una validación cruzada en la que se logra analizar la estabilidad de sus predicciones, mediante distintas iteraciones con bases de datos cambiantes. Por tanto, el modelo escogido como el mejor predictor correspondió a las redes neuronales, el cual obtuvo un mayor indicador de predicción posteriormente a con una

predicción acertada del 75%, valor aceptable, considerando la alta desproporción de la muestra entre ocupados y desocupados, lo cual dificulta altamente obtener unas mejores estimaciones.

Etapas 2: Análisis de escenarios

El criterio base para esta sección se basa en la simulación de mejoras en los niveles educativos de la población desempleada de la región y su posible impacto en la tasa de desempleo. Con el objetivo de determinar cuál enfoque resulta más prioritario (Mejora en idiomas o mejora en el nivel escolar) para cada uno de los clústers identificados.

Lo anterior tomando como supuestos que existe una demanda laboral exógena capaz absorber a los nuevos empleados y en segundo lugar que los ocupados mantienen su condición constante. Siguiendo, por tanto, con la prueba de 3 escenarios posibles, es decir se intenta predecir el impacto en el nivel desempleo si toda la población desempleada tuviera un dominio del 100% de un segundo idioma (escenario 1), o si la población desempleada tuviera secundaria completa o superior (escenario 2); o la combinación de ambos escenarios (escenario 3).

Tabla 21. Resultados de Predicciones de Tasas de Desempleo según Escenarios de Mejoras en Niveles Educativos para Región Central de Costa Rica, I-2022.

	Tasas de Desempleo			
	Original	Simulación 1: 100% Dominio Segundo Idioma	Simulación 2: Secundaria Completa o Superior	Simulación 3: Dominio Segundo Idioma y Secundaria Completa o Superior
Muestra Completa	14%	8%	9%	6%
Mujeres	18%	10%	13%	7%
Hombres	11%	6%	6%	6%
Zona Urbana	14%	8%	9%	7%
Zona Rural	13%	3%	11%	1%
25 años o menos	33%	18%	32%	19%
35 años o menos	22%	11%	17%	12%
36 años o más	8%	5%	3%	2%

Fuente: Elaboración Propia

Con la utilización del modelo predictivo calibrado anteriormente, se procede a estimar escenarios ficticios, en los que se someten las muestras de la fuerza de trabajo de la región a cambios en sus variables, como medidas de predecir el posible efecto de la aplicación de diversas políticas públicas enfocadas, para este caso, en mejorar la calificación de dichas personas, tanto en el dominio de un segundo idioma como a nivel de estudios de secundaria.

Tal y como se observa en la tabla 21 la población completa, las mujeres, la zona rural y las personas menores de 36 años reflejan un mayor efecto en la aplicación de mejoras en sus capacidades lingüísticas en comparación con la tenencia del título de secundaria. Por otro lado, los mayores de 35 años muestran un mayor aporte con la educación secundaria más que la búsqueda del dominio de otro idioma y por último los hombres y la zona urbana muestran comportamientos de mejora en la empleabilidad muy similares entre ambas políticas, lo que sugiere la necesidad de trabajar ambas capacidades para estos grupos.

Lo anterior, demuestra que, bajo los supuestos trabajados, el factor educativo es altamente influyente en la mejora de las posibilidades de empleabilidad sobre la población en estudio en la región central. Tanto, la obtención de la secundaria completa como el aprendizaje de un segundo idioma crean una fuerza laboral más cualificada, dando como resultado una reducción en la tasa de desempleo de la muestra de 8 p.p. pasando de un 14% a un 6%, como se observar en la tabla anterior.

4.4 Recomendaciones de Acciones de Política

En el presente apartado se abarca el cuarto objetivo del estudio en el que se presentan una serie de recomendaciones basadas en el análisis realizado en las secciones anteriores, con las cuales se pretende mediante un enfoque de mejora en las calificaciones educativas en cada uno de los 6 clústers encontrados, poder mejorar las condiciones de desempleabilidad en la región y con ellos lograr mejorar las condiciones socio económicas de su población.

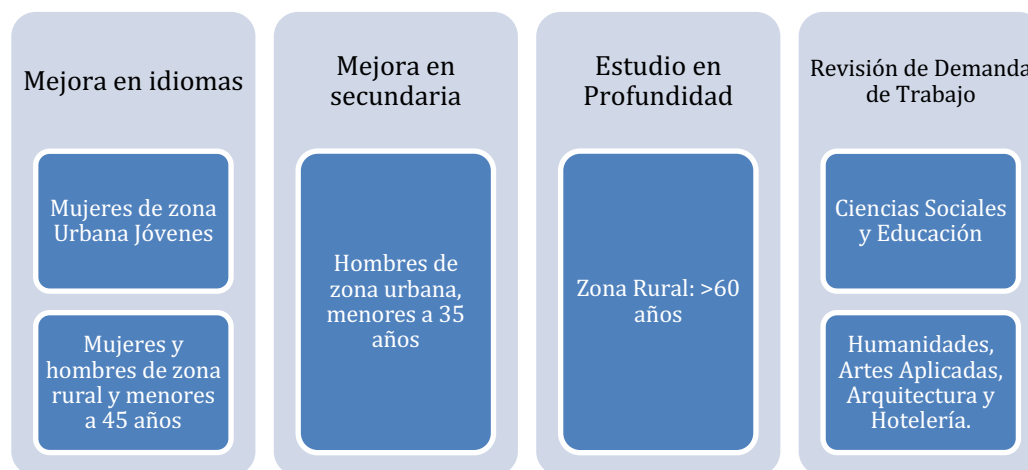
Tabla 22. Enfoque de Clústers de desempleados. Región Central de Costa Rica

Clúster	Nombre del Clúster	% del total	Enfoque Sugerido (Prioridad)		
			Mejorar Idiomas	Mejorar Niveles Educativos	Estudiar más en profundidad
1	Mujeres Urbanas menores a 45 años bajo calificadas	43%	X(1)	X(2)	
2	Hombres Urbanos menores a 35 años bajo calificados	32%	X(2)	X(1)	
3	Hombres Urbanos menores a 35 años medio/alto calificados	8%			X(1)
4	Mujeres Rurales menores a 45 años bajo calificadas	7%	X(1)	X(2)	
5	Mujeres Urbanas menores a 45 años medio/alto calificadas	5%			X(1)
6	Hombres Rurales menores a 35 años o mayores a 60 años bajo calificados	5%	X(2)	X(3)	X(1)

Fuente: Elaboración Propia

Con base en la analítica aplicada anteriormente, se recomienda tal y como se observa en la tabla 22 un enfoque basado en los clústers de desempleados identificados, así como en sus variables de mayor aporte determinadas con base en los modelos logísticos y predictivos estimados.

Figura 14: Sugerencia Plan de Acciones de Políticas Educativas Región Central



Fuente: Elaboración propia

El plan recomendado, tal y como se observa en la figura 14 se basa inicialmente en una estrategia de mejora de idiomas a tres grupos de la oferta laboral desempleada de la región. Siendo estos en primer lugar mujeres urbanas menores a 45 años bajo calificadas, las cuales al representar el 43% del total de desempleados se considera como la primera prioridad el mejorar sus habilidades lingüísticas y posteriormente buscar mejorar su nivel educativo.

Continuando con la estrategia de idiomas, esta se aplicaría además a dos grupos más minoritarios que residen en zonas rurales, que tiene muy baja calificación y que están conformados en un primer grupo por mujeres menores a 45 años (7%) y en un segundo por hombres (5%) con edades en los extremos (<35 o >60 años). Además, a este grupo en conjunto se sugiere una estrategia de capacitación extra en habilidades que les permitan insertarse al mercado laboral o desarrollar un emprendimiento propio.

Para el clúster 6, además se plantea el estudiar más en profundidad el grupo mayor a 60 años con el objetivo de identificar sus necesidades y el por qué su nivel de desempleabilidad viene en ascenso. Para así posteriormente poder diseñar un plan que les permita emplearse o conseguir algún tipo de ayuda social en caso de no puedan insertarse al mercado laboral.

El segundo enfoque del plan se dedica al grupo de desempleados conformado por hombres urbanos, menores a 35 años y con bajas calificaciones. Los cuales, debido a que representan el 32% del total de personas desocupadas, se requiere poner especial interés. Los modelos identificados para este caso recomiendan buscar inicialmente mejorar su nivel educativo es decir lograr que los que poseen estudios de secundaria incompleta o inferior, logren incrementar su calificación. Teniendo en cuenta además que, según las tendencias analizadas, sus mejoras educativas en los últimos años están siendo inferiores que las mujeres de dicha zona y edad.

Por último, el plan posee una tercera vertiente en la que se propone una futura investigación más profunda para dos grupos minoritarios (Clústers 3 y 5), los cuales residen en la zona urbana, son de edad joven y están medio/alto calificados con principalmente las especialidades mencionadas en la figura 14. Los cuales, teniendo en cuenta que teóricamente

cumplen las condiciones para tener altas probabilidades de empleabilidad, resulta importante determinar cuáles variables son las que dichas personas poseen, las cuales están dificultando su inserción laboral, tal y como puede ser si poseen una especialización con baja demanda laboral o si resulta ser por un tema de falta de experiencia.

Sumado a lo anterior, tal y como se observa en la tabla 23 se recomienda para el diseño de acciones de políticas de desempleo, utilizar un eje enfocado en el Derecho a un Empleo Digno, así como con trabajos en conjunto entre las instituciones de aprendizaje tanto públicas como privadas (INA, MEP, Centros de Enseñanza Superior), gobiernos locales y otras instituciones vinculadas con la población como Ministerios de Trabajo y de la Mujer.

Asimismo, se sugiere la evaluación de dichas acciones de política mediante un estudio de la evolución en las condiciones de empleabilidad de los grupos meta. Complementado además con un análisis histórico de los resultados reportados trimestralmente por la ECE para dicha región, como vía para identificar la efectividad real de estas en indicadores como con los niveles en las tasas de desempleo.

Tabla 23: Recomendaciones técnicas para una Política Pública de Empleo en la Región Central

Nombre	Eje	Población Meta	Ente encargado (Sugerido)	Detalle	Indicadores de Evaluación	Acciones para Cumplimiento (Encargado)
Programa Lingüístico de Empleabilidad Urbana	Derecho a un Empleo Digno	Mujeres desempleadas residentes de la zona urbana. menores a 45 años sin dominio de segundo idioma y bajo calificadas (Cluster 1)	Instituto Nacional de Aprendizaje (INA) + Ministerio Condición de la Mujer	Esta población representa el 43% del desempleo de la región, por lo que con esta política, pretendemos crear un programa de mejora en la calificación de este grupo mediante la enseñanza del inglés según las necesidades empresariales, con un trabajo multidisciplinario entre el INA, el Ministerio Condición de la Mujer y alianzas de empresa privada.	- Cantidad de mujeres graduadas del programa - Cantidad de mujeres empleadas gracias a la ayuda del programa	1) Identificación de candidatas al programa (Ministerio Condición de la Mujer) 2) Definir contenido del programa, duración y financiación (INA + Ministerio Condición de la Mujer + Alianzas Sector Privado) 3) Ejecutar el plan del piloto del programa (INA) 4) Coordinar estrategia de colocación laboral para las participantes del programa (Ministerio Condición de la Mujer)
Programa Técnico de Empleabilidad Urbana	Derecho a un Empleo Digno	Hombres desempleados residentes de la zona urbana menores a 35 años bajo calificadas (Cluster 2)	Ministerio de Educación Pública (MEP) + Gobiernos Locales (Municipalidades)	Esta población representa el 32% del desempleo de la región, por lo que con esta política, pretendemos crear un programa de mejora en la calificación de este grupo mediante educación secundaria/técnica para dicha población que decidió desertar, con un trabajo multidisciplinario entre el MEP, los gobiernos locales y alianzas de empresas privadas	-Cantidad de hombres graduados de secundaria técnica en el programa -Cantidad de hombres empleados gracias al programa	1) Identificación de candidatos al programa y necesidades técnicas según oferta laboral en la zona (Gobiernos Locales) 2) Definir Carreras Técnicas, horario de estudio según cada municipio (MEP + Gobierno Local + Alianzas Sector Privado) 3) Ejecutar el plan del piloto del programa (MEP) 4) Coordinar estrategia de colocación laboral para las participantes del programa (Gobiernos Locales)
Programa Lingüístico y de Capacitación Virtual Rural	Derecho a un Empleo Digno	Mujeres y hombres desempleados residentes de la zona rural, menores a 45 años sin dominio de segundo idioma y bajo calificados (Cluster 4 y 6)	Institutos Nacionales de Enseñanza + Gobiernos Locales (Municipalidades)	Esta población representa el 13% del desempleo de la región, por lo que con esta política, pretendemos crear un programa de mejora en la calificación de este grupo mediante la enseñanza del inglés y otras especialidades (Ej, Turismo) según las características de cada zona, con un trabajo multidisciplinario entre los gobiernos locales, alianzas de empresa privada y con Institutos Nacionales de Enseñanza. Se recomienda que sea de modalidad virtual, debido a las limitaciones de transporte.	-Cantidad de personas graduadas del programa -Cantidad de personas empleadas o con emprendimientos iniciados gracias a la ayuda del programa	1) Identificación de candidatos al programa y equipos tecnológicos requeridos (Gobiernos Locales) 2) Definir contenido del programa, duración y financiación (Gobiernos Locales + Alianzas Sector Privado + Institutos de Enseñanza) 3) Ejecutar el plan del piloto del programa (Institutos de Enseñanza + Gobiernos Locales) 4) Coordinar estrategia de colocación laboral para las participantes del programa (Gobiernos Locales)
Programa de Apoyo y Orientación Laboral Urbana	Derecho a un Empleo Digno	Mujeres y hombres desempleados residentes de la zona urbana, menores a 45 años medio/alto calificados (Cluster 3 y 5)	Ministerio de Trabajo y Seguridad Social (MTSS)	Esta población representa el 13% del desempleo de la región, por lo que con esta política, pretendemos crear un programa de orientación, guía y apoyo en la búsqueda de empleo para dicha población. A través de un plan en conjunto con el sector empresarial que les permita emplearse.	-Cantidad de personas integradas al programa -Cantidad de personas empleadas gracias a la ayuda del programa	1) Diseño, lanzamiento y difusión pública del programa y sus mecanismos de apoyo (MTSS) 2) Ejecución del programa a los interesados (MTSS) 3) Coordinar estrategia de colocación laboral para las participantes del programa (MTSS)

Fuente: Elaboración Propia

5. Conclusiones y Recomendaciones

5.1 Conclusiones

Los resultados demuestran en su primera etapa relacionada con la descomposición del problema, una región que pese a su menor tamaño geográfico es la que a nivel laboral absorbe la mayor oferta laboral y la cual durante el periodo analizado muestra una tendencia a aumento de edad, ya que los grupos de edad mayores a 35 cada año son más grandes relativamente.

Dicha oferta laboral en la región se destaca además por residir mayoritariamente en la zona urbana y a nivel educativo se identifican tres grandes grupos, inicialmente la mitad de ofertantes con secundaria incompleta o menos y la segunda mitad se lo dividen en parte iguales un grupo de universitarios graduados y en segundo lugar personas con secundaria.

Siendo las mujeres quienes poseen un mayor nivel educativo en comparación con los hombres. Por zona, la situación se agrava aún más en la zona rural con 6 de cada 10 ofertantes de empleo con niveles inferiores a secundaria. A nivel histórico, durante el periodo analizado, se muestra una mejora generalizada en la proporción universitaria y un importante incremento en el nivel secundaria en la zona rural, mientras que la zona urbana se muestra estancada en ese nivel.

Por dominio de un segundo idioma, esta habilidad solo se muestra en el 16% de la oferta, siendo las mujeres quienes más la poseen en comparación con los hombres y por zona, la rural se destaca por su muy leve dominio y urbana se distingue por mostrar un indicador levemente superior al valor nacional. Y por grupo de edad, los grupos extremos (<24 y >45 años) son quienes muestran mayores deficiencias con muy bajas tasas de dominio de una segunda lengua.

En lo que respecta a las condiciones laborales de la oferta laboral de dicha región, se concluye que los ocupados están colocándose principalmente en el sector terciario, siendo las actividades de información y comunicación, profesionales, científicas y técnicas e

inmobiliarias las cuales han mostrado los crecimientos más altos durante el periodo analizado.

La tasa de desempleo a nivel regional muestra un comportamiento a estabilizarse a sus valores prepandemia, con un valor del 14% para el 2022. Y en lo que respecta a sus valores granulares por grupo demográfico, la brecha por sexo continúa con desventaja para las mujeres. Por edad, la brecha continúa con los menores de 24 años quienes tienen las tasas más altas. Por zona, la situación es levemente mejor para la zona rural, sin embargo, esta brecha cada día resulta ser menor, ya que la rural se ha deteriorado considerablemente en el periodo analizado.

Por nivel educativo, los universitarios poseen menores tasas de desempleo en comparación con los grupos inferiores. Siendo a su vez el grupo con secundaria completa el nivel que más se ha deteriorado durante el periodo analizado. Por último, por dominio de segundo idioma existe una brecha en favor de los que poseen esa habilidad, siendo a nivel histórico esta brecha cada vez mayor.

Se han identificado posterior al análisis de clusterización, para el año 2022 seis grupos de desempleados, el primero en orden de importancia un grupo de mujeres urbanas sin dominio de segundo idioma y nivel educativo medio/bajo que abarca el 43% del total de desempleados. En segundo lugar, un grupo de hombres urbanos, sin dominio de idioma y nivel educativo medio/bajo de edad joven principalmente que abarcan el 32%.

Además, se identifican 4 grupos con cantidades entre el 5% y 7%, conformados por personas con características específicas, como son en primer lugar hombres jóvenes universitarios que representan el 8%, en segundo mujeres rurales con baja educación que son el 7%, en tercero mujeres urbanas altamente calificadas (5%) y por último hombres rurales con muy baja educación (5%).

Siendo a nivel de contraste entre ambos años analizados (2013 y 2022), lo más destacado que las características de los 6 grupos se mantienen considerablemente, a excepción del rango de edad, el cual tendió a aumentarse ligeramente en el año 2022 con respecto al 2013. Lo que refleja que estos grupos de personas con problemas de

empleabilidad estructurales se están manteniendo a lo largo del tiempo, pero que cada día conforme pasa el tiempo se va incrementando su edad.

Grupos para los que se han identificado tal y como se ha mostrado en los resultados, gracias al análisis descriptivo, la clusterización y los modelos predictivos, sus necesidades en temas de políticas que busquen mejorar sus posibilidades de emplearse, permitiendo así reducir la tasa de desempleo en la zona y contribuir así con mejoras en el desarrollo económico.

En fin, se puede confirmar que el desempleo en Costa Rica se encuentra en un estado estructural dada la investigación y comparación de datos realizado para el periodo 2013-2022. Además, si bien el gobierno ha tratado de implementar una serie de políticas para tratar de reducir el desempleo, a la fecha del estudio no se observan resultados muy significativos y concluimos la alta necesidad de revisar las deficiencias que poseen ciertos grupos de desempleados en comparación con las habilidades que demanda el mercado actualmente.

5.2 Recomendaciones

Ahora bien, con base a los resultados obtenidos a través de la aplicación de la metodología de Pensamiento Computacional en el análisis del mercado laboral y el desempleo en la Región Central de Costa Rica. Se recomiendan políticas públicas en tres categorías: educación, idiomas y empleabilidad, ver tabla 23. Sumado a ellos, se establece una última recomendación basada en la continuidad del uso de estos métodos cuantitativos de ciencia de datos para el análisis de problemas económicos.

5.2.1 Educación

En cuanto a la educación, se sugiere la implementación de programas de capacitación y formación en habilidades demandadas por el mercado laboral según cada zona, principalmente para hombre jóvenes residentes de la zona urbana, para así evitar el visible deterioro que se evidencia en dicho grupo de personas. Los cuales, tal y como se presentó en el análisis de la situación sus calificaciones han ido mejorando más lentamente conforme el tiempo en comparación con las mujeres.

Sumado a lo anterior, se recomiendan dirigir esfuerzos educativos incluyendo un fortalecimiento de la educación primaria y secundaria en la zona rural, ya que las diferencias en comparación con la urbana son significativas, siendo su principal debilidad la poca existencia de personas con calificaciones universitarias y un porcentaje mayor al 60% de personas que aún no concluyen la secundaria, por lo que es importante estudiar en profundidad dicho fenómeno y replantear la manera en que se intenta mejorar sus calificaciones.

5.2.2 Idiomas

En segundo lugar, se propone la promoción de la educación bilingüe, poniendo especial enfoque en mujeres tanto de zonas urbanas como rurales enfocada en edades medias, así como para hombres de la zona rural de todas las edades. Teniendo en cuenta que el dominio de un segundo idioma mostró ser un factor altamente determinante en las posibilidades de empleabilidad en ambas zonas para dichos grupos de personas.

5.2.3 Programas de Empleabilidad

En cuanto a la empleabilidad, se recomienda la implementación de programas de orientación laboral y de intermediación laboral principalmente para los grupos de la zona urbana que se identifican con altas calificaciones educativas pero que se mantienen como desempleados. Y se plantea además poner énfasis en la promoción de prácticas laborales y pasantías para jóvenes y personas con poca experiencia laboral, en ambas zonas y para ambos sexos. Teniendo en cuenta la elevada tasa de desempleo que mantienen las personas jóvenes en dicha región, en comparación con los otros grupos de edad.

La implementación de estas recomendaciones podría contribuir significativamente a reducir el desempleo estructural en la región y mejorar la calidad de vida de su población. Todo esto, con el fin de un objetivo conjunto de incentivar un mayor desarrollo económico de la zona en estudio y a su vez del país en general.

5.2.4 Aplicación de la Ciencia de Datos como Herramienta en Análisis Económicos

Se recomienda el uso de algoritmos de Aprendizaje Automático, para analizar cualquier problema económico en el que podamos obtener bases de datos con cantidades de datos importantes, tales como desempleo, pobreza, producción industrial, sector financiero, etc. Ya que el poder predictivo que estos muestran cada día resulta mejor y además, tal y como se ha planteado en el presente estudio, en importantes universidades y centros de investigación del mundo se están empezado a aplicar estas metodologías.

Para el caso de Costa Rica, además existe la oportunidad de acceder a las bases de datos de las encuestas realizadas por el INEC, por lo que las oportunidades son muchas. Tal y como es el área estudiada en dicho estudio, en la que se pueden analizar otras regiones, el país entero o grupos específicos de personas para identificar mediante modelaciones su evolución y tendencias a lo largo del tiempo, que permitan mejorar las políticas de empleo del país.

6. Anexos

Anexo 1: Estimaciones Índices de Desempleabilidad Región Central Costa Rica

	%	Tasa Desempleo I-2022	Índice de Desempleo
General	100%	14%	
Hombre	44%	11%	5%
Mujer	56%	18%	10%
Urbana	87%	14%	12%
Rural	13%	13%	2%
15 a 24 años	35%	36%	13%
25 a 34 años	28%	15%	4%
35 a 44 años	17%	10%	2%
45 a 59 años	16%	8%	1%
60 y más	4%	5%	0%
Universitario con título	12%	7%	1%
Secundaria Completa o Universitario sin título	38%	19%	7%
Secundaria Incompleta o menos	51%	14%	7%
Dominio Segundo Idioma	13%	11%	1%
Sin Dominio Segundo Idioma	87%	14%	12%

Fuente Elaboración Propia con datos de ECE-INEC I-2022

Anexo 2: Estimaciones Índices de Desempleabilidad Mujeres Región Central Costa Rica

	%	Tasa de Desempleo I-2022	Índice de Desempleo
General	100%	18%	
Urbana	87%	18%	16%
Rural	13%	22%	3%
15 a 14 años	29%	39%	11%
25 a 34 años	25%	18%	5%
35 a 44 años	27%	17%	5%
45 a 59 años	17%	13%	2%
60 y más	1%	3%	0%
Universitario con título	11%	7%	1%
Secundaria Completa o Universitario sin título	35%	23%	8%
Secundaria Incompleta o menos	54%	24%	13%
Dominio Segundo Idioma	10%	10%	1%
Sin Dominio segundo Idioma	90%	20%	18%

Fuente Elaboración Propia con datos de ECE-INEC I-2022

Anexo 3: Estimaciones Índices de Desempleabilidad Hombres Región Central Costa Rica

	%	Tasa de Desempleo I-2022	Índice de Desempleo
General	100%	11%	
Urbana	88%	11%	10%
Rural	12%	9%	1%
15 a 14 años	43%	34%	14%
25 a 34 años	32%	13%	4%
35 a 44 años	5%	3%	0%
45 a 59 años	13%	5%	1%
60 y más	7%	6%	0%
Universitario con título	12%	7%	1%
Secundaria Completa o Universitario sin título	41%	17%	7%
Secundaria Incompleta o menos	46%	9%	4%
Dominio Segundo Idioma	17%	12%	2%
Sin Dominio segundo Idioma	83%	10%	9%

Fuente Elaboración Propia con datos de ECE-INEC I-2022

Anexo 4: Repositorio Github Códigos

https://github.com/bolanos-m-5/Tesis_UNA_Bolanos_Mongrillo_2024/tree/main

Anexo 5: Resultados Modelos Logísticos

Modelo Muestra Completa 2013

Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.00919915	0.013759583	73.345184 0.000000e+00
idiomav21	0.27045609	0.010217264	26.470501 2.119419e-154
zonav21	0.24453035	0.009428557	25.935077 2.679444e-148
Sexov21	0.24067132	0.006296853	38.220889 0.000000e+00
calificacion_11	0.05426076	0.013151899	4.125698 3.696127e-05
calificacion_21	0.15163383	0.007850557	19.315042 4.014080e-83
calificacion_31	0.13282872	0.010393892	12.779498 2.134419e-37
edadv31	1.54100389	0.006356971	242.411653 0.000000e+00

Modelo Hombres 2013

Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.00529809	0.023305080	86.04554 0.000000e+00
idiomav21	0.41133156	0.013381387	30.73908 1.711065e-207
zonav21	0.49174653	0.013532091	36.33929 3.879819e-289
calificacion_11	-0.57643871	0.022910166	-25.16083 1.075817e-139
calificacion_21	0.04698986	0.010641591	4.41568 1.006931e-05
calificacion_31	-0.21770415	0.014198590	-15.33280 4.616419e-53
edadv31	1.32576332	0.008814116	150.41364 0.000000e+00

Modelo Mujeres 2013

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.34430985	0.017259941	19.948495	1.544674e-88
idiomav21	0.12682529	0.016023493	7.914959	2.473349e-15
zonav21	-0.08162053	0.013446839	-6.069868	1.280156e-09
calificacion_11	0.52727828	0.017054354	30.917516	6.946945e-210
calificacion_21	0.31225178	0.011705604	26.675409	9.081184e-157
calificacion_31	0.44009000	0.015325989	28.715275	2.459497e-181
edadv31	1.84942902	0.009511464	194.442099	0.000000e+00

Modelo Zona Urbana 2013

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.05548645	0.015416670	68.4639727	0.000000e+00
idiomav21	0.29693747	0.010633022	27.9259709	1.291106e-171
Sexov21	0.14895092	0.006759092	22.0371182	1.269658e-107
calificacion_11	-0.00844760	0.015022858	-0.5623164	5.739005e-01
calificacion_21	0.17217093	0.008402184	20.4912105	2.579009e-93
calificacion_31	0.04029939	0.010801909	3.7307651	1.908992e-04
edadv31	1.65967504	0.006806345	243.8423438	0.000000e+00

Modelo Zona Rural 2013

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.26260570	0.03127279	40.373936	0.000000e+00
idiomav21	0.05198111	0.03838670	1.354144	1.756905e-01
Sexov21	0.89664015	0.01765645	50.782594	0.000000e+00
calificacion_11	0.21108629	0.02766504	7.630075	2.346174e-14
calificacion_21	-0.10429399	0.02245532	-4.644511	3.408823e-06
calificacion_31	1.20598703	0.04456137	27.063507	2.648916e-161
edadv31	0.69988121	0.01877649	37.274335	4.275968e-304

Modelo Muestra Completa 2022

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.300484388	0.010935161	-27.478735	3.152359e-166
idiomav21	0.017487594	0.007884720	2.217909	2.656102e-02
zonav21	0.145113116	0.007174420	20.226461	5.726892e-91
Sexov21	0.778182940	0.004872591	159.706187	0.000000e+00
calificacion_11	0.479139705	0.009654083	49.630785	0.000000e+00
calificacion_21	-0.007312832	0.005814282	-1.257736	2.084872e-01
calificacion_31	0.887907134	0.008099589	109.623723	0.000000e+00
edadv31	1.430768764	0.005522406	259.084297	0.000000e+00

Modelo Hombres 2022

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.3072558	0.015239835	20.16136	2.139161e-90

idiomav21	-0.2695846	0.010866486	-24.80881	7.202486e-136
zonav21	0.3430239	0.010708217	32.03371	3.701430e-225
calificacion_11	0.6513661	0.014470549	45.01323	0.000000e+00
calificacion_21	-0.3234070	0.008443944	-38.30046	0.000000e+00
calificacion_31	0.4802972	0.012080414	39.75834	0.000000e+00
edadv31	1.7731619	0.007704602	230.14322	0.000000e+00

Modelo Mujeres 2022

Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.20669700	0.015025985	-13.755970 4.689498e-43
idiomav21	0.35172073	0.011669985	30.138918 1.498606e-199
zonav21	-0.06002213	0.009875143	-6.078103 1.216127e-09
calificacion_11	0.37443336	0.013353530	28.040029 5.285233e-173
calificacion_21	0.28113023	0.008083782	34.777067 5.404764e-265
calificacion_31	1.04832900	0.010897930	96.195240 0.000000e+00
edadv31	1.18258063	0.008207277	144.089277 0.000000e+00

Modelo Zona Urbana 2022

Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.38699039	0.011610874	-33.329997 1.420041e-243
idiomav21	-0.06693526	0.008208180	-8.154701 3.500443e-16
Sexov21	0.72169600	0.005191750	139.008242 0.000000e+00
calificacion_11	0.57619298	0.010361188	55.610708 0.000000e+00
calificacion_21	0.11766195	0.006281634	18.731107 2.761170e-78
calificacion_31	0.85364966	0.008590603	99.370164 0.000000e+00
edadv31	1.40071674	0.005940640	235.785492 0.000000e+00

Modelo Zona Rural 2022

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.20783114	0.03105407	6.692557 2.193036e-11	
idiomav21	0.53829902	0.03269120	16.466177 6.420133e-61	
Sexov21	1.17140850	0.01429968	81.918518 0.000000e+00	
calificacion_11	0.08314331	0.02794312	2.975449 2.925601e-03	
calificacion_21	-0.76447916	0.01594142	-47.955525 0.000000e+00	
calificacion_31	1.12185751	0.02516473	44.580555 0.000000e+00	
edadv31	1.57663289	0.01546246	101.965196 0.000000e+00	

Anexo 6: Comandos en R Studio para la Clusterización

Lectura de Datos

```
# La base de datos a utilizar es formato .sav (codificado para SPSS software)
y corresponde a un extracto de los microdatos de la Encuesta Continua de Empleo,
filtrando la región de planificación= Central

dataset_test = read.spss("~/Google Drive/Otras computadoras/Dell 5379/Documentos/
UNA/Tesis version final/Central_2010_2022_v2.sav", to.data.frame=TRUE)

# La base de datos del INEC contiene más de 300 variables, por lo que se procede a
filtrar las variables a analizar

dataset_test2 <- dataset_test[, c(1,2,7,8,9,10,22, 26, 183, 184,185, 196, 198,
, 199, 201, 204, 206)]

#Se filtra el periodo a analizar

data_2022 <- subset(dataset_test2, dataset_test2$ID_AMO == "2022" & dataset_test2$ID_TRIMESTRE == "1")

# Se filtran las variables de año y trimestre, además de la variable que no se utilizará
en el análisis.

data_2022 <- data_2022[, c(-1, -2, -6)]

# se filtran solamente las observaciones con condición = desempleado, ya que son las de
interés de analizar.

data_2022 <- subset(data_2022, data_2022$Condicion_actividad == "Desempleado")

#Se extrae la variable factor de expansión, que posteriormente permite expandir la muestra a
valores poblacionales. Además, se filtran las variables de especialización de educación y
experiencia las cuales se utilizarán posteriormente en el análisis.

Fact_expand2 <- data_2022[, c(4,8,11,14)]

# Se definen las variables finales para la clusterización y se ajustan los formatos de la
edad, nivel educativo e idioma

desempleadov3 <- data_2022[, c(-1,-4,-6,-8,-9,-11,-12,-13,-14)]

desempleadov3$Edad <- as.numeric(levels(desempleadov3$Edad))[desempleadov3$Edad]

desempleadov3$Nivel_educativo <- factor(desempleadov3$Nivel_educativo, ordered = TRUE)

desempleadov3$Idioma <- factor(desempleadov3$Idioma)

#Se imprime un detalle de la estructura de la base de datos final a clusterizar

str(desempleadov3)

#summary(desempleadov3)
```

Matriz de distancias

```
# Los métodos de clusterización, cuando existen variable no numéricas (Ej, Nivel Educativo), necesitan aplicar un ajuste matemático en el que dichas variables se transforman en números para poder identificar las similitudes entre observaciones.

# Por lo que se utiliza la función "Daisy" que permite hacer dicho ajuste con el método "gower"

d2 <- daisy(desempleadov3, metric = "gower")

str(d2)
```

Optimización de los parámetros

```
# Se definen los métodos de Clusterización a utilizar en el ciclo de optimización

metodos <- c("ward.D",
            "ward.D2", "single", "complete", "average", "mcquitty", "median", "centroid")

# Se definen las matrices en blanco donde posteriormente se almacenaran los resultados de clusterización por método y cantidad de clústers

index <- matrix(nrow = 3, ncol = 8)

best <- matrix(nrow = 8, ncol = 2)

# Se inicia un ciclo donde se prueban los 8 métodos establecidos anteriormente e con cantidad de clústers entre 4 y 6, para buscar las mayor precisión.

# Se imprimen los resultados en las matrices previamente creadas y al final se filtra el método y cantidad de clústers con mejor precisión.

for (i in 1:8) {

  datos <- NbClust(d2, diss = d2, distance = NULL, method = metodos[i],
                 index = "silhouette", min.nc = 4, max.nc = 6)

  index[,i] <- datos$All.index
  best[i,] <- datos$Best.nc
}

row.names(index) <- c("4_clusters", "5_clusters", "6_clusters")
colnames(index) <- metodos
print(index)

row.names(best) <- metodos
colnames(best) <- c("Number_of_Clusters", "Silhouette_Index")

best <- data.frame(best)

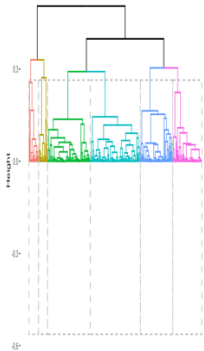
best <- best[order(best$Silhouette_Index, decreasing = TRUE),]
```

```
print(best)
best.model <- row.names(best)[1]
best.clusters <- best[1,1]
```

Modelo Final con base en la optimización anterior

```
# Se realiza la partición de las observaciones según el mejor modelo obtenido anteriormente
modelo1 <- hclust(d2, method = best.model)

# Se realiza un grafico, para ilustrar la distribución de la clusterización, con base en lo optimizado anteriormente
fviz_dend(modelo1,
           k = best.clusters,
           cex = 1.3,
           color_labels_by_k = FALSE,
           rect = TRUE,
           main = " ", show_labels = FALSE)
```



División de la base de datos

```
#Con el modelo y la cantidad optima de clústers, se procede a asignar un cluster a cada observación de la base de datos y se le agrega la columna del factor de expansión y otras del dataset Fact_expand2.
grupos <- cutree(modelo1, k = best.clusters)
NDatos <- cbind(desempleadov3, Fact_expand2 , grupos)
cluster <- NDatos$grupos
sel.cluster <- list()
datos.cluster <- list()
```

```

for (i in 1:best.clusters) {
  sel.cluster[[i]] <- match(cluster, i, 0)

  datos.cluster[[i]] <- NDatos[sel.cluster[[i]] > 0,]
}

```

Aplicar el Factor de Expansión a cada grupo y exportar archivos

```

# Cada observación se expande según el valor de factor de expansión asignado por el INEC.

# Además, cada clúster se guarda en formato csv con el nombre datos.cluster.i.csv, para un posterior análisis.

for (i in 1:best.clusters) {
  datos.cluster[[i]] <- expandRows(datos.cluster[[i]], "Factor_ponderacion")
  name <- paste("~/Google Drive/Otras computadoras/Dell 5379/Documentos/UNA/Tesis version final/Datos_Cluster_2022_jun24_",i, ".csv")
  archivo <- ("Datos_Cluster")
  data.grupo <- data.frame(datos.cluster[[i]])
  #write.csv(data.grupo, as.character(name))
  #print(name)
  #print(summary(datos.cluster[[i]]))
}

```

Anexo 7. Comandos R Studio para Modelos Logísticos

Lectura de Datos

```

# La base de datos a utilizar es format .sav (codificado para SPSS software) y corresponde a un extracto de los microdatos de la Encuesta Continua de Empleo, filtrando la region de planificación= Central y Fuerza de trabajo = Verdadero.

dataset2 = read.spss("FT_2013_1.sav", to.data.frame=TRUE)

#dataset2 = read.spss("FT_CENTRAL_I_2022_v1.sav", to.data.frame=TRUE)

```

Ajustar Variables (excluyendo Nivel Educativo)

```

# se realiza un ajuste de la variable Edad, convirtiéndolas en aptas para la modelación. Pasándola de tipo nivel a tipo número entero.

dataset2$Edad <- droplevels(dataset2$Edad)
dataset2$Edad <- as.character(dataset2$Edad)
dataset2$Edad <- as.integer(dataset2$Edad)

```

```

# Las variables Idioma, zona, sexo y condicion de actividad se convierten en
tipo dicotómicas (1 o 0), para poder ser analizadas en los modelos.

dataset2$Idioma <- droplevels(dataset2$Idioma)

dataset2$idiomav2 <- as.factor(ifelse( dataset2$Idioma == "Sí", 1, 0))

dataset2$zonav2 <- as.factor(ifelse( dataset2$Zona == "Urbana", 0, 1))

dataset2$Sexov2 <- as.factor(ifelse(dataset2$Sexo == "Mujer", 0, 1))

dataset2$Condicion_actividad <- droplevels(dataset2$Condicion_actividad)

dataset2$Condicion_actividadv2 <- as.factor(ifelse(dataset2$Condicion_activi
dad == "Ocupado", 1, 0))

```

Ajustar Variable Nivel Educativo (Segùn Modelo a estimar)

```

# La variable Nivel Educativo, se ajusta en tres nuevas variables según dicho
s niveles:

# 1) calificación 1 = Primaria completa o superior
# 2) calificación 2 = Secundaria completa o superior
# 3) calificación 3 = Universitarion con título

# Además, estas se transforman en tipo dicotómicas

dataset2$calificacion_1 <- as.factor(ifelse(dataset2$Nivel_educativo == "Univ
ersitario con título" |dataset2$Nivel_educativo == "Universitario sin título"
| dataset2$Nivel_educativo == "Secundaria completa" |dataset2$Nivel_educativ
o == "Secundaria incompleta" |dataset2$Nivel_educativo == "Primaria completa
", 1, 0))

dataset2$calificacion_2 <- as.factor(ifelse(dataset2$Nivel_educativo == "Univ
ersitario con título" |dataset2$Nivel_educativo == "Universitario sin título"
| dataset2$Nivel_educativo == "Secundaria completa" , 1, 0))

dataset2$calificacion_3 <- as.factor(ifelse(dataset2$Nivel_educativo == "Univ
ersitario con título", 1, 0))

```

Extraer factor de expansión y expandir el dataset

```

# La base de datos del INEC viene con base en la muestra y un factor de expan
sión que permite expandir la muestra a valores estimados de la población por
lo que se extrae dicha variable y además se crea un nuevo dataset con dicho f
actor aplicado

factor_expand <- dataset2[, 6]

dataset3 <- expandRows(dataset2, "Factor_ponderacion")

```

Filtrar Variables Requeridas Dataset y crear valor de edad ajustado

```

# Se filtran las columnas destinadas para ser testeadas en los modelos logit

datasetv2 <- dataset3[, c(3,11,12,13,14,15,16,17)]

```

```
# Se crea una nueva variable = Edadv3, la cual ajusta la edad en tipo dicotómico
datasetv2$edadv3 <- as.factor(ifelse(datasetv2$Edad>25, 1,0))
str(datasetv2)
```

Crear función de Pseudo R2 para los métodos a utilizar

```
r2 <- function(created_model) {
  # Desviación nula y desviación del modelo
  dev <- created_model$deviance
  null_dev <- created_model$null.deviance
  # Cantidad de variables
  model_n <- length(created_model$fitted.values)
  # Hosmer and Lemeshow
  R_l <- 1 - dev / null_dev
  # Cox and Snell
  R_cs <- 1 - exp(-(null_dev - dev) / model_n)
  # Nagelkerke
  R_n <- R_cs / (1 - exp(-(null_dev / model_n)))
  return( cbind(R_l, R_cs, R_n ) )
}
```

Modelos Logit

Modelo con Muestra Completa

```
# Se filtran variables requeridas
dataset_final <- datasetv2[, c(-1)]
# Estimación del modelo logit
glm_model <- glm(Condicion_actividadv2~. , data = dataset_final, family = "binomial")
# Extracción coeficientes
summary(glm_model)$coefficients
# Extracción de ODDS Ratios
exp(glm_model$coefficients)
# Extracción de desviaciones
glm_model$null.deviance
glm_model$deviance
# Estimación Pseudo R-2
```

```
r2(glm_model)
```

Modelos por sexo

Hombres

```
# Se filtran solamente hombres
data_hombres <- filter(dataset_final, Sexov2 == 1)
data_hombres <- data_hombres[, -c(3)]

# Estimación del modelo logit
glm_model_h <- glm(Condicion_actividadv2~. , data = data_hombres, family = "b
inomial")

# Extracción coeficientes
summary(glm_model_h)$coefficients

#Extracción de ODDS Ratios
exp(glm_model_h$coefficients)

# Extracción de desviaciones
glm_model_h$null.deviance
glm_model_h$deviance

#Estimación Pseudo R-2
r2(glm_model_h)
```

Mujeres

```
# Selección de solo mujeres
data_mujeres <- filter(dataset_final, Sexov2 == 0)
data_mujeres <- data_mujeres[, -c(3)]
str(data_mujeres)

# Estimación del modelo logit
glm_model_m <- glm(Condicion_actividadv2~. , data = data_mujeres, family = "b
inomial")

# Extracción coeficientes
summary(glm_model_m)$coefficients

#Extracción de ODDS Ratios
exp(glm_model_m$coefficients)

# Extracción de desviaciones
glm_model_m$null.deviance
glm_model_m$deviance

#Estimación Pseudo R-2
```

```
r2(glm_model_m)
```

Modelos por zona

Zona Urbana

```
# Filtración solamente zona urbana
data_urbana <- filter(dataset_final, zonav2 == 0)
data_urbana <- data_urbana[, -c(2)]
str(data_urbana )

# Estimación del modelo logit
glm_model_u <- glm(Condicion_actividadv2~. , data = data_urbana, family = "binomial")

# Extracción coeficientes
summary(glm_model_u)$coefficients

#Extracción de ODDS Ratios
exp(glm_model_u$coefficients)

# Extracción de desviaciones
glm_model_u$null.deviance
glm_model_u$deviance

#Estimación Pseudo R-2
r2(glm_model_u)
```

Zona Rural

```
#Filtración zona rural
data_rural <- filter(dataset_final, zonav2 == 1)
data_rural <- data_rural[, -c(2)]
str(data_rural )

# Estimación del modelo logit
glm_model_r <- glm(Condicion_actividadv2~. , data = data_rural, family = "binomial")

# Extracción coeficientes
summary(glm_model_r)$coefficients

#Extracción de ODDS Ratios
exp(glm_model_r$coefficients)

# Extracción de desviaciones
glm_model_r$null.deviance
glm_model_r$deviance
```

```
#Estimación Pseudo R-2
```

```
r2(glm_model_r)
```

Anexo 8. Comandos R Studio para Modelo Predictivo

Lectura de Datos

```
# La base de datos a utilizar es format .sav (codificado para SPSS software)
y corresponde a un extracto de los microdatos de la Encuesta Continua de Empleo,
filtrando la región de planificación= Central y Fuerza de trabajo = Verdadero.
```

```
dataset2 = read.spss("FT_CENTRAL_I_2022_v1.sav", to.data.frame=TRUE)
```

Ajustar Variables

```
# se realiza un ajuste de la variable Edad, convirtiéndolas en aptas para la
modelación. Pasándola de tipo nivel a tipo número entero.
```

```
dataset2$Edad <- droplevels(dataset2$Edad)
```

```
dataset2$Edad <- as.character(dataset2$Edad)
```

```
dataset2$Edad <- as.integer(dataset2$Edad)
```

```
# se realiza un ajuste de las variables Condicion actividad, idioma y nivel
educativo, convirtiéndolas en factores. Siendo el nivel educativo además ordenado,
debido a que dichos niveles poseen un orden lógico (Ex. Secundaria > Primaria)
```

```
dataset2$Condicion_actividad <- droplevels(dataset2$Condicion_actividad)
```

```
dataset2$Idioma <- droplevels(dataset2$Idioma)
```

```
dataset2$Nivel_educativo <- factor(dataset2$Nivel_educativo, ordered = TRUE)
```

Expandir Data según factor de ponderación

```
# La base de datos del INEC viene con base en la muestra y un factor de expansión
que permite expandir la muestra a valores estimados de la población por lo que se
crea un nuevo dataset con dicho factor aplicado.
```

```
dataset3 <- expandRows(dataset2, "Factor_ponderacion")
```

Filtrar Variables Requeridas

```
# Se filtran solamente las variables a utilizar en la modelación
```

```
dataset3 <- dataset3[,c(1,2,4,6,9,10)]
```

```
str(dataset3)
```

```
summary(dataset3)
```

Particionar dataset para entrenar (75% de datos) y testear modelos (25% de datos)

```

# se utiliza la librería "CreateDataPartition", permite dividir la muestras e
n 2 submuestras, con la misma proporción de la variable condición de activida
d. En este caso se divide en una muestra que contiene el 75% de datos y otra
que contiene el 25% de datos.

muestra <- createDataPartition(dataset3$Condicion_actividad, p = 0.75, list =
F)

# Posterior a la partición, se crean dos datasets: el de entrenamiento y el d
e testeo

ttraining <- dataset3[muestra,]
ttesting <- dataset3[-muestra,]

str(ttraining)
str(ttesting)

```

Entrenar Modelo de Redes Neuronales con dataset de training

```

# Se ejecuta un modelo de redes neuronales, con la base de datos de entrenami
ento a predecir la condición de actividad con las demás variables como predic
toras.

# Los parámetros de dicho modelo: Size, Maxnwts, rang, decay fueron previamen
te testeados hasta lograr resultados aceptables.

modelo_nnet <- train.nnet(Condicion_actividad~ ., data = ttraining, size = 15
0, MaxNWts = 10000, rang = 0.01, decay = 5e-4, maxit = 150, trace = TRUE)

```

Predecir datos en tabla de testing

```

# Posteriormente a definir el modelo, este se pone a predecir la base de dato
s de testeo para medir su precisión.

# Además, al asignar el parámetro type = prob, el modelo no despliega un resu
ltado definido (Empleado o desempleado), sino que despliega una probabilidad
dada de empleado (ej. 95%)

prediccion_nnet <- predict(modelo_nnet, ttesting, type = "prob")

prediccion_nnet

```

Optimizar Corte de Probabilidad

```

# Posterior a ejecutar el modelo en la table de testeo y como sus resultados
son probabilidades, es necesario ajustar el corte del modelo para asignar seg
ún un valor de probabilidad una observación como empleada o desempleada.

# Inicialmente se definen listas en blanco donde se van a almacenar los resul
tados según cada valor de corte testeado.

indices <- c()
cortes <- c()
indices.si <- c()
indices.no <- c()

```

```

# Se define una secuencia de números para testear los cortes, en este caso va
n de 0.10 a 0.50, con una secuencia de 0.01 (Ej. 0.10, 0.11, 0.12, etc)

cort <- seq(0.10, 0.50, by = 0.01)

# Se diseña un ciclo que recorre dicha secuencia de números, realiza el corte
en la base de datos de testeo y prueba su precisión vs los valores originales
de condición de actividad.

for (i in 1:50) {
Corte <- cort[i]
Clase <- ttesting[,2]
Score <- prediccion_nnet$prediction[,2]
Prediccion <- ifelse(Score > Corte, "Desempleado", "Ocupado")
MC <- table(Clase, Pred = factor(Prediccion, levels = c("Ocupado", "Desemplea
do")))
cortes [i] <- Corte
index <- general.indexes(mc = MC)
indices[i] <- index$overall.accuracy
indices.si[i] <- index$category.accuracy[1]
indices.no[i] <- index$category.accuracy[2]
}

# posteriormente se unen los resultados del ciclo en un dataset el cual poste
riormente se grafica

resultados <- cbind(cortes, indices, indices.si, indices.no)
resultados <- as.data.frame(resultados)
ggplot(data = resultados, mapping = aes( x= cortes, y = indices)) +
  geom_line(mapping = aes( x= cortes, y = indices.si, color = "Indices
si")) +
  geom_line(mapping = aes( x= cortes, y = indices.no , color = "Indices
no")) +
  geom_line(mapping = aes( x= cortes, y = indices , color = "Indices g
eneral"))
print(resultados)

```

Guardar modelo

```

# Finalmente el modelo finalmente creado y con el valor de corte ajustado, se
procede a guardar en la memoria de la computadora

#saveRDS(modelo_nnet, "./final_model4.rds")

```

Estimar estabilidad del modelo con validación cruzada con 5 grupos distintos

```

#Asignar Corte Optimo (Previamente Optimizado)

```

```

Corte <- 0.15

#Setear Validación cruzada, en este se realizarán 5 particiones distintas en
los datos.

numero.filas <- nrow(dataset2)
cantidad.validacion.cruzada <- 1
cantidad.grupos <- 5

# Setear listas vacías donde se van a almacenar los resultados.
indice <- c()
indices.si<- c()
indices.no <- c()

# Iniciar Validación Cruzada, con un ciclo que recorre la cantidad de validac
iones cruzadas a realizar, para este caso solamente es 1.

for (i in 1:cantidad.validacion.cruzada) {

# Posteriormente se crean los 5 grupos distintos a validar mediante validació
n cruzada y se asignan los valores iniciales de las listas donde se van a im
primir los resultados

grupos <- createFolds(1:numero.filas, cantidad.grupos)
indice <-0
indices.si<- 0
indices.no <- 0

# Se inicia un ciclo que recorre los 5 grupos, corre el modelo con los paráme
tros datos, realiza el corte con el valor dado.

for (k in 1:cantidad.grupos) {
  muestra <- grupos[[k]]
  ttesting <- dataset2[muestra, ]
  ttraining <- dataset2[-muestra, ]

  modelo <- train.nnet(Condicion_actividad~ ., data = ttraining, size = 150
, MaxNWts = 10000, rang = 0.001, decay = 5e-4, maxit = 150, trace = TRUE)
  prediccion <- predict(modelo, ttesting, type = "prob")
  Clase <- ttesting[,2]
  Score <- prediccion$prediction[,2]
  Prediccion <- ifelse(Score > Corte, "Desempleado", "Ocupado")
  MC <- table(Clase, Pred = factor(Prediccion, levels = c("Ocupado", "Desem
pleado")))
  index <- general.indexes(mc = MC)
  indice <- index$overall.accuracy
  indices.si <- index$category.accuracy[1]

```

```

    indices.no <- index$category.accuracy[2]
  }
  indice[i] <- indice
  indices.si[i] <- indices.si
  indices.no[i] <- indices.no
}
# Posteriormente se asignan los resultados obtenidos en todos los grupos en n
# uevos datasets y se presentan todos en su conjunto.
resultados.indice <- data.frame(
  "redes_nnet" = indice)
resultados.indice.si <- data.frame(
  "redes_nnet" = indices.si)
resultados.indice.no <- data.frame(
  "redes_nnet" = indices.no)
resultados <- rbind(resultados.indice, resultados.indice.si, resultados.indic
e.no)
resultados

```

Análisis de Casos

Leer modelo optimizado

```

# Se lee el modelo creado y ajustado
super_model <- readRDS("./final_model4.rds")
print(super_model)

```

Setear base de datos a analizar

```

# Se divide la base de datos en 2: desempleados y ocupados
desempleado <- subset(dataset3, dataset3$Condicion_actividad == "Desempleado"
)
ocupado <- subset(dataset3, dataset3$Condicion_actividad == "Ocupado")
# El modelo va aplicar solamente a los desempleados, para el caso de los ocup
# ados su condición de actividad se mantendrá constante.
ocupado$Prediccion <- as.factor("Ocupado")
# Resumen de ambos datasets
summary(desempleado)
summary(ocupado)

```

Caso 1: Dominio de Segundo Idioma

Simulación 1 Toda la muestra con variable idioma ajustada

```

# Se crea una nueva base de datos para la simulación 1
desempleado.1 <- desempleado

# Se asigna a la variable idioma = 1 (Todos con dominio de idioma) y se ajust
a como un factor.
desempleado.1$Idioma <- 1
desempleado.1$Idioma <-factor(desempleado.1$Idioma, levels= c("1", "2"),
labels=c("Sí", "No"))

# Se realiza la predicción con el modelo leído, la nueva base de datos y con
el parámetro que determina probabilidades
prediccion1 <- predict(super_model, desempleado.1, type = "prob")

# Se aplica el corte, según lo optimizado anteriormente
Corte <-0.15

# Se crea la matriz de confusión y sus respectivos indicadores de precisión
Clase <- desempleado.1[,2]
Score <- prediccion1$prediction[,2]
Prediccion <- ifelse(Score > Corte, "Desempleado", "Ocupado")
MC <- table(Clase, Pred = factor(Prediccion, levels = c("Ocupado", "Desemplea
do")))
index <- general.indexes(mc = MC)
indice <- index$overall.accuracy
indices.si <- index$category.accuracy[1]
indices.no <- index$category.accuracy[2]

# Se guarda la predicción en una variable y se le añade al dataset desemplead
o
Prediccion <- as.factor(Prediccion)
resultados1 <- cbind (desempleado.1, Prediccion)

# Se une el dataset de desempleado con la predicción con el dataset de los oc
upados, para medir el impacto de la simulación
resultados1.1 <- rbind(resultados1, ocupado)
tab1(resultados1.1$Prediccion, sort.group = "decreasing", graph = FALSE)
tab1(resultados1.1$Condicion_actividad, sort.group = "decreasing", graph = FA
LSE)

# Por último, se procede a guardar la simulación en un archivo csv
#write.csv(resultados1.1, "Casos1_simulado.csv")

```

Simulación 1.1: Mujeres

```
desempleado.1.2 <- subset(resultados1.1, Sexo == "Mujer")
```

```
tab1(desempleado.1.2$Prediccion, sort.group = "decreasing", graph = FALSE)
tab1(desempleado.1.2$Condicion_actividad, sort.group = "decreasing", graph =
FALSE)
```

Simulación 1.2: Hombres

```
desempleado.1.3 <- subset(resultados1.1, Sexo == "Hombre")
tab1(desempleado.1.3$Prediccion, sort.group = "decreasing", graph = FALSE)
tab1(desempleado.1.3$Condicion_actividad, sort.group = "decreasing", graph =
FALSE)
```

Simulación 1.3: Zona Urbana

```
desempleado.1.4 <- subset(resultados1.1, Zona == "Urbana")

tab1(desempleado.1.4$Prediccion, sort.group = "decreasing", graph = FALSE)

tab1(desempleado.1.4$Condicion_actividad, sort.group = "decreasing", graph =
FALSE)
```

Simulación 1.4: Zona Rural

```
desempleado.1.5 <- subset(resultados1.1, Zona == "Rural")
tab1(desempleado.1.5$Prediccion, sort.group = "decreasing", graph = FALSE)
tab1(desempleado.1.5$Condicion_actividad, sort.group = "decreasing", graph =
FALSE)
```

Simulación 1.5: 25 años o menos

```
desempleado.1.6 <- subset(resultados1.1, Edad <= 25)
tab1(desempleado.1.6$Prediccion, sort.group = "decreasing", graph = FALSE)
tab1(desempleado.1.6$Condicion_actividad, sort.group = "decreasing", graph =
FALSE)
```

Simulación 1.6: 35 años o menos

```
desempleado.1.7 <- subset(resultados1.1, Edad <= 35)
tab1(desempleado.1.7$Prediccion, sort.group = "decreasing", graph = FALSE)
tab1(desempleado.1.7$Condicion_actividad, sort.group = "decreasing", graph =
FALSE)
```

Simulación 1.7: 36 años o mas

```
desempleado.1.8 <- subset(resultados1.1, Edad > 35)
tab1(desempleado.1.8$Prediccion, sort.group = "decreasing", graph = FALSE)
```

```
tab1(desempleado.1.8$Condicion_actividad, sort.group = "decreasing", graph = FALSE)
```

Caso 2: Todos Secundaria Completa o Superior

Simulación 2: Toda la muestra con variable educación ajustada

```
# Se crea una nueva base de datos para la simulacion 1
desempleado.2 <- desempleado

# Se ajusta la variable Nivel educativo, dejando a todas las personas con nivel inferior a secundaria completa, con secundaria completa y las demás constantes.
desempleado.2$Nivel_educativo <- as.numeric(desempleado.2$Nivel_educativo)
desempleado.2$Nivel_educativo <- ifelse(desempleado.2$Nivel_educativo<=5, 5,
desempleado.2$Nivel_educativo)

desempleado.2$Nivel_educativo <-factor (desempleado.2$Nivel_educativo, levels
= c("1", "2", "3", "4", "5", "6", "7"), labels=c("Ninguno", "Primaria incompleta",
"Primaria completa", "Secundaria incompleta", "Secundaria completa", "Universi
tario sin título", "Universitario con título"))

summary(desempleado.2)

# Se realiza la predicción con el dataset ajustado y se le asigna el corte
prediccion2 <- predict(super_model, desempleado.2, type = "prob")
Corte <-0.15

# Se calcula la matriz de confusión y los indicadores de precisión del modelo
Clase <- desempleado.2[,2]
Score <- prediccion2$prediction[,2]
Prediccion <- ifelse(Score > Corte, "Desempleado", "Ocupado")
MC <- table(Clase, Pred = factor(Prediccion, levels = c("Ocupado", "Desemplea
do")))
index <- general.indexes(mc = MC)
indice <- index$overall.accuracy
indices.si <- index$category.accuracy[1]
indices.no <- index$category.accuracy[2]

# Se convierte la predicción en una variable y se le añade al dataset desempl
eado simulado

Prediccion <- as.factor(Prediccion)
resultados2 <- cbind (desempleado.2, Prediccion)

# se une el dataset desempleado con el ocupado y se evalúa el impacto de la s
imulación
resultados2.1 <- rbind(resultados2, ocupado)
```

```
tab1(resultados2.1$Prediccion, sort.group = "decreasing", graph = FALSE)
tab1(resultados2.1$Condicion_actividad, sort.group = "decreasing", graph = FALSE)
# Por último, se guarda la simulación en tipo csv
#write.csv(resultados2.1, "Caso2_simulado.csv")
```

Simulación 2.1: Mujeres

```
desempleado.2.2 <- subset(resultados2.1, Sexo == "Mujer")
tab1(desempleado.2.2$Prediccion, sort.group = "decreasing", graph = FALSE)
tab1(desempleado.2.2$Condicion_actividad, sort.group = "decreasing", graph = FALSE)
```

Simulación 2.2: Hombres

```
desempleado.2.3 <- subset(resultados2.1, Sexo == "Hombre")
tab1(desempleado.2.3$Prediccion, sort.group = "decreasing", graph = FALSE)
tab1(desempleado.2.3$Condicion_actividad, sort.group = "decreasing", graph = FALSE)
```

Simulación 2.3: Zona Urbana

```
desempleado.2.4 <- subset(resultados2.1, Zona == "Urbana")
tab1(desempleado.2.4$Prediccion, sort.group = "decreasing", graph = FALSE)
tab1(desempleado.2.4$Condicion_actividad, sort.group = "decreasing", graph = FALSE)
```

Simulación 2.4: Zona Rural

```
desempleado.2.5 <- subset(resultados2.1, Zona == "Rural")
tab1(desempleado.2.5$Prediccion, sort.group = "decreasing", graph = FALSE)
tab1(desempleado.2.5$Condicion_actividad, sort.group = "decreasing", graph = FALSE)
```

Simulación 2.5: 25 años o menos

```
desempleado.2.6 <- subset(resultados2.1, Edad <= 25)
tab1(desempleado.2.6$Prediccion, sort.group = "decreasing", graph = FALSE)
tab1(desempleado.2.6$Condicion_actividad, sort.group = "decreasing", graph = FALSE)
```

Simulación 2.6: 35 años o menos

```
desempleado.2.7 <- subset(resultados2.1, Edad <= 35)
  tab1(desempleado.2.7$Prediccion, sort.group = "decreasing", graph = FALSE
)
```

```
tab1(desempleado.2.7$Condicion_actividad, sort.group = "decreasing", graph = FALSE)
```

Simulación 2.7: 36 años o mas

```
desempleado.2.8 <- subset(resultados2.1, Edad > 35)
tab1(desempleado.2.8$Prediccion, sort.group = "decreasing", graph = FALSE)
tab1(desempleado.2.8$Condicion_actividad, sort.group = "decreasing", graph = FALSE)
```

Caso 3: Todos Secundaria Completa o Superior y Dominio de Segundo Idioma

Simulación 3: Toda la muestra con variables educación e idioma Ajustadas

```
# Se crea un nuevo dataset para la simulación
desempleado.3 <- desempleado

# Se asigna a la variable idioma = 1 (Todos con dominio de idioma) y se ajusta
a como un factor.
desempleado.3$Idioma <- 1
desempleado.3$Idioma <- factor(desempleado.3$Idioma, levels= c("1", "2"),
labels=c("Sí", "No"))

# Se ajusta la variable Nivel educativo, dejando a todas las personas con nivel
inferior a secundaria completa, con con secundaria completa y las demás constantes.
desempleado.3$Nivel_educativo <- as.numeric(desempleado.3$Nivel_educativo)
desempleado.3$Nivel_educativo <- ifelse(desempleado.3$Nivel_educativo<=5, 5,
desempleado.3$Nivel_educativo)

desempleado.3$Nivel_educativo <- factor (desempleado.3$Nivel_educativo, levels
= c("1", "2", "3", "4", "5", "6", "7"), labels=c("Ninguno", "Primaria incompleta",
"Primaria completa", "Secundaria incompleta", "Secundaria completa", "Universi
tario sin título", "Universitario con título"))

summary(desempleado.3)

# Se realiza la predicción con el modelo previamente ajustado y el tipo = pro
babilidades
prediccion3 <- predict(super_model, desempleado.3, type = "prob")

Corte <- 0.15
Clase <- desempleado.3[,2]
Score <- prediccion3$prediction[,2]
Prediccion <- ifelse(Score > Corte, "Desempleado", "Ocupado")

MC <- table(Clase, Pred = factor(Prediccion, levels = c("Ocupado", "Desemplea
do")))

index <- general.indexes(mc = MC)
indice <- index$overall.accuracy
```

```

indices.si <- index$category.accuracy[1]
indices.no <- index$category.accuracy[2]
# Se guarda la prediccion en una variable y se le añade al dataset de desempleados
Prediccion <- as.factor(Prediccion)
resultados3 <- cbind (desempleado.3, Prediccion)
# Se unen los datasets de desempleados y ocupados
resultados3.1 <- rbind(resultados3, ocupado)
# Se evalúan los cambios con las simulaciones
tab1(resultados3.1$Prediccion, sort.group = "decreasing", graph = FALSE)
tab1(resultados3.1$Condicion_actividad, sort.group = "decreasing", graph = FALSE)
# se guardan los resultados de la simulación en tipo csv
#write.csv(resultados3.1, "Caso3_simulado.csv")

```

Simulación 3.1: Mujeres

```

desempleado.3.2 <- subset(resultados3.1, Sexo == "Mujer")
tab1(desempleado.3.2$Prediccion, sort.group = "decreasing", graph = FALSE)
tab1(desempleado.3.2$Condicion_actividad, sort.group = "decreasing", graph = FALSE)

```

Simulación 3.2: Hombres

```

desempleado.3.3 <- subset(resultados3.1, Sexo == "Hombre")
tab1(desempleado.3.3$Prediccion, sort.group = "decreasing", graph = FALSE)
tab1(desempleado.3.3$Condicion_actividad, sort.group = "decreasing", graph = FALSE)

```

Simulación 3.3: Zona Urbana

```

desempleado.3.4 <- subset(resultados3.1, Zona == "Urbana")
tab1(desempleado.3.4$Prediccion, sort.group = "decreasing", graph = FALSE)
tab1(desempleado.3.4$Condicion_actividad, sort.group = "decreasing", graph = FALSE)

```

Simulación 3.4: Zona Rural

```

desempleado.3.5 <- subset(resultados3.1, Zona == "Rural")
tab1(desempleado.3.5$Prediccion, sort.group = "decreasing", graph = FALSE)
tab1(desempleado.3.5$Condicion_actividad, sort.group = "decreasing", graph = FALSE)

```

Simulación 3.5: 25 años o menos

```
desempleado.3.6 <- subset(resultados3.1, Edad <= 25)
tab1(desempleado.3.6$Prediccion, sort.group = "decreasing", graph = FALSE)
tab1(desempleado.3.6$Condicion_actividad, sort.group = "decreasing", graph =
FALSE)
```

Simulación 3.6: 35 años o menos

```
desempleado.3.7 <- subset(resultados3.1, Edad <= 35)
tab1(desempleado.3.7$Prediccion, sort.group = "decreasing", graph = FALSE)
tab1(desempleado.3.7$Condicion_actividad, sort.group = "decreasing", graph =
FALSE)
```

Simulación 3.7: 36 años o mas

```
desempleado.3.8 <- subset(resultados3.1, Edad > 35)
tab1(desempleado.3.8$Prediccion, sort.group = "decreasing", graph = FALSE)
tab1(desempleado.3.8$Condicion_actividad, sort.group = "decreasing", graph =
FALSE)
```

7. Referencias

- Acemoglu, D y Restrepo, P. (2020). *Robots and Jobs: Evidence from US Labor Markets*. Journal of Political Economy, MIT. <https://economics.mit.edu/sites/default/files/publications/Robots%20and%20Jobs%20-%20Evidence%20from%20US%20Labor%20Markets.p.pdf>
- Amat, R. (2016). Validación de modelos predictivos: Cross-validation, OneLeaveOut, Bootstrapping. Sitio web: https://www.cienciadedatos.net/documentos/30_cross-validation_oneleaveout_bootstrap#K-Fold_Cross-Validation
- Athey, S (2018). *The Impact of Machine Learning on Economics*. Stanford University Sitio web: <https://www.gsb.stanford.edu/faculty-research/publications/impact-machine-learning-economics>
- Becker, G. (1975). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education, Second Edition*. NBER. Sitio web: <https://www.nber.org/books/beck75-1>
- Becker, G. (1993). *Nobel Lecture: The Economic Way of Looking at Behavior*. Journal of Political Economy. <https://www.jstor.org/stable/2138769>
- Benavides, J. (2020). *Tópicos en Investigación: Odds Ratio*. Fundación Universitaria Sanitas. Sitio web: <https://revistas.unisanitas.edu.co/index.php/rms/article/view/171/133>
- Blanchard, O. y Johnson, D. (2013). *Macroeconomics*. Pearson.
- Brown, S y Sessions, J. (1997). *A Profile of UK Unemployment: Regional versus Demographic Influences*. Regional Studies. Sitio web: <https://www.tandfonline.com/action/showCitFormats?doi=10.1080%2F00343409750132964>
- Burn, I. Button, P. y Neumark, D. (2017). *Age Discrimination and Hiring of Older Workers*. FRBSF Economic Letter, Federal Reserve Bank of San Francisco. <https://www.frbsf.org/wp-content/uploads/sites/4/el2017-06.pdf>
- Burt, R. (2000). *The Network Structure of Social Capital*. Research in Organizational Behavior. <https://www.sciencedirect.com/science/article/abs/pii/S0191308500220091>
- Ceballos, O. (2019). *Principios básicos de microeconometría y del uso de variables*

- instrumentales para la inferencia causal*. De Análisis Económico. Sitio web: https://www.researchgate.net/publication/334071127_Principios_basicos_de_microeconomia_y_del_uso_de_variables_instrumentales_para_la_inferencia_causal/link/5d1f7c3e458515c11c151b90/download
- Coa, M & Ponsot, E. (2019). *Funciones de enlace alternativas en modelos de respuesta binomial*. 2019, de Research Gate Sitio web: https://www.researchgate.net/profile/Ernesto_Ponsot_Balaguer/publication/332684248_Alternative_link_functions_in_binomial_response_models/links/5cc320d24585156cd7b44a72/Alternative-link-functions-in-binomial-response-models.pdf
- Contreras, O. (2019). *Support Vector Machines for Classification*. Towards Data Science. Sitio web: <https://towardsdatascience.com/support-vector-machines-for-classification-fc7c1565e3>
- Dhiman. D, Saraogi, V. Devatha, V. (2019). Modeling Youth Unemployment. Sitio web: <https://towardsdatascience.com/https-medium-com-vikramdevatha-modeling-youth-unemployment-d0f7cbcd078a>
- ENZYME. (s.f). *Redes Neuronales Artificiales y Deep Learning, explicado para dummies*. ENZYME Advising Group. Sitio web: <https://blog.enzymeadvisinggroup.com/redes-neuronales-artificiales-y-deep-learning>
- Fernández, E. (2022). Lineamientos para la Empleabilidad en la Región Central de Costa Rica: Sector de Servicios. Universidad Nacional de Costa Rica. Sitio web: https://repositorio.una.ac.cr/bitstream/handle/11056/24476/TFG_Erick%20Fern%C3%A1ndez%20D_%28092022%29.pdf?sequence=1&isAllowed=y
- Gallardo. (2020). *Métodos Jerárquicos de Análisis Cluster*. Universidad de Granada. Recuperado de: <https://www.ugr.es/~gallardo/pdf/cluster-3.pdf>
- Glauber, R. (2007). *Marriage and the Motherhood Wage Penalty Among African Americans, Hispanics, and Whites*. Journal of Marriage and Family. https://www.researchgate.net/publication/229724149_Marriage_and_the_Motherhood_Wage_Penalty_Among_African_Americans_Hispanics_and_Whites

- Grabois, J. (2014). *Precariedad laboral, exclusión social y economía popular*. Pontifical Academy of Sciences. Sitio web: <http://www.pas.va/content/dam/accademia/pdf/es41/es41-grabois.pdf>
- Hacienda. (2019). *Déficit 2018 baja en 1,2% del PIB en relación con lo proyectado*. 2019, de Ministerio de Hacienda Sitio web: <https://www.hacienda.go.cr/noticias/14921-deficit-2018-baja-en-12-del-pib-en-relacion-con-lo-proyectado>
- Heilman, M. (2012). *Gender stereotypes and workplace bias*. Research in Organizational Behavior. <https://www.sciencedirect.com/science/article/abs/pii/S0191308512000093>
- Hernández, R. (2014). *Metodología de la investigación*. México: McGraw-Hill / Interamericana Editores S.A. de C.V.
- Hosmer, D & Lemeshow, S. (2004). *Applied Logistic Regression*. https://ftp.idu.ac.id/wp-content/uploads/ebook/ip/REGRESI%20LOGISTIK/epdf.pub_applied-logistic-regression-wiley-series-in-probab.pdf
- INEC. (2012). *Encuesta Continua de Empleo: Métodos y Procedimientos*. De Instituto Nacional de Estadística y Censos Sitio Web: https://www.inec.cr/sites/default/files/documentos-biblioteca-virtual/meecemetodos_01.pdf
- INEC. (2022). *Encuesta Continua de Empleo*. De INEC Sitio web: http://sistemas.inec.cr/pad4/index.php/catalog/246/get_microdata
- Kiziryan, M. (2018). *Mercado Laboral*. De Economipedia Sitio web: <https://economipedia.com/definiciones/mercado-laboral.html>
- Kutuk, Y. (2019). Prediction of Transition Probabilities from Unemployment to Employment for Turkey Via Machine Learning and Econometrics: A Comparative Study. Sitio web: <https://dergipark.org.tr/en/pub/jore/issue/44898/559016>
- Loria, M & Martínez, J. (2016). El desempleo en Costa Rica: evolución reciente y principales características. De Academia de Centroamérica Sitio web: <https://www.academiaca.or.cr/wp-content/uploads/2017/02/Desempleo-en-Costa-Rica.pdf>

- Marín, J. (2014). *Regresión Logística*. 2019, de Universidad Carlos III de Madrid Sitio web: <http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/DM/tema2dm.pdf>
- Márquez, C. (2015). *Determinantes del desempleo en las urbes mexicanas. Continuidades y rupturas en el periodo de crisis*. De Papeles de Población, Toluca. Sitio Web: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-74252015000100005
- Martínez, I. (2018). *Predicting, explaining, and understanding risk of long-term unemployment*. De Neural Information Processing Systems Sitio web: https://aiforsocialgood.github.io/2018/pdfs/track1/97_aig_neurips2018.pdf
- Martínez, J y Solórzano, M. (2018). *Determinantes del desempleo juvenil en el Ecuador*, año 2016. Universidad Central del Ecuador. Recuperado de: <http://www.dspace.uce.edu.ec/bitstream/25000/14452/1/T-UCE-0005-ES011-2018.pdf>
- Mc Gregor, A. (1978). Unemployment Duration and Re-employment Probability. The Economic Journal Vol 88 No 352, Diciembre, pp 693-706. <https://www.jstor.org/stable/pdf/2231973.pdf?refreqid=excelsior%3Aa7caca67e43c19b8a578ffb9d2d7acad>
- MIDEPLAN. (2018). *Plan Nacional de Desarrollo y de Inversión Pública del Bicentenario. 2019*, de MIDEPLAN Sitio web: <https://observatorioplanificacion.cepal.org/sites/default/files/plan/files/Costa%20Rica%20PNDIP%20%202019-2022.pdf>
- Morales, N y Segura, R (2018) *Barreras al acceso al mercado laboral y predicción de movilidad laboral entre sectores económicos con enfoque de género*, del Informe Estado de la Nación 2018. Sitio web: <http://repositorio.conare.ac.cr/handle/20.500.12337/2972>
- Morales, R. (2017). *Desempleo un Problema Estructural en Costa Rica*. 2019, de CAMPUS UNA Sitio web: http://www.campus.una.ac.cr/ediciones/2017/noviembre/2017noviembre_pag23a.html

- MTSS. (2014). Estrategia Nacional de Empleo y producción. 2019, de MTSS Sitio web: <http://www.mtss.go.cr/elministerio/despacho/Estrategia%20Nacional%20de%20Empleo%20y%20Produccion.pdf>
- Nagelkerke, N. (1991). A Note on a General Definition of the Coefficient of Determination *Biometrika*.
https://www.cesarzamudio.com/uploads/1/7/9/1/17916581/nagelkerke_n.j.d._1991_-_a_note_on_a_general_definition_of_the_coefficient_of_determination.pdf
- Núñez, F. Rodríguez, A y Usabiaga ,C. (2014). Segmentación laboral y análisis de clusters una aplicación al mercado de trabajo cualificado andaluz. De Centro de Estudios Andaluces. Recuperado de <https://dialnet.unirioja.es/servlet/articulo?codigo=5769340>
- OCDE. (2016). *OECD Economic Surveys: Costa Rica 2016*. 2019, de OCDE Sitio web: https://read.oecd-ilibrary.org/economics/oecd-economic-surveys-costa-rica-2016_eco_surveys-cri-2016-en#page1
- OECD. (2014). *¿Los jóvenes de 15 años son creativos a la hora de resolver problemas? PISA IN FOCUS 38*. En: <http://www.mecd.gob.es/dctm/inee/pisa-in-focus/pisa-in-focus-n38-esp-1-4-2014.pdf?documentId=0901e72b81904916>
- Organización Internacional del Trabajo. (2015). *Trabajo decente*. OIT. Recuperado de: <https://www.ilo.org/global/topics/decent-work/lang--es/index.htm>
- Pacheco, J. (2013). *Desempleo Juvenil en Costa Rica*. 2019, de Estadonacion.or.cr Sitio web: https://estadonacion.or.cr/files/biblioteca_virtual/019/Pacheco_2013.pdf
- Parra, A. (2019). *El Pensamiento Computacional*. Rock Studio. En <https://rockcontent.com/es/blog/pensamiento-computacional/>
- Reich, M, Gordon,D y Edwards,R. (1973). *Dual Labor Markets: A Theory of Labor Market Segmentation*. *American Economic Review* Vol 63. Sitio web: https://www.researchgate.net/publication/4721732_Dual_Labor_Markets_A_Theory_of_Labor_Market_Segmentation

- Reich, M. Gordon, D. y Edwards, C. (1973). *A Theory of Labor Market Segmentation*. The American Economic Review. <https://www.jstor.org/stable/1817097>
- Reich, R. (1991). *The Work of Nations: Preparing Ourselves for 21st Century Capitalism*. Alfred A. Knopf.
- Salamanca, S. (2015). *Duración y factores determinantes del desempleo en Medellín*. Universidad Nacional de Colombia. Recuperado de <http://bdigital.unal.edu.co/52343/1/sandraisabelsalamancagil.2015.pdf>
- Sánchez, M. (2014). Determinantes de las fluctuaciones del desempleo en Costa Rica a partir de la estimación de la curva de Beveridge y de un Índice de desemparejamiento de habilidades. 2019, de Universidad de Costa Rica Sitio web: <http://repositorio.sibdi.ucr.ac.cr:8080/jspui/bitstream/123456789/2364/1/37565.pdf>
- Sancho, F. (2017). *Introducción al Aprendizaje Automático*. De Universidad de Sevilla Sitio web: <http://www.cs.us.es/~fsancho/?e=75>
- Sen, A. (2000). *Desarrollo y Libertad*. Buenos Aires, Argentina. Editorial Planeta.
- Serneels, P. (2001). *Unemployment duration in urban Ethiopia*. Center for the study of African Economics, Oxford University. Sitio Web: https://www.researchgate.net/publication/238703341_Unemployment_Duration_in_Urban_Ethiopia
- Shutaywi, M y Kachouie, N. (2021). *Silhouette Analysis for Performance Evaluation in Machine Learning with Applications to Clustering*. National Library of Medicine. Sitio Web: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8234541/>
- Van Winkle, Z., Taylor, J., y Higgins, M. (2018). *The Role of Rural Communities in the Global Supply of Labor: Evidence from Chinese Coastal Villages*. Journal of Development Economics, 135, 349-367

- Varian, H. (2014). *Big Data: New Tricks for Econometrics*. De Journal of Economic Perspectives. Recuperado de: <https://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.28.2.3>
- Vicente, S. (2015). *Introducción a la macroeconomía*. Apartado 8.4 “tipos de desempleo”, páginas: 29-31. De la Universidad del país vasco. Sitio web: <https://ocw.ehu.eus/mod/folder/view.php?id=3762>
- Villegas, B y Ulate, J. (2013). Análisis dinámico del desempleo en Costa Rica para el período 2010-2011. 2019, de Universidad de Costa Rica Sitio web: <https://revistas.ucr.ac.cr/index.php/economicas/article/download/10618/10015/>
- W.H, Thomas y Feldman, D. (2011). *Affective organizational commitment and citizenship behavior: Linear and non-linear moderating effects of organizational tenure*. Journal of Vocational Behavior. <https://www.sciencedirect.com/science/article/abs/pii/S0001879111000303?via%3Dihub>
- Wing, J. (2006). *Computational Thinking*. Carnegie Mellon University School of Computer Science. Sitio Web: <https://www.cs.cmu.edu/~15110-s13/Wing06-ct.pdf>
- Winters, J. (2010). *Human Capital and Population Growth in Non-Metropolitan U.S. Counties: The Importance of College Student Migration*. Munich Personal RePEc Archive. https://mpa.ub.unimuenchen.de/25592/1/Human_Capital_and_Population_Growth_in_Non-Metropolitan_Counties.pdf
- Xu, N, Finkelman, R, Dai, S y Peng, M. (2021). *Average Linkage Hierarchical Clustering Algorithm for Determining the Relationships between Elements in Coal*. National Library of Medicine. Sitio web: <https://pubmed.ncbi.nlm.nih.gov/33718711/>