

# **Análisis de la demanda laboral de profesionales TI mediante algoritmos de machine learning**

Joao Carlo Ayzanoa Solano



# Indice

➤ INTRODUCION	01
➤ DESARROLLO	02
➤ Estado del arte	2.1
➤ Metodología y aporte teórico	2.2 05
➤ Aporte practico	2.3
➤ Resultados y analisis	2.4
➤ CONCLUSIONES	03
➤ Referencias bibliograficas	





## 01. INTRODUCCION

La demanda de profesionales en Tecnologías de la Información (TI) crece constantemente. Este trabajo analiza esa demanda utilizando algoritmos de machine learning, con el objetivo de identificar habilidades requeridas y tendencias del mercado laboral actual.



## 1.1 Planteamiento del problema

Existe poca información automatizada sobre qué habilidades y perfiles TI demanda el mercado. Esto dificulta la toma de decisiones para profesionales, instituciones y empresas.

## 1.2 Objetivo específicos:

- Recopilar datos de ofertas laborales TI.
- Aplicar machine learning para clasificarlas.
- Identificar habilidades y tendencias del mercado.



## 02. DESARROLLO

### 2.1 Estado del arte

Se revisaron tres artículos que aplican machine learning para analizar el mercado laboral TI. Todos destacan el uso de algoritmos para predecir empleabilidad, clasificar habilidades y detectar tendencias del sector.

N	Autores	Título	Fuente	Año
1	Zoraida Mamani Rodríguez	Proceso de <i>machine learning</i> para determinar la demanda social de puestos de empleo de profesionales de TI	<i>Industrial Data</i> , Vol. 25(2), pp. 275–287	2022
2	Gehad ElSharkawy, Yehia Helmy, Engy Yehia	Employability Prediction of Information Technology Graduates Using <i>Machine Learning</i> Algorithms	<i>International Journal of Advanced Computer Science and Applications (IJACSA)</i> , Vol. 13(10), pp. 359–366	2022
3	Nik Dawson, Marian-Andrei Rizoiu, Benjamin Johnston, Mary-Anne Williams	Predicting Skill Shortages in Labor Markets: A <i>Machine Learning</i> Approach	Proceedings of the IEEE International Conference on Big Data (BigData 2020)	2020

## 2.2 Metodología y aporte teórico

### Selección y justificación de la(s) tecnología(s) (Comparación de las tecnologías encontradas en la revisión del estado del arte)

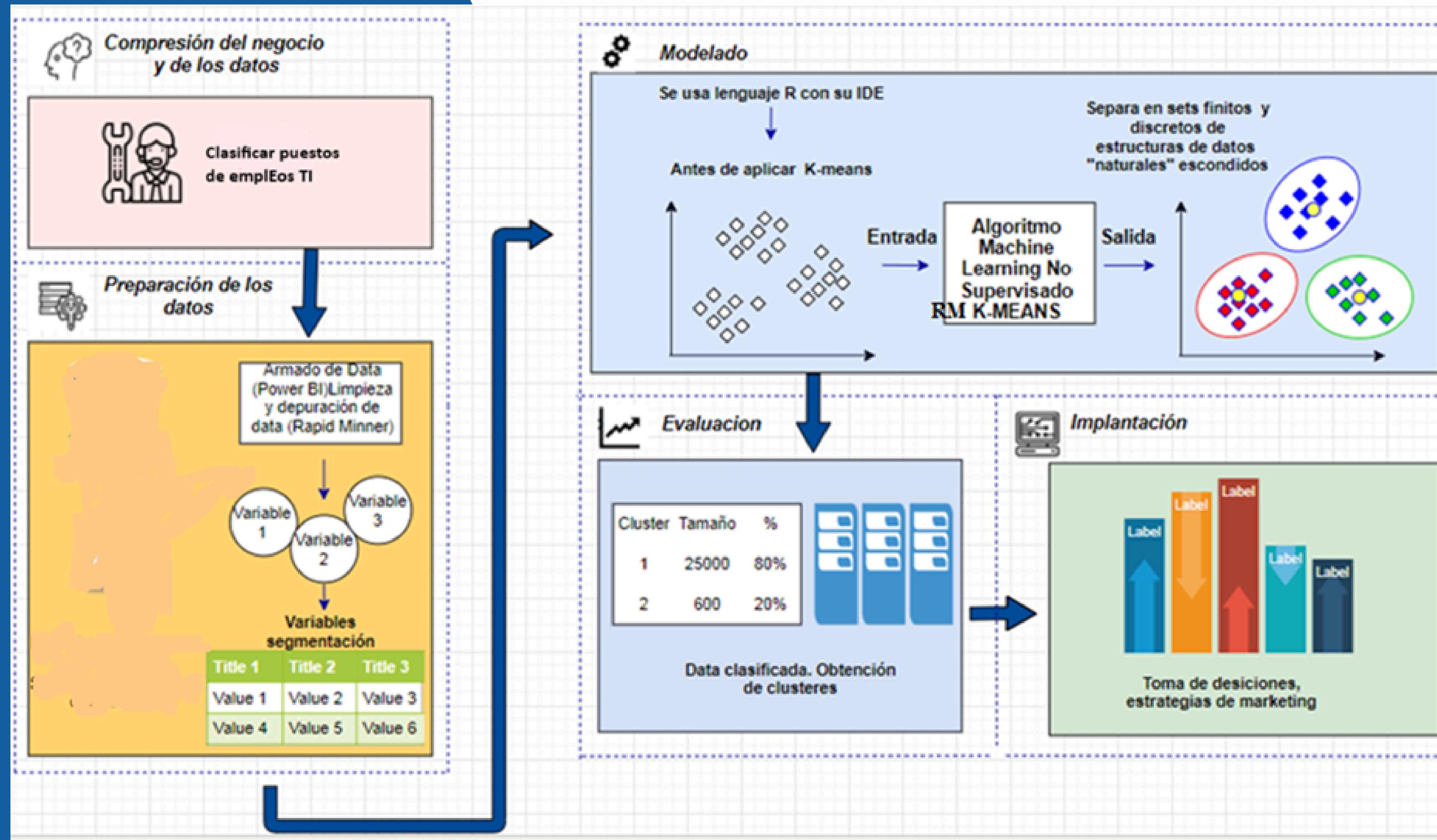
Autores	Objetivo del trabajo	Métodos/ Técnicas utilizados	Área de aplicación o contexto
Mamani Rodríguez, Zoraida (2022)	Diseñar un proceso de machine learning que permita determinar la demanda social de puestos laborales en el sector de TI.	Machine Learning No Supervisado: Aplicación del algoritmo k-means	Empleabilidad, mercado laboral en tecnología de la información (TI).
ElSharkawy, Helmy & Yehia (2022)	Predecir la empleabilidad de egresados de TI mediante algoritmos de aprendizaje automático.	Algoritmos supervisados como Decision Tree, Naïve Bayes, SVM, Logistic Regression y Random Forest.	Educación, empleabilidad en Tecnologías de la Información.
Dawson, Rizou, Johnston & Williams (2020)	Predecir la escasez de habilidades en el mercado laboral utilizando datos históricos mediante técnicas de aprendizaje automático.	Algoritmos de ML como LSTM, Random Forest y técnicas de series de tiempo; uso de representaciones semánticas para clasificar habilidades.	Ánalysis de mercado laboral, planificación de políticas públicas.

#### Clustering con k-means (Mamani, 2022):

- Agrupa puestos TI por similitud.
- Método no supervisado (sin datos etiquetados).
- Datos reales + herramientas accesibles (Python, Power BI).

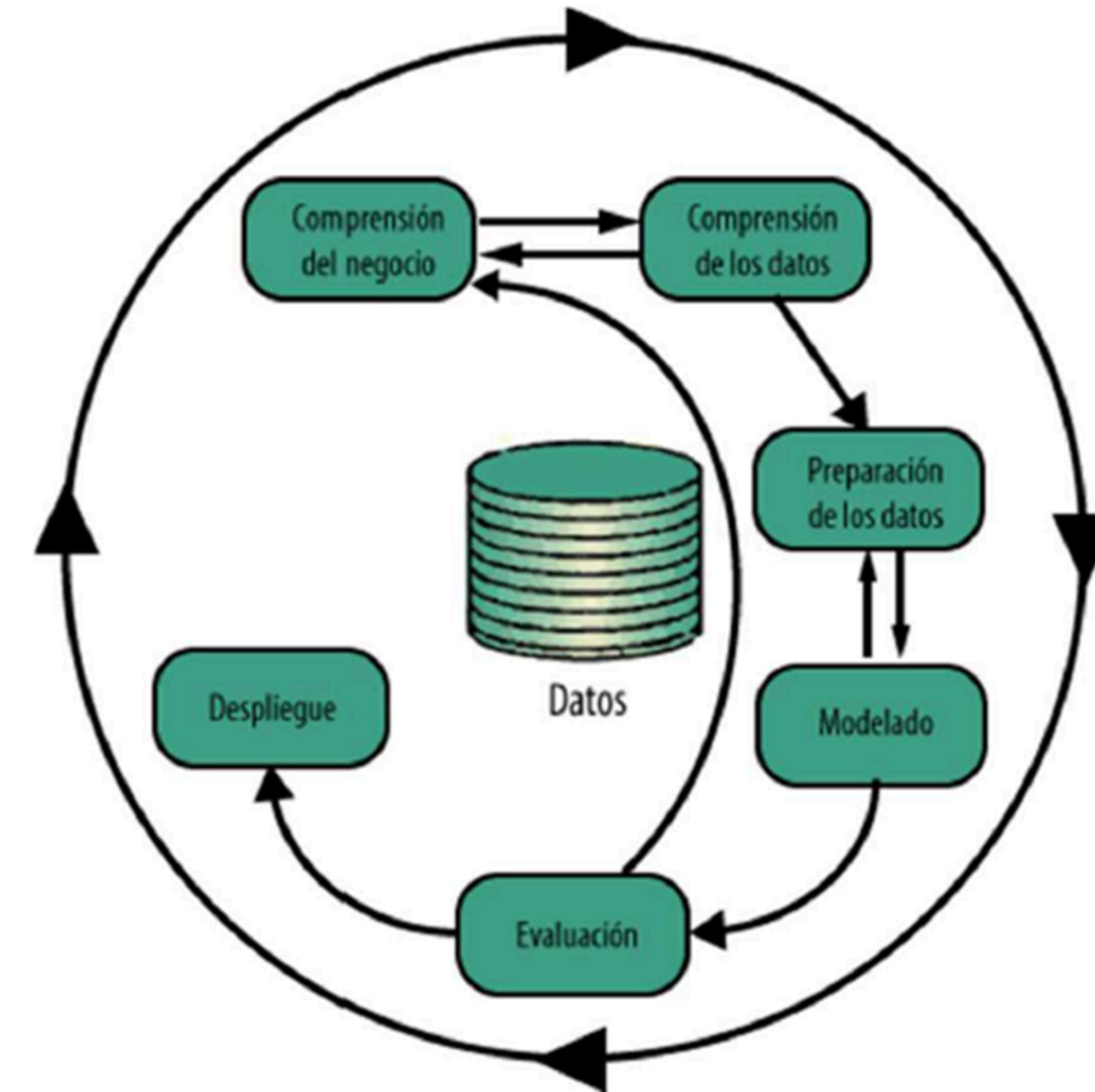
## 2.2 Metodología y aporte teorico

### Definición de la propuesta de solución (Adaptación de la tecnología)



## 2.2 Metodología y aporte teorico

### Metodología a seguir para el desarrollo



## 2.3 Aporte practico

Para el diseño e implementación del aporte se usaron las siguientes herramientas:



## Comprensión del negocio y de los datos.

### Preparación de los datos.

#### Modelado

A	B	C	D	E	F	G
1	Puesto,UbicaciÃ³n,Funciones,Competencias,Habilidades					
2	Tester QA,Lima,GestiÃ³n de redes,C#,PrecisiÃ³n					
3	Ingeniero de Software,Trujillo,Desarrollo e implementaciÃ³n de software,React,ComunicaciÃ³n					
4	Programador Python,Piura,Desarrollo web,SAP,EmpatÃ-a					
5	Desarrollador Fullstack,Piura,DiseÃ±o de sistemas,SQL,Liderazgo					
6	Especialista en Ciberseguridad,Lima,Pruebas funcionales,Windows Server,AutonomÃ-a					
7	Administrador de Redes,Cusco,DiseÃ±o de sistemas,SQL,EmpatÃ-a					
8	Consultor SAP,Piura,Pruebas funcionales,Selenium,ComunicaciÃ³n					
9	Programador Python,Arequipa,Seguridad informÃitica,Firewall,AnÃlisis					
10	Consultor SAP,Trujillo,GestiÃ³n de redes,SQL,AnÃlisis					
11	Ingeniero de Software,Piura,AutomatizaciÃ³n de scripts,Java,Creatividad					
12	Analista de Datos,Trujillo,ImplementaciÃ³n de ERP,Selenium,Pensamiento crÃ-tico					
13	Programador Python,Arequipa,Seguridad informÃitica,C#,Liderazgo					
14	Desarrollador Frontend,Lima,ImplementaciÃ³n de ERP,Firewall,ResoluciÃ³n de problemas					
15	Consultor SAP,Trujillo,Desarrollo e implementaciÃ³n de software,React,Pensamiento crÃ-tico					
16	Administrador de Redes,Cusco,Seguridad informÃitica,SQL,AnÃlisis					
17	Consultor SAP,Piura,Desarrollo e implementaciÃ³n de software,Windows Server,AutonomÃ-a					
18	Soporte TÃ©cnico,Trujillo,DiseÃ±o de sistemas,Firewall,PrecisiÃ³n					
19	Tester QA,Cusco,Desarrollo e implementaciÃ³n de software,C#,ComunicaciÃ³n					
20	Administrador de Redes,Lima,Soporte a usuarios,Java,ComunicaciÃ³n					
21	Desarrollador Frontend,Lima,ImplementaciÃ³n de ERP,SAP,AnÃlisis					
22	Desarrollador Fullstack,Piura,DiseÃ±o de sistemas,SQL,EmpatÃ-a					

**• Para este parte luego de haber hecho scrapping y tener nuestra data limpia y ordenada, lo vamos a tener en un archivo.csv**

**• Esa seria nuestra data con 30 ofertas laborales extraída luego de los pasos anteriores**

**• Esto lo llevamos a R, tenemos q cargar ese archivo y luego vamos a proceder a usar cÃdigo R, para llevar a cabo esta clasificaciÃ³n automÃtica.**

## Modelado

Código en R usando algoritmos kmeans y el archivo Excel donde tenemos nuestra data limpia lista para entrar al modelo:

```
1 # Instalar solo si no lo tienes (una sola vez)
2 install.packages("tidyverse")
3 install.packages("cluster")
4 install.packages("factoextra")
5
6 # Cargar librerías
7 library(tidyverse)
8 library(cluster)
9 library(factoextra)
10
11 # Leer archivo CSV (debe estar en la misma carpeta)
12 datos <- read.csv("ofertas_laborales_30.csv")
13
14 # Ver primeros datos
15 head(datos)
16
17 # Codificar variables categóricas como variables numéricas
18 datos_cod <- datos %>%
19   select(Puesto, Ubicación, Funciones, Competencias, Habilidades) %>%
20   mutate(across(everything(), as.factor)) %>%
21   model.matrix(~ . -1, data = .) %>%
22   as.data.frame()
23
24 # Determinar número óptimo de clusters (método del codo)
25 fviz_nbclust(datos_cod, kmeans, method = "wss")
26
27 # Aplicar k-means con 3 clusters
28 set.seed(123)
29 modelo <- kmeans(datos_cod, centers = 3, nstart = 25)
30
31 # Mostrar resumen
32 print(modelo)
33
34 # Visualizar clusters
35 fviz_cluster(modelo, data = datos_cod)
```

## 2.4 Resultados y análisis

K-means clustering with 3 clusters of sizes 6, 11, 13

Cluster means:

	PuestoAdministrador de Redes	PuestoAnalista de Datos	PuestoConsultor SAP
1	0.0000000	0.16666667	0.33333333
2	0.0000000	0.00000000	0.18181818
3	0.2367592	0.07692308	0.07692308

	PuestoDesarrollador Frontend	PuestoDesarrollador Fullstack
1	0.0000000	0.00000000
2	0.0000000	0.09090909
3	0.1538462	0.00000000

	PuestoEspecialista en Ciberseguridad	PuestoIngeniero de Software
1	0.0000000	0.16666667
2	0.0000000	0.09090909
3	0.07692308	0.07692308

	PuestoProgramador Python	PuestoSoporte Técnico	PuestoTester QA	UbicaciónCusc
0	0.0000000	0.33333333	0.0000000	0.0000000
1	0.54545455	0.09090909	0.0000000	0.0909090
2	0.37692308	0.07692308	0.1538462	0.3846153
3	0.0000000	0.0000000	0.0000000	0.0000000
4	0.0000000	0.0000000	0.0000000	0.0000000
5	0.0000000	0.0000000	0.0000000	0.0000000
6	0.0000000	0.0000000	0.0000000	0.0000000
7	0.0000000	0.0000000	0.0000000	0.0000000
8	0.0000000	0.0000000	0.0000000	0.0000000

	UbicaciónLima	UbicaciónPiura	UbicaciónTrujillo	FuncionesAutomatización de scripts
1	0.0000000	0.6000000	1	0.000
2	0.0000000	0.7272727	0	0.090
3	0.4615385	0.0000000	0	0.000
4	0.0000000	0.0000000	0	0.000
5	0.0000000	0.0000000	0	0.000
6	0.0000000	0.0000000	0	0.000
7	0.0000000	0.0000000	0	0.000
8	0.0000000	0.0000000	0	0.000

FuncionesDesarrollo e implementación de software FuncionesDesarrollo web

Clustering vector:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
27	28																									
3	1	2	2	3	3	2	2	1	2	1	2	3	1	3	2	1	3	3	3	2	3	2	2	1	3	
3	3																									
29	30																									
2	3																									

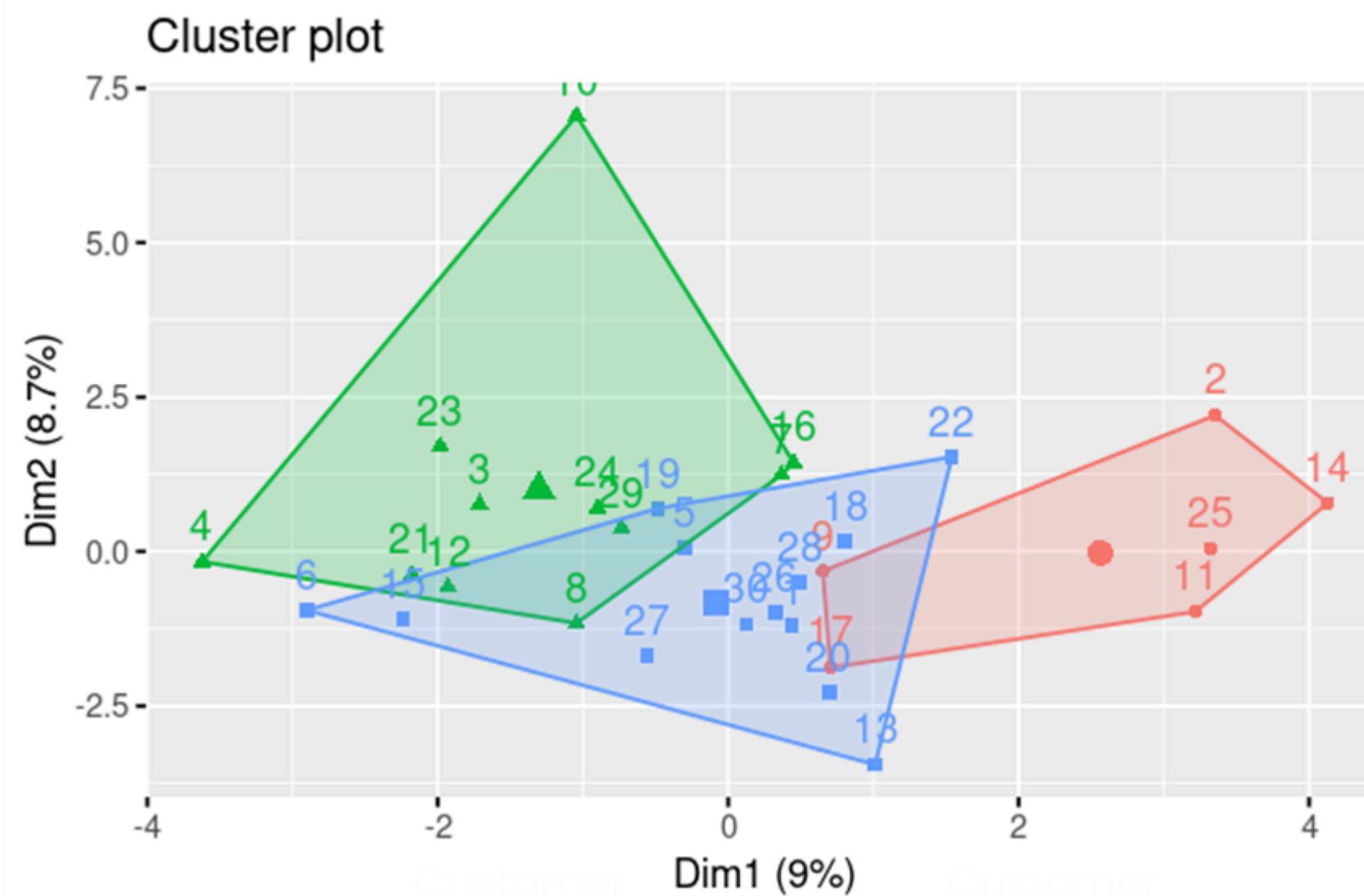
Within cluster sum of squares by cluster:

[1] 15.33333 33.27273 44.15385  
(between\_SS / total\_SS = 17.9 %)

Available components:

[1] "cluster" "centers" "totss" "withinss" "tot.withinss"  
[6] "betweenss" "size" "iter" "ifault"  
> # Visualizar clusters  
> fviz\_cluster(modelo, data = datos\_cod)  
Session restored from your saved work on 2025-Jun-23 08:12:17 UTC (2 hours ago)  
>  
> View(modelo)

## 2.4 Resultados y análisis



Cada punto representa una oferta laboral.  
El color representa el cluster (perfil laboral).  
Los grupos muestran tendencias en habilidades y competencias.

cluster	Cluster	Características principales	Perfil detectado
1	1	SAP, pensamiento crítico, React	Backend avanzado
2	2	Python, liderazgo, Piura	Datos + gestión
3	3	Soporte, trabajo en equipo, Lima	Técnico operativo

- El mercado laboral muestra tres perfiles dominantes.
  - Hay una clara segmentación entre roles técnicos, de datos y de soporte.
  - Esta metodología puede ayudar a orientar la formación profesional y la toma de decisiones en talento humano.

# 03. Conclusiones

Reflexión sobre la importancia del análisis automatizado del mercado laboral TI.

Valor del enfoque propuesto con ML.

Potencial aplicación en asesoría vocacional, políticas públicas, o desarrollo curricular.



## **REFERENCIAS BIBLIOGRAFICAS**

- Mamani Rodríguez, Z. (2022).  
Proceso de machine learning para determinar la demanda social de puestos de empleo de profesionales de TI. *Revista Industrial Data*, 25(2), 275–300.  
DOI: 10.15381/idata.v25i2.21643
- Arthur, D., & Vassilvitskii, S. (2007).  
k-means++: The advantages of careful seeding. *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms*, 1027–1035.  
URL: <https://dl.acm.org/doi/10.5555/1283383.1283494>
- Dawson, N., Rizoiu, M.-A., Johnston, B., & Williams, M.-A. (2020).  
Predicting Skill Shortages in Labor Markets: A Machine Learning Approach. *arXiv:2004.01311v3*.  
URL: <https://arxiv.org/abs/2004.01311>
- ElSharkawy, G., Helmy, Y., & Yehia, E. (2022).  
Employability Prediction of Information Technology Graduates Using Machine Learning Algorithms. *International Journal of Advanced Computer Science and Applications*, 13(10).  
DOI: 10.14569/IJACSA.2022.0131036

# THANK YOU!

