

## Predicción del empleo formal en la Argentina aplicando algoritmos de aprendizaje automático

Fernando Alvarez<sup>1</sup> 0009-0006-3903-8948

Universidad Nacional del Nordeste, Resistencia, Chaco, Argentina  
`fernando.alvarez@comunidad.unne.edu.ar`

**Resumen** El empleo formal mejora la calidad de vida al garantizar derechos y beneficios legales. En Argentina, donde gran parte de la población trabaja informalmente, los datos oficiales sobre el mercado laboral presentan retrasos significativos, obstaculizando la implementación oportuna de políticas públicas. Este estudio aplica algoritmos de Aprendizaje Automático a datos de la Encuesta Permanente de Hogares para predecir la condición de formalidad laboral. Siguiendo la metodología CRISP-DM, se implementó imputación múltiple para datos faltantes y se compararon diversos modelos predictivos. XGBoost destacó por su precisión y equilibrio entre sensibilidad y especificidad. Los resultados demuestran la utilidad de modelos predictivos avanzados para anticipar cambios en el empleo formal, facilitando la toma de decisiones más ágiles y efectivas en materia de políticas laborales.

**Palabras clave:** empleo, rotación laboral, predicción, algoritmos de aprendizaje automático, empleo formal

## Predicting Formal Employment in Argentina Using Machine Learning Algorithms

**Abstract.** Formal employment is associated with improved quality of life, as it ensures access to legal protections and benefits. In Argentina, a large proportion of the workforce is in the informal sector, and official labor market data experience significant delays in publication, hindering timely policy implementation. This work applies Machine Learning algorithms to Permanent Household Survey (EPH) data to predict formal employment status. Following CRISP-DM methodology, we implemented multiple imputation for missing data and compared various predictive models. XGBoost demonstrated superior performance with balanced sensitivity and specificity metrics. The results highlight the value of advanced predictive models for anticipating changes in formal employment, enabling more agile and effective labor policy decisions.

**Keywords:** employment, labor turnover, prediction, machine learning, formal employment

## 1 Introducción

El empleo formal suele asociarse con una mejor calidad de vida, dado que está vinculado a una serie de derechos y beneficios amparados por la legislación vigente. Entre estos se destacan un mayor acceso a servicios de salud mediante la cobertura de obras sociales, salarios más altos, la percepción de un sueldo anual complementario (aguinaldo) y otros beneficios que favorecen a quienes integran este sector de la población.

## 2 Revisión de la literatura

### 2.1 Formalidad laboral y rotación laboral

Es fundamental distinguir entre dos conceptos clave: formalidad empresarial y formalidad laboral. La formalidad empresarial se refiere al registro de actividades de una empresa ante el Estado, lo que implica el cumplimiento de normativas fiscales y el pago de impuestos. Por otro lado, la formalidad laboral está asociada con el acceso de los trabajadores a derechos y protecciones sociales garantizados por la ley, como cobertura médica, salarios complementarios y niveles salariales superiores, entre otros beneficios (Rofman, 2007).

La formalidad laboral, que constituye el eje de esta investigación, no suele estar explícitamente detallada en las encuestas. En su lugar, se utiliza un indicador común para inferirla: la presencia de descuentos jubilatorios. Si un trabajador reporta tener descuentos jubilatorios, esto generalmente indica que su empleo es formal, ya que estas contribuciones están asociadas con trabajos registrados y formalizados dentro del sistema de seguridad social.

En cuanto a la rotación laboral, este concepto se entiende como el cambio de un trabajador entre puestos internos o entre empresas, así como la transición entre el empleo y el desempleo, ya sea de forma temporal o definitiva (Abbasi & Hollman, 2000). Para nuestro análisis, nos interesan especialmente los casos en los que una persona desempleada pasa a ser empleada de manera formal, y aquellos en los que un trabajador formal pierde su empleo.

La rotación laboral puede analizarse desde dos ejes importantes: la tasa de entrada, que se refiere a la cantidad de personas que pasan de un estado de “no formal” a un estado de formalidad, y la tasa de salida, que mide la cantidad de personas que dejan un trabajo formal. Los trabajadores pueden abandonar o perder su puesto de trabajo por una variedad de factores, agrupados en tres categorías: factores personales (edad, sexo, estado civil, nivel educativo), factores organizacionales (cultura organizacional, beneficios, salarios) y factores socio-económicos (nivel de actividad económica) (Zhang, 2016).

### 2.2 Aprendizaje Automático aplicado al mercado laboral

El uso de algoritmos de Aprendizaje Automático ha crecido significativamente en los últimos años, extendiéndose a diversas áreas, incluido el mercado laboral.

Samuel (1959) lo define como “el campo de estudio que otorga a las computadoras la habilidad de aprender sin ser programadas explícitamente”. En una definición más reciente, se entiende al Aprendizaje Automático como “una subdisciplina de la inteligencia artificial que se centra en el desarrollo de sistemas que pueden aprender de los datos para realizar predicciones o tomar decisiones sin depender de reglas programadas explícitamente” (Jordan & Mitchell, 2015).

Un ejemplo destacado es el estudio de Cengiz et al. (2022), que analiza el impacto del salario mínimo en el mercado laboral de Estados Unidos utilizando modelos de Aprendizaje Automático. Este enfoque presenta ventajas significativas frente a los modelos tradicionales, ya que se basa en un enfoque impulsado por los datos (data-driven), aprovechando directamente la información contenida en ellos, en lugar de depender de supuestos subjetivos.

En América Latina, también se encuentran ejemplos del uso de Aprendizaje Automático en el análisis del mercado laboral. Un estudio en Colombia destaca su implementación para pronosticar indicadores del mercado laboral en un contexto post-pandemia (Orozco-Castañeda et al., 2024). En Argentina, existen trabajos que aplican Aprendizaje Automático para estimar la probabilidad de que una mujer desempleada encuentre empleo (Staudt & Heredia, 2021) y para analizar transiciones en el mercado laboral (Staudt, 2022).

Entre los modelos específicos utilizados para pronosticar la salida laboral (turnover), Esmaieeli Sikaroudi et al. (2015) aplicaron diversos algoritmos a datos de una empresa fabricante de autopartes en Irán. Los resultados mostraron que Random Forest logró la mejor precisión (90.6 %), seguido por Naive Bayes (88.8 %) y redes neuronales tipo MLP (82 %). Por otro lado, Chakraborty et al. (2021) utilizaron datos de IBM, donde el método Gradient Booster mostró los mejores resultados (87.5 %), seguido por Random Forest (87.3 %).

### 3 Datos

#### 3.1 Fuentes de datos

Para la construcción del modelo de clasificación sobre la formalidad laboral, se utilizó como fuente principal la EPH, una base de datos representativa de la población argentina. La EPH abarca una amplia cobertura geográfica, centrándose en los principales aglomerados del país, y proporciona información detallada sobre diversos aspectos socioeconómicos de los individuos.

La variable objetivo de este estudio es predecir el estado formal de empleo a un año ( $t + 1$ ) en comparación con el año presente ( $t$ ) del registro del individuo. Para ello, se aprovechó la estructura de panel de la EPH, que permite hacer un seguimiento de los mismos individuos a lo largo del tiempo. En particular, se utilizaron los datos de las personas que permanecieron en la muestra de un año a otro, lo que permitió identificar los casos de aquellos individuos que al año siguiente se encontraron empleadas formalmente.

Para incorporar el contexto económico, se añadió al modelo la variable Producto Bruto Interno (PBI), que es una medida clave del estado de la economía.

Además, se incorporaron los índices de precios al consumidor (IPC) regionales para ajustar las variables monetarias a precios constantes, eliminando el efecto distorsionador de la inflación sobre los datos económicos.

### 3.2 Composición de la muestra

Se puede apreciar en la Figura 1 que en el periodo de análisis de 2016 a 2024, aproximadamente 260 mil individuos se mantuvieron en la muestra de un año a otro, lo que representa el 13.2 % de la muestra total de ese periodo (1.97 millones de encuestados). De estos, el 43.5 % (113 mil personas) transitó hacia el empleo formal en el año siguiente.



**Figura 1.** Muestra de la EPH utilizada. Período 2016-2024.

### 3.3 Procesamiento de los datos

La primera etapa consistió en consolidar las bases de datos de la EPH, publicadas trimestralmente, en una base de datos continua que abarca el periodo comprendido entre el segundo trimestre de 2016 y el segundo trimestre de 2024.

Para el tratamiento de los valores faltantes en variables numéricas, especialmente en aquellas relacionadas con los ingresos, se aplicó el método de imputación múltiple por ecuaciones encadenadas (MICE). Esta técnica genera múltiples imputaciones para los valores ausentes, reflejando la incertidumbre inherente al proceso logrando así imputaciones más consistentes (Azur et al., 2011).

Por su parte, para las variables categóricas con múltiples categorías, se utilizó la técnica de Pesos de la Evidencia o WoE (por sus siglas en inglés), cuya finalidad es transformar las variables categóricas en variables continuas según la probabilidad de ocurrencia del evento objetivo. Esta técnica no solo reduce la dimensionalidad del modelo y mejora su eficiencia, sino que también contribuye a evitar el sobreajuste al generalizar la información contenida en categorías poco frecuentes (Zdravevski et al., 2011).

## 4 Metodología

### 4.1 Enfoque metodológico CRISP-DM

Este trabajo sigue la metodología CRISP-DM (Cross Industry Standard Process for Data Mining), ampliamente utilizada en el ámbito de la minería de datos. Esta metodología comprende seis fases en un ciclo iterativo:

1. **Comprensión del Negocio:** Entender a fondo el problema del mercado laboral formal, sus factores clave y variables influyentes.
2. **Comprensión de los Datos:** Examinar los datos disponibles, identificar fuentes, explorar su calidad y determinar las variables relevantes.
3. **Preparación de los Datos:** Transformar y preparar los datos para el análisis, incluyendo limpieza, normalización y creación de nuevas características.
4. **Modelado:** Crear modelos predictivos utilizando algoritmos de Aprendizaje Automático.
5. **Evaluación:** Validar el desempeño de los modelos mediante métricas como precisión, recall, F1-score y AUC.
6. **Despliegue:** Implementar el modelo para realizar predicciones con datos reales.

#### 4.2 Variables e indicadores

Las variables utilizadas en la construcción de los modelos se seleccionaron con base en la literatura consultada y múltiples pruebas. Las mismas se observan en el Cuadro 1 y estas incluyen características personales (sexo, edad, estado civil), del hogar (relación dentro del hogar), educativas (nivel educativo), temporales (año, trimestre), territoriales (región), de actividad (estado de actividad, ingresos), laborales (antigüedad, tamaño de empresa, sector, ingresos laborales, calificación, intensidad) y de contexto económico (variación interanual del PBI).

La variable objetivo (estado formal) se construye indirectamente a partir de otras variables como la categoría ocupacional (si es trabajador dependiente) y si se les descuenta de su salario un aporte para su jubilación.

#### 4.3 Modelos implementados

Se seleccionaron siete modelos para su implementación y comparación:

1. **Regresión logística (Logit):** Modelo tradicional que estima la probabilidad de un evento binario mediante la función logística.
2. **Naive Bayes:** Algoritmo probabilístico basado en el teorema de Bayes que asume independencia entre las variables predictoras.
3. **Árbol de decisión:** Estructura jerárquica que segmenta el espacio de características mediante decisiones binarias.
4. **Random Forest:** Ensamble de múltiples árboles de decisión que mejora la precisión y reduce el sobreajuste.
5. **XGBoost:** Algoritmo de boosting que construye árboles secuencialmente, donde cada árbol corrige los errores del anterior.
6. **Support Vector Machine (SVM):** Algoritmo que busca el hiperplano óptimo que separa las clases maximizando el margen, haciendo uso de diversos kernels (lineal, polinomial y de base radial (RBF)).
7. **Redes neuronales:** Modelo inspirado en el cerebro humano que procesa la información a través de capas de neuronas interconectadas.

**Cuadro 1.** Variables e indicadores.

Unidad de análisis	Variable	Indicador
Individuo encuestado en la EPH en el periodo 2016-2024	Persona	Sexo
		Edad
		Estado civil
	Hogar	Relación dentro del hogar (Jefe/a del hogar, cónyuge, hijo, etc.)
	Educación	Nivel educativo máximo alcanzado
	Tiempo	Año
		Trimestre
	Territorio	Región INDEC
	Actividad	Estado de actividad (Empleado, desempleado, inactivo, etc.)
		Ingresos totales
	Trabajo	Antigüedad laboral
		Tamaño empresa
		Sector laboral
		Ingresos laborales
		Calificación laboral
		Intensidad laboral (subocupados, normal o sobreocupado)
		Sector institucional (público o privado)
		Estado formal
	Contexto económico	Variación interanual del PBI

#### 4.4 Optimización de los hiperparámetros

Para los modelos más complejos (XGBoost, SVM, redes neuronales), se utilizó la optimización bayesiana para encontrar los valores óptimos de los hiperparámetros. Esta técnica construye un modelo probabilístico que representa la creencia sobre el rendimiento del algoritmo en función de los hiperparámetros, utilizando información de evaluaciones previas para hacer más eficiente la búsqueda.

Como función objetivo para la optimización se utilizó el Área Bajo la Curva ROC (AUC), que mide la capacidad discriminativa del modelo independientemente del umbral de clasificación elegido.

#### 4.5 Métricas de evaluación

Con el objetivo de evaluar de forma rigurosa el rendimiento de los modelos, se dividió la base de datos en dos subconjuntos: entrenamiento (70 %) y prueba (30 %). El primero se utilizó para ajustar los modelos, mientras que el segundo permitió evaluar su capacidad de generalización sobre datos no observados previamente.

Para comparar el desempeño de los modelos sobre el conjunto de prueba, se recurrió a diversas métricas de evaluación:

1. **Accuracy (exactitud):** Proporción de predicciones correctas sobre el total de observaciones.
2. **AUC:** Área bajo la curva ROC, que refleja la capacidad del modelo para distinguir entre clases.
3. **Recall (sensibilidad):** Proporción de casos positivos correctamente identificados.
4. **Specificity (especificidad):** Proporción de casos negativos correctamente identificados.
5. **Precision (precisión):** Proporción de predicciones positivas que resultaron correctas.
6. **F1-score:** Media armónica entre precisión y sensibilidad.

### 5 Resultados

Una vez implementados y optimizados los siete modelos propuestos, se procedió a evaluar su desempeño utilizando las métricas definidas previamente. El Cuadro 2 presenta un resumen comparativo de los resultados obtenidos.

El modelo XGBoost se destaca como el más robusto, alcanzando la mayor exactitud (84.8 %) y el valor más alto de AUC (91.6 %), lo cual evidencia una excelente capacidad para discriminar entre clases. Además, exhibe un equilibrio notable entre sensibilidad y especificidad (ambas del 84.8 %), lo que refleja una buena capacidad para identificar tanto los casos positivos como los negativos.

El modelo Random Forest también muestra un desempeño destacado, con una exactitud del 84.5 % y una AUC de 91.2 %, junto con métricas sólidas como

**Cuadro 2.** Resumen de métricas de evaluación para los modelos ajustados.

Modelo	Accuracy	AUC	F1	Sensibilidad	Especificidad	Precisión
Logit	80.9 %	88.0 %	79.0 %	82.6 %	79.6 %	75.7 %
Naive Bayes	79.6 %	86.3 %	76.8 %	77.3 %	81.5 %	76.3 %
Árbol de decisión	82.6 %	87.5 %	84.8 %	86.1 %	78.0 %	83.6 %
<b>Random Forest</b>	84.5 %	91.2 %	<b>86.4 %</b>	<b>86.9 %</b>	81.4 %	<b>85.9 %</b>
<b>XGBoost</b>	<b>84.8 %</b>	<b>91.6 %</b>	83.0 %	84.8 %	<b>84.8 %</b>	81.2 %
Redes neuronales	82.3 %	89.8 %	80.4 %	83.3 %	81.4 %	77.6 %
SVM Lineal	81.1 %	88.1 %	78.6 %	79.7 %	82.1 %	77.4 %
SVM Polinomial	81.5 %	87.8 %	78.0 %	75.3 %	86.2 %	80.8 %
SVM RBF	82.9 %	89.2 %	80.0 %	78.4 %	86.4 %	81.6 %

el F1 Score (86.4 %) y la precisión (85.9 %). Aunque altamente competitivo, presenta un rendimiento levemente inferior al de XGBoost en términos de balance general.

El modelo Logit, utilizado como referencia base, obtuvo una exactitud del 80.9 % y una AUC de 88.0 %. Si bien su desempeño es más modesto, constituye un punto de partida útil para comparar mejoras logradas mediante modelos más complejos. Modelos como Árbol de Decisión, SVM con kernel RBF y Redes Neuronales evidencian mejoras notables respecto al Logit, en particular por su capacidad de capturar relaciones no lineales en los datos.

En el Cuadro 3 podemos apreciar la importancia por variable para el modelo XGBoost medida a través de tres variables Gain (o ganancia), Cover (cobertura) y Frequency (frecuencia). Gain mide la mejora en la precisión del modelo al dividir un nodo con la variable en cuestión, es decir, cuánto contribuye cada variable a reducir el error y mejorar la predicción. Cover calcula el número de observaciones que se ven afectadas por la variable durante las divisiones del árbol, lo que indica el alcance de la variable en términos de la cantidad de datos que influye. Por último, Frequency mide la frecuencia de aparición de la variable a lo largo de todos los árboles, es decir, cuán a menudo se utiliza esta variable para hacer divisiones dentro del conjunto de árboles del modelo.

El Cuadro 3 revela que el tamaño de la empresa constituye la variable más decisiva en el modelo, con un valor Gain de 0.35, seguida por el ingreso real transformado en logaritmos (0.13). En tercer y cuarto lugar, con un impacto significativo aunque de menor magnitud, se encuentran el sector de actividad (0.11) y la pertenencia al sector público (0.10). Es particularmente interesante observar que, a pesar de que la cobertura de datos (Cover: 0.12) y la frecuencia de aparición (Frequency: 0.10) de la variable tamaño de empresa son relativamente bajas en comparación con el ingreso real (Cover: 0.17; Frequency: 0.20), esta mantiene su posición predominante. Esto refuerza la importancia intrínseca de dicha variable, ya que demuestra que, incluso con menor presencia en los distintos árboles del modelo, su capacidad para aportar información relevante a la predicción es considerablemente superior.



**Cuadro 3.** Importancia de las variables en el modelo final de XGBoost.

<b>Variable</b>	<b>Gain</b>	<b>Cover</b>	<b>Frequency</b>
Tamaño empresa	0.35	0.12	0.10
Log Ingreso Real	0.13	0.17	0.20
Sector de actividad	0.11	0.10	0.10
Sector público	0.10	0.02	0.02
Log Horas de trabajo	0.06	0.10	0.10
Edad	0.06	0.12	0.12
Antigüedad laboral	0.05	0.04	0.04
Nivel educativo	0.03	0.04	0.04
Intensidad laboral	0.03	0.03	0.03
Calificación laboral	0.02	0.04	0.04
Region	0.02	0.04	0.04
Var. i.a. PBI	0.01	0.04	0.05
Parentezco hogar	0.01	0.02	0.02
Tamaño empresa 2	0.01	0.02	0.01
Año	0.01	0.03	0.03
Estado civil	0.01	0.02	0.03
Desempleado	0.00	0.01	0.01
Sexo	0.00	0.01	0.01
Trimestre	0.00	0.02	0.02

En síntesis, el análisis comparativo de los modelos permite concluir que XGBoost ofrece el mejor equilibrio entre precisión y capacidad discriminativa para predecir la formalidad laboral, superando significativamente al modelo logístico tradicional. Asimismo, las variables relacionadas con las características de la empresa y los ingresos emergen como los predictores más importantes.

## 6 Conclusiones

Este trabajo desarrolló un enfoque basado en Ciencia de Datos para pronosticar el nivel de empleo formal en Argentina en tiempo real, utilizando datos de la Encuesta Permanente de Hogares. A través de la implementación de diversos modelos de Aprendizaje Automático y su comparación con el modelo logístico tradicional, se evidenció que los modelos avanzados tienen una capacidad predictiva notablemente superior.

De los modelos evaluados, el XGBoost se destacó como el mejor predictor, logrando un accuracy de 84.8 % y un AUC de 91.6 %. Estas métricas indican una alta capacidad discriminativa, lo que lo convierte en el modelo más preciso para predecir el empleo formal. El método de Random Forest también mostró un desempeño destacado, con un accuracy de 84.5 % y un AUC de 91.2 %, consolidándose como una alternativa robusta.

El análisis de la importancia de las variables reveló que el tamaño de la empresa es el factor más determinante para predecir si un individuo estará empleado formalmente en el año siguiente. Otras variables relevantes incluyen el ingreso real, el sector de actividad, la pertenencia al sector público, la edad y la antigüedad laboral.

Los resultados subrayan la capacidad superior de los algoritmos de Aprendizaje Automático para abordar problemas complejos de pronóstico, particularmente en el contexto del empleo formal. Además, resaltan la importancia de integrar herramientas de Ciencia de Datos en la toma de decisiones públicas y empresariales, ya que pueden mejorar significativamente la precisión de las proyecciones y facilitar la implementación de políticas más oportunas y efectivas.

La adopción de estos modelos podría reducir el retraso en la disponibilidad de datos sobre empleo formal, mejorar las proyecciones y ayudar en la formulación de políticas públicas más informadas, especialmente aquellas orientadas a la promoción de la formalización laboral.

Como líneas futuras de investigación, se sugiere ampliar la validación de los modelos incorporando datos provenientes de fuentes adicionales, como las bases del Sistema Integrado Previsional Argentino (SIPA). Asimismo, sería valioso explorar el efecto de nuevas variables explicativas y evaluar la aplicabilidad de estos modelos en contextos internacionales, con el objetivo de desarrollar herramientas predictivas que puedan generalizarse a economías con características similares.

## Agradecimientos

Quiero expresar mi más sincero agradecimiento al ITBA por su apoyo y por proporcionarme los recursos necesarios para el desarrollo de este artículo, que se basa en mi tesis de maestría. La oportunidad de colaborar con una institución tan prestigiosa ha sido invaluable en el proceso de investigación.

Mi más profundo agradecimiento a mi directora, María Juliana Gambini, por su constante orientación, paciencia y valiosos comentarios a lo largo de mi tesis. Su guía experta y su dedicación fueron fundamentales para el desarrollo de este trabajo.

## Referencias

- Abbasi, S. M., & Hollman, K. W. (2000). Turnover: The real bottom line. *Public Personnel Management*, 29(3), 333-342. <https://doi.org/10.1177/009102600002900303>
- Azur, M. J., Stuart, E. A., Frangakis, C., & Leaf, P. J. (2011). Multiple imputation by chained equations: what is it and how does it work? *International Journal of Methods in Psychiatric Research*, 20(1), 40-49. <https://doi.org/10.1002/mpr.329>
- Cengiz, D., Dube, A., Lindner, A. S., & Zentler-Munro, D. (2022). Seeing beyond the trees: Using machine learning to estimate the impact of minimum wages on labor market outcomes. *Journal of Labor Economics*, 40(S1), S203-S247. <https://doi.org/10.3386/w28399>
- Chakraborty, R., Mridha, K., Shaw, R. N., & Ghosh, A. (2021). Study and prediction analysis of the employee turnover using machine learning approaches. *IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON)*. <https://doi.org/10.1109/GUCON50781.2021.9573759>
- Esmiaeeli Sikaroudi, A. M., Ghousi, R., & Sikaroudi, A. (2015). A data mining approach to employee turnover prediction (case study: Arak automotive parts manufacturing). *Journal of industrial and systems engineering*, 8(4), 106-121.
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260. <https://doi.org/10.1126/science.aaa8415>
- Orozco-Castañeda, J. M., Sierra-Suárez, L. P., & Vidal, P. (2024). Labor market forecasting in unprecedented times: A machine learning approach. *Bulletin of Economic Research*, 76(4), 893-915. <https://doi.org/10.1111/boer.12451>
- Rofman, R. (2007). La informalidad laboral y su rol en la determinación de la condiciones de vida de los trabajadores. Algunas observaciones en base a datos del Gran Buenos Aires. *IX Jornadas Argentinas de Estudios de Población*.

- Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3), 210-229. <https://doi.org/10.1147/rd.33.0210>
- Staudt, A. (2022). *Predicción de transiciones laborales de la actividad a la inactividad en el mercado laboral argentino: un enfoque de aprendizaje automático* [Tesis doctoral]. Universidad Nacional de La Plata. <https://doi.org/10.35537/10915/139421>
- Staudt, A., & Heredia, J. L. (2021). Predicción de la probabilidad de inserción ocupacional de los desocupados en Argentina (2003-2019). *Visión de futuro*, 25(2), 1-21.
- Zdravevski, E., Lameski, P., & Kulakov, A. (2011). Weight of evidence as a tool for attribute transformation in the preprocessing stage of supervised learning algorithms. *The 2011 International Joint Conference on Neural Networks (IJCNN)*, 181-188. <https://doi.org/10.1109/IJCNN.2011.6033219>
- Zhang, Y. (2016). A review of employee turnover influence factor and counter-measure. *Journal of Human Resource and Sustainability Studies*, 4(2), 85-91. <https://doi.org/10.4236/jhrss.2016.42010>