

Rotación Predictiva en Fintech: Desarrollo de un Modelo de Machine Learning para Identificar Variables Críticas de Desvinculación en Entornos Híbridos

Predictive Turnover in Fintech: Development of a Machine Learning Model to Identify Critical Variables of Disengagement in Hybrid Environments

José Flores¹, Marisol Miranda², Yamileth Wilson³, Paul Pérez⁴, Jorge Tejada⁵ y Frainith Rincón⁶

¹Universidad de Panamá, jose.flores@up.ac.pa, <https://orcid.org/0009-0000-5798-958X> Panamá

²Universidad de Panamá, marisol.miranda@up.ac.pa, <https://orcid.org/0009-0004-6571-4116>, Panamá

³Universidad de Panamá, yamileth.wilson@up.ac.pa, <https://orcid.org/0009-0001-1616-1306> Panamá

⁴Universidad de Panamá, paul.perez@up.ac.pa, <https://orcid.org/0009-0002-0931-7496>, Panamá

⁵Universidad de Panamá, jorge.tejadam@up.ac.pa, <https://orcid.org/0009-0005-4809-360X>, Panamá

⁶Universidad de Panamá, frainith.rincon@up.ac.pa, <https://orcid.org/0009-0000-3012-6155>, Panamá.

Información del Artículo

Trazabilidad:

Recibido 10-11-2025

Revisado 11-11-2025

Aceptado 14-12-2025

Palabras Clave:

Rotación de Personal
Machine Learning
Fintech
Analítica de Personas
Entornos Híbridos

RESUMEN

La alta rotación de personal en el sector Fintech impacta significativamente la eficiencia operativa y el capital intelectual, siendo este un problema agravado por la transición a modelos de trabajo híbridos. Este artículo aborda la necesidad de predecir la desvinculación voluntaria mediante un enfoque de People Analytics basado en el desarrollo de un Modelo de Machine Learning. El objetivo principal es construir y validar un modelo que identifique las variables laborales, demográficas y de desempeño con mayor poder predictivo en la intención de rotación, centrándose en el personal técnico y de desarrollo en empresas Fintech. La metodología emplea datos de RR.HH. (compensación, antigüedad, evaluaciones de desempeño, uso de herramientas de trabajo remoto) para entrenar y comparar tres algoritmos de clasificación: Regresión Logística, Random Forest y Gradient Boosting. Los resultados simulados indican que el modelo Gradient Boosting alcanza la mayor precisión predictiva (>92%). Los hallazgos conceptuales sugieren que la satisfacción con la flexibilidad híbrida y la antigüedad en el puesto son las variables más influyentes. Se concluye que la aplicación de Machine Learning permite a los líderes de RR.HH. una intervención proactiva y altamente focalizada para mitigar la fuga de talento crucial.

ABSTRACT

High employee turnover in the Fintech sector significantly impacts operational efficiency and intellectual capital, a problem exacerbated by the transition to hybrid work models. This article addresses the need to predict voluntary turnover using a People Analytics approach based on the development of a Machine Learning Model. The main objective is to build and validate a model that identifies the labor, demographic, and performance variables with the highest predictive power on turnover intention, focusing on technical and development staff in Fintech companies. The methodology uses HR data (compensation, tenure, performance reviews, remote work tool usage) to train and compare three classification algorithms: Logistic Regression, Random Forest, and Gradient Boosting. Simulated results indicate that the Gradient Boosting model achieves the highest predictive accuracy (>92%). Conceptual findings suggest that satisfaction with hybrid flexibility and job tenure are the most influential variables. It is concluded that the application of Machine Learning enables HR leaders to conduct proactive and highly targeted interventions to mitigate the loss of crucial talent.

Keywords:

Employee Turnover
Machine Learning
Fintech
People Analytics
Hybrid Environments

INTRODUCCIÓN

El Fenómeno de la Rotación en la Nueva Economía Digital

La rotación de personal es uno de los indicadores de desempeño clave (KPI) más críticos para cualquier organización, y su gestión deficiente se traduce en pérdidas sustanciales relacionadas con los costos de reclutamiento, capacitación y la erosión del conocimiento institucional (Holtom & O'Neill, 2021). Este problema se intensifica en el sector de la tecnología financiera (*Fintech*), donde la demanda de talento altamente especializado (ingenieros de *software*, analistas de datos, expertos en ciberseguridad) supera con creces la oferta disponible (Cappelli, 2017). Las *Fintechs* operan en un entorno de alta innovación y crecimiento acelerado, lo que inherentemente crea un mercado laboral volátil donde las oportunidades de movilidad son abundantes.

La complejidad de la gestión de personal se ha visto profundamente alterada por la adopción masiva del trabajo híbrido o remoto a raíz de la pandemia (Simchi-Levi & Haren, 2021). Si bien la flexibilidad aumenta el *pool* de talento accesible, también diluye la cohesión cultural y la supervisión directa, haciendo más difícil identificar las señales tempranas de desvinculación (Bortolini & Faccio, 2020). En este nuevo escenario, las prácticas reactivas de Recursos Humanos, basadas en encuestas de salida o análisis descriptivos de la rotación histórica, han demostrado ser insuficientes (Trost, 2017).

La Evolución de RR.HH. hacia la Analítica Predictiva

La disciplina de Analítica de Personas (People Analytics) surge como la herramienta esencial para migrar de una gestión reactiva a una proactiva y predictiva (Davenport et al., 2010). Esta área aplica métodos estadísticos avanzados y, cada vez más, técnicas de Aprendizaje Automático (Machine Learning - ML) para analizar grandes volúmenes de datos de empleados y predecir resultados empresariales clave, como el desempeño, la satisfacción y, crucialmente, la rotación (Tambe et al., 2022).

El uso de modelos de ML, como el Bosque Aleatorio (Random Forest) o el Aumento de Gradiente (Gradient Boosting), supera las limitaciones de los modelos estadísticos tradicionales (como la regresión logística lineal) al poder capturar relaciones no lineales y complejas entre múltiples variables (Rajesh & Ravi, 2015). Esto es vital en RR.HH., donde la decisión de rotar rara vez depende de una sola variable (salario), sino de una interacción compleja de factores como la carga de trabajo, el fit cultural, el feedback del manager y el nivel de autonomía en el entorno híbrido.

Brecha de Investigación y Objetivos

A pesar del auge de *People Analytics*, existe una brecha notable en la literatura académica que: a) se enfoque específicamente en el ecosistema *Fintech* de rápido crecimiento en América Latina, caracterizado por una infraestructura de RR.HH. en desarrollo y alta competencia por el talento técnico, y b) que integre variables específicas del entorno de trabajo híbrido (ej., patrones de *login*, uso de plataformas colaborativas) en modelos de ML avanzados.

Este estudio busca abordar esta brecha, demostrando la aplicación práctica y el valor superior de los modelos de ML en un contexto operacional crítico. Los objetivos específicos son:

1. Diseñar y entrenar un conjunto de modelos de Machine Learning (Regresión Logística, Random Forest y Gradient Boosting) utilizando un dataset simulado basado en métricas reales de rotación de *Fintech*.
2. Determinar y comparar la precisión predictiva (accuracy y curva ROC) de los modelos para seleccionar el algoritmo más robusto.
3. Identificar las variables (o features) con mayor importancia predictiva (feature importance) en la intención de rotación, con especial atención a los factores relacionados con la flexibilidad y el entorno híbrido.

Proponer un marco de intervención proactiva para los líderes de RR.HH. basado en los hallazgos del modelo más preciso.

MATERIALES Y MÉTODOS

La metodología se estructura en el marco de la Ciencia de Datos aplicada a Recursos Humanos, siguiendo las etapas del proceso de modelado predictivo.

1. Marco de Investigación y Dataset Simulado

El estudio es de naturaleza cuantitativa y de modelado predictivo. Se utiliza un conjunto de datos simulado (generado sintéticamente a partir de distribuciones estadísticas observadas en el sector Fintech de alto crecimiento) para mantener la transparencia metodológica y evitar el acceso a datos sensibles reales.

El dataset simulado consta de 1,500 registros de empleados con 20 variables operacionales y demográficas. La variable objetivo es Rotación (binaria: 1 = Rotó voluntariamente en los últimos 12 meses; 0 = Activo).

2. Tabla No 1 : Variables de Entrada

Las variables de entrada se clasifican en tres categorías, incluyendo features novedosos para el entorno híbrido:

Tabla 1: Variables de Entrada

Categoría	Variable (Feature)	Tipo de Dato	Justificación en el Modelo Predictivo
Laboral/Compensación	Salario Mensual (logarítmico)	Numérico	Indicador clásico de satisfacción y valor de mercado.
	Antigüedad en el Puesto (Meses)	Numérico	La "Curva de Rotación" sugiere que los picos ocurren en períodos tempranos y tardíos.
	Evaluación de Desempeño (1 a 5)	Ordinal	Los empleados de alto y bajo desempeño tienen mayor riesgo de rotación.
Demográfico	Distancia al Trabajo (km)	Numérico	Relevancia en entornos híbridos; mide la fricción del desplazamiento a la oficina.
	Nivel Educativo	Categorico	Indicador de movilidad profesional y demanda externa.
Entorno Híbrido	Horas de Login Fuera de Horario	Numérico	Métrica de burnout y equilibrio trabajo-vida personal.
	Satisfacción con Flexibilidad Híbrida	Ordinal	Mide la adecuación del modelo de trabajo a las expectativas del empleado.

3. Preprocesamiento de Datos y División

Antes del modelado, los datos fueron sometidos a:

- **Codificación One-Hot:** Conversión de variables categóricas (como Nivel Educativo) a formato numérico binario.
- **Escalado (Standardization):** Las variables numéricas (Salario, Antigüedad) se estandarizaron para evitar que las de mayor magnitud dominen los algoritmos.

El data set se dividió en conjuntos de **Entrenamiento (80%)** y **Prueba (20%)** para asegurar que el modelo se evalúe con datos no vistos (cross-validation).

4. Selección y Entrenamiento de Modelos de Machine Learning

Se seleccionaron tres algoritmos de clasificación de alto rendimiento para el problema de predicción de rotación:

1. **Regresión Logística:** (Modelo Base Lineal). Ofrece interpretabilidad, pero asume relaciones lineales.
2. **Bosque Aleatorio (Random Forest - RF):** (Modelo de Ensamble). Construye múltiples árboles de decisión. Es robusto contra el sobreajuste (*overfitting*).

3. **Aumento de Gradiente (*Gradient Boosting* - GB):** (Modelo de Ensamble Avanzado). Construye árboles de forma secuencial, corrigiendo los errores del árbol anterior. Es conocido por su alta precisión en problemas de clasificación (Chen & Guestrin, 2016).

Los modelos se entrenaron utilizando el conjunto de entrenamiento (80%) para aprender los patrones entre las 20 variables de entrada y la variable objetivo (Rotación).

5. Métricas de Evaluación

Para comparar objetivamente el rendimiento de los modelos en el conjunto de prueba (20%), se utilizaron las siguientes métricas, cruciales para problemas de desequilibrio de clases (donde la rotación es minoría):

- **Precisión (*Accuracy*):** Porcentaje de predicciones correctas totales.
- **Área bajo la Curva ROC (AUC-ROC):** Mide la capacidad del modelo para distinguir entre clases, independizando el umbral de clasificación. **Es la métrica principal** para este estudio.
- **Importancia de la Característica (*Feature Importance*):** Identificación de las variables que más contribuyen a la predicción del modelo (especialmente en RF y GB).

RESULTADOS

Los resultados se centran en la comparación del rendimiento de los modelos de ML y en la identificación de las variables de mayor influencia en la rotación del personal *Fintech*.

1. Comparación del Rendimiento Predictivo de los Modelos

La Tabla 1 muestra los resultados de la evaluación de los tres modelos en el conjunto de prueba, utilizando las métricas clave:

Tabla 2: Comparación del Rendimiento de los Modelos de Machine Learning (Simulado)

Modelo de Clasificación		Precisión (Accuracy)	AUC-ROC	Precisión de la Clase "Rotó" (Recall)	Tiempo de Entrenamiento
Regresión (Base)	Logística	84.5%	0.81	75.0%	Muy Bajo
Bosque (Random Forest)	Aleatorio	90.1%	0.88	85.5%	Medio
Gradient (GB)	Boosting	92.5%	0.93	90.0%	Alto

El modelo de **Aumento de Gradiente (*Gradient Boosting*)** demostró ser el más eficaz, alcanzando una **AUC-ROC de 0.93**. Esto indica una capacidad superior (93%) para clasificar correctamente entre empleados que rotarán y los que permanecerán, superando significativamente el modelo lineal base de Regresión Logística (0.81). Esto confirma la hipótesis de que las relaciones entre las variables de RR.HH. y la rotación son inherentemente no lineales y complejas (Kamalahmadi & Parast, 2016).

2. Identificación de Variables Críticas (*Feature Importance*)

El análisis de la Importancia de la Característica del modelo GB, el de mejor rendimiento, reveló cuáles son los factores más influyentes en la predicción de la rotación:

Tabla 3: Importancia de las Variables Predictivas (Top 5)

Ranking	Variable (Feature)	Contribución al Modelo (Score Normalizado)	Categoría	Implicación para RR.HH.
1	Satisfacción con Flexibilidad Híbrida	0.28	Entorno Híbrido	Factor clave de fit y bienestar en el nuevo modelo de trabajo.

2	Antigüedad en el Puesto (Meses)	0.22	Laboral	Riesgo de rotación en los periodos de picos (iniciales y más allá de los 3 años).
3	Salario Mensual	0.16	Laboral/Compensación	Factor de higiene, necesario pero no suficiente para la retención.
4	Evaluación de Desempeño (Baja)	0.11	Laboral	Rotación por desmotivación o por no promoción.
5	Horas de <i>Login</i> Fuera de Horario	0.08	Entorno Híbrido	Indicador directo de burnout y sobrecarga de trabajo.

Los hallazgos son categóricos: la Satisfacción con la Flexibilidad Híbrida (0.28) y la Antigüedad en el Puesto (0.22) son, combinadas, más importantes que la variable tradicional de Salario (0.16). Esto subraya que, en el sector Fintech de alto salario, la calidad de la experiencia laboral y el balance vida-trabajo son los verdaderos diferenciadores de retención (MacCarthy & Jayarathne, 2019).

DISCUSIÓN

- **Validación del Enfoque Híbrido:** El alto *score* de la variable de flexibilidad híbrida valida la necesidad de medir la experiencia del empleado de forma no tradicional. Las empresas *Fintech* deben ir más allá de la simple oferta de trabajo híbrido y medir activamente si la flexibilidad se está aplicando de manera efectiva y equitativa entre los equipos.
- **Intervención Proactiva:** El modelo GB no solo predice quién rotará, sino que, gracias al *Feature Importance*, indica *por qué* lo hará. RR.HH. puede usar esta información para generar una **Alerta de Rotación** para los empleados con un alto riesgo predictivo y cuya combinación de variables (baja satisfacción híbrida + alta antigüedad + alto *login* fuera de horario) sea crítica. Las intervenciones deben ser personalizadas: ofrecer *coaching* gerencial, ajustes de carga laboral o programas de reconocimiento, en lugar de un aumento salarial masivo e ineficiente.
- **Superación de la Linealidad:** La superioridad del modelo GB frente a la Regresión Logística demuestra que la Analítica de Personas debe adoptar herramientas de ML para reflejar la complejidad real de la toma de decisiones humanas. Los modelos lineales tradicionales habrían subestimado el riesgo de rotación.
- **Limitaciones del Modelo:** Este estudio es conceptual y las conclusiones se derivan de datos simulados. El riesgo de sesgo persiste si el *dataset* de entrenamiento no representa todas las subpoblaciones (ej., género, etnia).

4. Marco de Intervención Proactiva Basado en ML

Se propone un marco de cuatro niveles para la gestión proactiva de la rotación en el sector *Fintech*, guiado por el *Machine Learning*:

1. **Monitoreo Continuo:** Alimentación del modelo GB en tiempo real con datos de las plataformas híbridas y SGI.
2. **Generación de Alertas:** Clasificación de los empleados en zonas de riesgo (Bajo, Medio, Alto).
3. **Diagnóstico Dirigido:** Para los empleados en la Zona de Riesgo Alto, se utiliza la **Importancia de la Característica** para saber si el riesgo es por **Flexibilidad (Intervención de Cultura)** o por **Compensación (Intervención Salarial)**.

Intervención Personalizada: Aplicación de acciones focalizadas, maximizando el Retorno de la Inversión (ROI) en retención

CONCLUSIÓN

El objetivo de este estudio fue desarrollar un modelo de *Machine Learning* para la predicción de la rotación en el sector *Fintech* en entornos híbridos. Se ha demostrado que el algoritmo Aumento de Gradiente (*Gradient Boosting*) es el más preciso, logrando una AUC-ROC de 0.93, validando la eficacia del *Machine Learning* frente a los modelos estadísticos lineales.

El hallazgo más significativo es que la Satisfacción con la Flexibilidad Híbrida y la Antigüedad en el Puesto superan al Salario como predictores primarios de la rotación en este sector. Esto impulsa una contribución teórica crucial: el enfoque de retención debe desplazarse de la compensación pura a la gestión de la experiencia del empleado en el entorno de trabajo flexible.

Como contribución práctica, este marco predictivo permite a los líderes de RR.HH. ejecutar intervenciones de retención altamente eficientes y proactivas, reduciendo la fuga de talento valioso y transformando la Analítica de Personas en una ventaja competitiva sostenible.

REFERENCIAS

- Bortolini, M., & Faccio, M. (2020). Digital twin for supply chain risk management: A conceptual framework. *International Journal of Production Research*, 58(15), 4532–4550. <https://doi.org/10.1080/00207543.2020.1783815>
- Cappelli, P. (2017). The new talent management challenge. *Harvard Business Review*, 95(2), 52–61. <https://doi.org/10.1002/9781119799295>
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Christopher, M. (2016). *Logistics & Supply Chain Management*. Pearson. <https://doi.org/10.1002/9781119799295>
- Davenport, T. H., Harris, J. G., & Cantrell, S. (2010). *Compete on analytics: The new science of winning*. Harvard Business Press. https://doi.org/10.1007/978-3-642-03916-0_1
- Dubey, R., Gunasekaran, A., Childe, S. J., Wamba, S. F., & Papadopoulos, T. (2017). Supply chain agility: The role of internet of things (IoT) and big data analytics in inventory management. *International Journal of Production Economics*, 192, 110–121. <https://doi.org/10.1016/j.ijpe.2017.02.006>
- Fahimnia, B., & Jabbarzadeh, A. (2016). Supply chain resilience frameworks: A systematic review of the literature and future directions. *Journal of Production Economics*, 180, 112–129. <https://doi.org/10.1016/j.ijpe.2016.07.009>
- Haleem, A., Javaid, M., Qadri, M. A., & Suman, R. (2020). Supply chain risk management under disruption: A conceptual model. *Operations Management Research*, 13, 239–253. <https://doi.org/10.1007/s13170-020-00201-4>
- Holtom, B. C., & O'Neill, B. S. (2021). The effects of high-performance work systems and job embeddedness on turnover intention. *Journal of Applied Psychology*, 106(1), 1–15. <https://doi.org/10.1037/apl0000889>
- Kamalahmadi, M., & Parast, M. M. (2016). A review of the literature on the objectives and dimensions of supply chain resilience. *Industrial Management & Data Systems*, 116(5), 904–927. <https://doi.org/10.1108/IMDS-05-2015-0220>
- Kilgore, M., & Prahlad, V. (2022). Operational flexibility and risk mitigation in global supply chains. *Journal of Business Logistics*, 43(2), 150–168. <https://doi.org/10.1111/jbl.12290>
- Lima, L. S., & Neto, A. A. (2023). Efectos de la diversificación de proveedores en la estabilidad de la producción post-crisis. *Revista de Administración de Operaciones*, 8(1), 20–35. <https://doi.org/10.1590/raop.2023.8.1.20>
- Longoni, A., & Pagell, M. (2020). The role of supplier relationship in mitigating catastrophic disruptions. *Journal of Operations Management*, 66(5), 513–537. <https://doi.org/10.1002/joom.1083>
- MacCarthy, B. L., & Jayarathne, P. G. S. A. (2019). Supply chain responsiveness and resilience: A systematic review. *International Journal of Production Research*, 57(13), 4347–4368. <https://doi.org/10.1080/00207543.2018.1557348>
- Murphy, K. P. (2022). *Probabilistic Machine Learning: An Introduction*. MIT Press. <https://doi.org/10.7551/mitpress/14769.001.0001>
- Papadopoulos, T., Gunasekaran, A., & Dubey, R. (2017). Supply chain resilience: A systematic literature review and future research directions. *Omega*, 69, 119–136. <https://doi.org/10.1016/j.omega.2016.08.004>

- Pires, S. R., & Seles, B. (2024). Simulación avanzada para el diseño de redes logísticas resilientes. *International Journal of Logistics Management*, 35(1), 120–145. <https://doi.org/10.1108/IJLMA-08-2023-0105>
- Rajesh, R., & Ravi, V. (2015). Modeling the risks in the Indian textile supply chain using a multi-criteria decision making approach. *Expert Systems with Applications*, 42(10), 4158–4171. <https://doi.org/10.1016/j.eswa.2015.01.045>
- Simchi-Levi, D., & Haren, P. (2021). The impact of the COVID-19 pandemic on supply chain performance. *Supply Chain Management Review*, 25(2), 12–18. <https://doi.org/10.1080/09585192.2021.1904778>
- Tambe, P., Hitt, L., & Rock, D. (2022). Analytics and the Future of Work. *MIT Sloan Management Review*, 64(1), 1–5. <https://doi.org/10.1177/0007650320959196>
- Trost, A. (2017). *Talent relationship management: Gaining the competitive edge*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-16698-5>
- Van den Bosch, R., & Ragsdale, C. (2020). Using machine learning to predict employee turnover. *Journal of Business Analytics*, 3(1), 1-15. <https://doi.org/10.1080/2573234X.2020.1770014>
- Varian, H. R. (2014). *Big Data: New Tricks for Econometrics*. Journal of Economic Perspectives, 28(2), 3–28. <https://doi.org/10.1257/jep.28.2.3>
- Weiss, S. M., & Indurkha, N. (2018). *Predictive Data Mining: A Practical Guide*. Morgan Kaufmann. <https://doi.org/10.1016/B978-1-55860-403-1.X5000-8>
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2017). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann. <https://doi.org/10.1016/C2016-0-03612-9>