

Proyecto: Prevención de Fraude en Retiros por Desempleo utilizando Inteligencia Artificial en el IMSS y PENSIONISSSTE

Javier Horacio Pérez Ricárdez

Octubre del 2024

1 Introducción

El fraude en el sistema de retiro por desempleo es un problema que afecta tanto al **Instituto Mexicano del Seguro Social (IMSS)** como a **PENSIONISSSTE**, dos de las principales instituciones encargadas de gestionar los ahorros y las pensiones de los trabajadores en México. La Comisión Nacional del Sistema de Ahorro para el Retiro ha identificado casos de fraude en los que trabajadores, con la ayuda de asesores previsionales de las Afores, solicitan retiros por desempleo declarando salarios más altos de los que realmente perciben. Este problema, que afecta el equilibrio del sistema de pensiones, requiere una solución integral que incorpore herramientas de inteligencia artificial (IA) para detectar y prevenir fraudes.

2 Objetivo

El objetivo de este proyecto es desarrollar un sistema basado en IA para el **IMSS** y **PENSIONISSSTE**, que permita la detección y prevención de fraudes en retiros por desempleo mediante el análisis de patrones salariales, aportaciones y comportamientos sospechosos. La colaboración entre estas dos instituciones es fundamental, ya que ambas gestionan una gran cantidad de trabajadores y pueden compartir datos para crear un sistema más robusto.

3 Descripción del Proyecto

El proyecto se compone de las siguientes etapas clave, enfocadas en el **IMSS** y **PENSIONISSSTE**:

3.1 1. Detección de Anomalías en los Salarios Declarados

Descripción: Utilizando modelos de aprendizaje no supervisado, se analizarán los datos de salarios declarados en ambas instituciones para detectar patrones anómalos o discrepancias entre los salarios históricos y los reportados antes de un retiro por desempleo. Esto es especialmente relevante para **PENSIONISSSTE**, que administra los fondos de los trabajadores del sector público, y para el **IMSS**, que gestiona los fondos del sector privado.

Técnicas:

- *Isolation Forest*: Permite identificar casos de salarios anómalos en las bases de datos del IMSS y PENSIONISSSTE.
- *Autoencoders*: Ayuda a encontrar comportamientos atípicos en la evolución salarial de los trabajadores.
- *RNN (Redes Neuronales Recurrentes)*: Analiza secuencias temporales de salarios para detectar incrementos inusuales previos a los retiros.

3.2 2. Modelos Predictivos para Evaluar el Riesgo de Fraude

Descripción: Se desarrollarán modelos predictivos que, utilizando los datos de salarios, aportaciones y retiros por desempleo, clasificarán las solicitudes de retiro en función de su riesgo de fraude. ****PENSIONISSSTE**** y el ****IMSS**** compartirán esta información para evaluar los riesgos de manera conjunta y más eficiente.

Técnicas:

- *Random Forest*: Se entrenará con datos de contribuciones y salarios de ambas instituciones para clasificar el riesgo de fraude.
- *Gradient Boosting*: Mejorará las predicciones de riesgo con base en datos históricos de fraudes detectados.
- *Redes Neuronales*: Capaz de identificar relaciones complejas entre las variables de los trabajadores de ambas instituciones.

3.3 3. Análisis Comparativo entre Contribuciones y Salarios

Descripción: Este análisis comparará automáticamente las contribuciones al ****PENSIONISSSTE**** con los salarios declarados por los trabajadores en distintos momentos. Cualquier discrepancia significativa, como un aumento abrupto del salario previo a un retiro por desempleo, será reportada como sospechosa.

Técnicas:

- *Algoritmos de matching*: Comparación entre los registros del IMSS y PENSIONISSSTE para detectar inconsistencias.
- *Análisis de discrepancias*: Se usarán técnicas de detección de anomalías para identificar diferencias en las contribuciones reportadas en distintos periodos.

3.4 4. Procesamiento de Lenguaje Natural (NLP) en Comunicaciones

Descripción: En caso de detectar la participación de asesores previsionales en el fraude, se analizarán correos electrónicos, mensajes y otras comunicaciones entre trabajadores y asesores del ****IMSS**** y ****PENSIONISSSTE****. Esto permitirá identificar indicios de asesorías fraudulentas.

Técnicas:

- *BERT (Bidirectional Encoder Representations from Transformers)*: Un modelo que analizará grandes volúmenes de texto para detectar lenguaje sospechoso.
- *Análisis de sentimientos*: Para identificar el tono de las comunicaciones y posibles indicios de fraude.

3.5 5. Implementación de Alertas Basadas en Reglas

Descripción: Las reglas de negocio del ****IMSS**** y ****PENSIONISSSTE**** permitirán generar alertas automáticas cuando las solicitudes de retiro superen ciertos umbrales o cuando se detecten discrepancias salariales significativas. Estas reglas, combinadas con el análisis de IA, garantizarán un monitoreo efectivo.

Técnicas:

- *Sistemas basados en reglas*: Definición de políticas específicas del IMSS y PENSIONISSSTE para generar alertas.
- *Monitoreo en tiempo real*: Verificación de las solicitudes de retiro mientras son procesadas.

3.6 6. Monitoreo en Tiempo Real

Descripción: La IA se integrará en los sistemas del ****IMSS**** y ****PENSIONISSSTE**** para monitorear en tiempo real las solicitudes de retiro por desempleo, validando las contribuciones y salarios declarados antes de la aprobación de cualquier retiro.

Técnicas:

- *Sistemas de monitoreo en tiempo real*: Similares a los usados en la banca para prevenir fraudes financieros.

3.7 7. Identificación de Redes Fraudulentas

Descripción: Utilizando aprendizaje en grafos, se identificarán las relaciones entre trabajadores y asesores previsionales que colaboren en fraudes. Este análisis permitirá detectar redes organizadas que operen de manera recurrente en el ****IMSS**** o ****PENSIONISSSTE****.

Técnicas:

- *Graph Machine Learning*: Análisis de grafos para identificar nodos fraudulentos basados en sus relaciones y comportamientos.
- *Análisis de redes sociales*: Identificación de patrones de comportamiento en grupos relacionados con fraudes.

4 Infraestructura Tecnológica Necesaria

Para implementar un sistema de detección y prevención de fraudes en tiempo real, se requiere una infraestructura tecnológica robusta y especializada, que debe incluir los siguientes componentes:

- **Sistemas de procesamiento de datos de alto rendimiento:** Capacidad significativa de procesamiento para analizar grandes volúmenes de datos de salarios, contribuciones y retiros en tiempo real. Se sugiere el uso de servidores potentes en una arquitectura distribuida.
- **Almacenamiento de datos escalable:** Solución de almacenamiento que permita guardar datos históricos y actuales, así como modelos de IA. Bases de datos distribuidas como Hadoop o Spark son adecuadas para este propósito.
- **Herramientas de análisis de datos en tiempo real:** Plataformas de análisis de streaming como Apache Kafka o Apache Flink para ingerir, procesar y analizar datos a medida que llegan.
- **Integración con sistemas existentes:** La infraestructura debe integrarse con los sistemas del IMSS y PENSIONISSSTE para acceder a datos de trabajadores y solicitudes de retiro en tiempo real, lo que puede requerir el desarrollo de APIs o conectores específicos.
- **Sistemas de monitoreo y alerta:** Sistemas que monitoricen los resultados del análisis en tiempo real y generen alertas cuando se detecten posibles fraudes, dirigidas a los equipos de seguridad y fraude para su investigación.
- **Seguridad de los datos, redundancia y alta disponibilidad:** Consideraciones críticas para garantizar un funcionamiento continuo y confiable del sistema.

5 Recursos Necesarios

- **Equipo multidisciplinario:** Incluir expertos en IA, análisis de datos, ingenieros de software, así como especialistas en seguridad y cumplimiento normativo de ambos organismos.
- **Datos de calidad:** Acceso a bases de datos históricas y actuales sobre salarios, contribuciones, retiros por desempleo y registros de fraudes identificados. Este acceso debe ser garantizado por el IMSS y PENSIONISSSTE.
- **Financiamiento:** La obtención de recursos económicos para la implementación del sistema, mantenimiento y capacitación del personal involucrado será fundamental para el éxito del proyecto.

6 Beneficios Esperados

La implementación de este sistema de detección de fraudes en retiros por desempleo para el ****IMSS**** y ****PENSIONISSSTE**** tiene múltiples beneficios, que incluyen:

- **Reducción del fraude:** La implementación de un sistema automatizado de detección y prevención disminuirá el fraude en los retiros por desempleo, protegiendo los fondos de pensiones.

- **Aumento de la confianza en el sistema:** Al prevenir el fraude, se generará mayor confianza en el sistema de pensiones por parte de los trabajadores.
- **Mejora en la gestión de riesgos:** La identificación temprana de patrones de fraude permitirá a ambos organismos gestionar mejor los riesgos asociados con los retiros por desempleo.
- **Ahorro de costos:** La prevención de fraudes ahorrará recursos que de otro modo se perderían, permitiendo su re-inversión en servicios para los trabajadores.
- **Innovación en el uso de IA:** La implementación de este sistema será un caso de éxito en el uso de inteligencia artificial en el sector público, sentando un precedente para otros proyectos futuros.

7 Conclusiones

La combinación de inteligencia artificial y el acceso a datos del **IMSS** y **PENSIONISSSTE** ofrece una oportunidad única para detectar y prevenir el fraude en retiros por desempleo. Este proyecto no solo beneficiará a los trabajadores, sino que también protegerá los fondos de pensiones, asegurando un sistema más justo y sostenible.