

Fundamentos algorítmicos y ética en IA: curso de la teoría a la práctica

Herramienta para sesiones sincrónicas

UC3 | Algoritmos y sus limitaciones
Estudio de caso

Estudio de caso



Estudio de caso

Descripción para capacitadores

	Descripción
Descripción de la tarea	Los estudiantes analizarán un caso ficticio (ver la siguiente diapositiva) que involucra a una empresa de tecnología financiera (FinTech) que utilizó un sistema algorítmico para evaluar la solvencia crediticia para la aprobación de préstamos. La empresa tenía como objetivo proporcionar servicios de préstamos rápidos y eficientes a una demografía más amplia. Sin embargo, se descubrió más tarde que el algoritmo favorecía consistentemente a los solicitantes de ciertos tramos de ingresos y antecedentes raciales, lo que resultó en prácticas de préstamo discriminatorias. Los estudiantes deben investigar las causas raíz de este problema, evaluar las consecuencias en las comunidades afectadas y sugerir estrategias efectivas de mitigación.
Descripción de cómo realizar la tarea	Trabajo en grupo de 3-4 personas. Investigación y Análisis de Antecedentes • Revisar los materiales del curso y la literatura sobre sesgo algorítmico y equidad. • Entender los factores específicos que contribuyeron a los resultados discriminatorios de FinTech Innovations, incluyendo sesgos en los datos, características utilizadas y la ausencia de métricas de equidad. Identificación de Causas Raíz • Analizar las formas en que los datos de entrenamiento y las características de FinTech Innovations podrían haber incorporado sesgos sociales. • Investigar la ausencia de métricas de equidad y la falta de aportes de las partes interesadas. Recomendaciones de Mitigación Desarrollar un conjunto de recomendaciones para abordar los sesgos identificados, incluyendo diversificación de datos, nuevos procedimientos de validación, directrices éticas y auditorías continuas. Redacción y Entrega Organizar los hallazgos en un informe estructurado, proporcionando un resumen claro, análisis de las causas raíz y estrategias de mitigación detalladas. Incluir visuales, gráficos o tablas donde sea relevante para apoyar sus puntos.
Tiempo estimado para realizar la tarea	30-60 minutos.
Sugerencia de fuentes para realizar la tarea	Materiales del curso, bases de datos académicas, sistemas de IA (es decir, ChatGPT, Copilot, etc.) y búsquedas en internet.
Descripción detallada de cómo entregar la tarea	Escribir un informe resumen de 500-750 palabras.
Información sobre el plazo de entrega de la tarea	Durante la sesión sincrónica.
Información de contacto o cómo aclarar dudas	El profesor debe proporcionar una forma de contacto (puede ser una dirección de correo electrónico, un número de teléfono, etc.).

Caso ficticio para estudiantes

La industria de FinTech es conocida por su enfoque innovador para proporcionar servicios financieros rápidos y eficientes, a menudo a través del uso de sistemas algorítmicos para evaluar la solvencia y aprobar préstamos. Este caso ficticio gira en torno a una empresa de FinTech que empleó un avanzado algoritmo de puntuación crediticia para evaluar a los solicitantes y agilizar su proceso de aprobación de préstamos. La misión de la empresa era democratizar el acceso a los recursos financieros y apoyar a los clientes que a menudo son desatendidos por los bancos tradicionales. Sin embargo, su algoritmo resultó inesperadamente en prácticas de préstamo discriminatorias, favoreciendo sistemáticamente a solicitantes de ciertos tramos de ingresos y antecedentes raciales.

Análisis de la causa raíz

Antecedentes de la empresa

La empresa ficticia de FinTech, FinTech Innovations, fue establecida para interrumpir el sector bancario tradicional al ofrecer un sistema de aprobación de préstamos completamente en línea y sin fisuras. Este sistema utilizó un algoritmo propietario para evaluar el riesgo crediticio basado en varios factores, como ingresos, historial laboral, ubicación geográfica y antecedentes educativos. FinTech Innovations tenía como objetivo acceder a mercados ignorados por los bancos tradicionales, particularmente entre individuos más jóvenes, grupos minoritarios y aquellos con fuentes de ingresos no convencionales.

Identificación del problema

A pesar de sus objetivos declarados de inclusividad, el algoritmo de puntuación crediticia de FinTech Innovations favorecía consistentemente a los solicitantes de tramos de ingresos más altos y de antecedentes raciales específicos. Esto resultó en tasas desproporcionadamente altas de rechazo de préstamos para solicitantes de bajos ingresos y aquellos de minorías raciales. Las discrepancias en las tasas de aprobación de préstamos entre grupos demográficos generaron preocupaciones entre los grupos de defensa del cliente, lo que llevó a una auditoría externa que confirmó el comportamiento discriminatorio del algoritmo.

El análisis reveló varias causas raíz del sesgo. Primero, los datos históricos utilizados para entrenar el algoritmo reflejaban las disparidades sociales existentes, incorporando sesgos sistémicos en el modelo. Por ejemplo, utilizó la ubicación geográfica como una característica, que correlacionaba fuertemente con los ingresos y la raza. El algoritmo también dependía en gran medida de los datos de empleo tradicionales, pasando por alto a aquellos con fuentes de ingresos no convencionales como freelancers o trabajadores independientes. Finalmente, la falta de métricas de equidad durante la validación del modelo permitió que estos sesgos persistieran sin ser detectados.

Evaluación del impacto

Las prácticas discriminatorias de FinTech Innovations tuvieron repercusiones sociales y financieras significativas para los grupos afectados. Las tasas de rechazo de préstamos para minorías y solicitantes de bajos ingresos eran casi el doble que las de solicitantes de altos ingresos de antecedentes privilegiados. Esta exclusión resultó en oportunidades perdidas para el crecimiento financiero y reforzó aún más las desigualdades existentes. La empresa enfrentó daños a su reputación, escrutinio regulatorio y desafíos legales.

Estudio de caso

Caso ficticio para estudiantes

Estrategias de Mitigación

Para abordar estos problemas, FinTech Innovations implementó varias medidas correctivas:

- **Diversificación de Datos**

- Ampliaron el conjunto de datos de entrenamiento para incluir solicitantes diversos de diferentes niveles de ingresos, razas y tipos de empleo.

- **Métricas de Equidad**

- Incorporaron métricas de equidad para garantizar que las predicciones del algoritmo no fueran sesgadas y fueran representativas.

- **Medidas de Transparencia**

- La empresa se comprometió a compartir más información sobre cómo su algoritmo evaluaba la solvencia crediticia.

- **Guías Éticas**

- Nuevas pautas requerían monitoreo continuo, evaluación de impacto y participación de las partes interesadas.

Conclusión y Recomendaciones

Este caso ilustra las consecuencias no intencionadas de la toma de decisiones algorítmica en FinTech. Subraya la importancia de utilizar datos de entrenamiento diversos, validar algoritmos con métricas de equidad y mantener la transparencia para mitigar sesgos y discriminación. Los estándares éticos y los marcos regulatorios de la industria son esenciales para fomentar prácticas FinTech justas e inclusivas.

GRACIAS

Número de proyecto |
2022-1-ES01-KA220-HED-000085257



El apoyo de la Comisión Europea para la producción de esta publicación no constituye la aprobación de los contenidos, que reflejan únicamente las opiniones de los autores, y la Comisión no puede ser considerada responsable de ningún uso que se pueda hacer de la información contenida en ella.

