



Tecnológico de Monterrey

Campus:
Monterrey

Inteligencia Artificial Avanzada para la Ciencia de Datos (Gpo 102)

Curso:
TC3006C.102

Portafolio Análisis

Estudiante:

Jose David de la Garza Salas | A00834760

Lugar y Fecha:
Monterrey, Nuevo León
07/09/2024

1. Introducción

En este portafolio se desarrolla el proceso de análisis y evaluación de un modelo que machine learning y, Random Forest, junto al set de datos de Astronomical Data, acerca de la información de distintas estrellas y como las categorizamos dependiendo de su características esenciales, como este reporte solo se enfoca en el diagnóstico de los resultados y porque se decidió hacer cambios en el modelo en base a los resultados anteriores se da por sentado la limpieza y el procesamiento de los datos en el modelo. Este dataset(Astronomical Data) se ha seleccionado desde la página de Kaggle y a continuación se desglosa cada uno de los aspectos normativos acerca de su uso y restricciones, para conocer cuáles son nuestras limitaciones para su uso y distribución con el objetivo de poder utilizar el conjunto de datos sin miedo a violar terminos legales o alguna brecha de ley.

2. Análisis

Investigación de la Normativa Asociada al Dataset de Estrellas(Astronomical Data)

Para un dataset de datos astronómicos, es importante entender que este tipo de datos generalmente no incluye información personal o datos sensibles que podrían estar sujetos a las leyes de protección de datos personales. Sin embargo, aún se deben considerar las normativas de licencias de datos y propiedad intelectual, como la licencia Apache 2.0, bajo la cual se ha distribuido el dataset en Kaggle.

1. Licencia Apache 2.0

El dataset cuenta con la **Licencia Apache** que es una licencia de software de código abierto que permite el uso, la modificación y la distribución de los datos, siempre que se cumplan ciertas condiciones:

- **Atribución:** Debes proporcionar la atribución adecuada al creador original del dataset.
- **Distribución de código fuente:** Si modificas los datos y los distribuyes, debes hacerlo bajo la misma licencia y con el código fuente disponible.
- **Sin garantías:** La licencia establece que no hay garantías asociadas al uso de los datos.

2. Normativa Asociada a Datos de Estrellas

Los datos de estrellas, al ser datos científicos de observaciones astronómicas, no contienen información personal o sensible. Por lo tanto, las leyes de protección de datos personales como el GDPR en Europa, la CCPA en California, o la LGPD en Brasil no aplican directamente a este dataset.

Aspectos a considerar:

- **Uso científico o académico:** Los datos astronómicos son generalmente recopilados para investigación científica, y su uso está sujeto a prácticas éticas de investigación, como la atribución adecuada y el respeto por los términos de la licencia de datos.
- **Apache License 2.0:** Permite el uso libre de los datos siempre que se cumplan con los términos de la licencia.

3. Medidas Tomadas para Asegurar el Cumplimiento de la Normativa en el modelo

Para asegurar de no haber violado la normativa establecida bajo la Apache License 2.0, se han tomado las siguientes medidas:

- **Uso apropiado de los datos:** El uso del dataset se ha limitado a fines de análisis y aprendizaje, respetando las condiciones de la licencia. No se ha intentado monetizar o distribuir los datos sin atribución.
- **Atribución adecuada:** En todos los materiales derivados o trabajos que utilicen el dataset, se proporciona atribución completa a los autores originales tal como lo exige la licencia.

- **No hay modificaciones distribuidas sin fuente:** Si se realizan modificaciones al dataset y se comparten públicamente, se hará de acuerdo con los requisitos de la licencia, asegurando que el código fuente esté disponible.
- **Revisión de políticas de uso de datos:** Se ha revisado la política de uso de datos de Kaggle y la licencia Apache 2.0 para asegurarse de que se cumplan todas las condiciones establecidas.

Mi modelo cumple con la normativa por neutralidad de los datos. Mi uso de datos de estrellas son datos objetivos que no contienen atributos personales ni categorías sensibles que podrían llevar a sesgos injustos o discriminatorios.

Algoritmo Apropriado: El uso de mi modelo de Random Forest es adecuado para datos numéricos y categóricos y no introduce sesgos inherentes más allá de los que puedan estar presentes en el dataset original.

Transparencia en el Modelo: Random Forest es un algoritmo relativamente interpretable en comparación con otros modelos de machine learning más complejos. Esta transparencia ayuda a identificar posibles problemas y mitigar cualquier sesgo inadvertido en mi predicción.

Medidas para Mitigar Faltas Éticas

- **Documentación Clara y Transparente:** Asegúrate de proporcionar documentación clara sobre cómo se entrenó el modelo, sus limitaciones, y los supuestos realizados.
- **Licencia y Uso Restringido:** Si publicas el modelo, puedes especificar bajo qué condiciones puede ser utilizado, y dejar claro que cualquier mal uso está en contra de los términos de uso.
- **Validación Continua:** Realiza validaciones periódicas del modelo para asegurarte de que sigue siendo preciso y no se esté utilizando de manera inapropiada o con datos no adecuados.

3. Referencias

Apache License, Version 2.0, January 2004.

Licensed under the Apache License, Version 2.0 (the "License");

<http://www.apache.org/licenses/LICENSE-2.0>

Kaggle Astronomical Data

[Astronomical Data \(kaggle.com\)](#)

Normativa y legislación en PDP

Leyes en México para la protección de datos personales

https://micrositios.inai.org.mx/marcocompetencias/?page_id=370