1. OBJETIVO

Describir los aspectos metodológicos, fases y actividades predominantes, que hacen parte de los ejercicios de analítica avanzada en Agencia Distrital para la Educación Superior, la Ciencia y la Tecnología - ATENEA, incluyendo la formulación, implementación, seguimiento y evaluación, bajo el cumplimiento de los principios éticos y legales, que maximizan el impacto en la toma de decisiones basada en datos.

2. ALCANCE

La presente guía puede ser aplicada los ejercicios de analítica avanzada que realice cualquier dependencia de ATENEA, incluyendo los proyectos de análisis descriptivo, predictivo y prescriptivo. Por lo tanto, es una referencia que ayuda a garantizar la consistencia metodológica, el cumplimiento de principios éticos y legales, y la alineación con los objetivos estratégicos de la entidad.

3. DEFINICIONES

Artefacto: métodos o procesos de desarrollo específicos, es un producto tangible resultante del proceso de desarrollo de software

Analytics Solutions Unified Method (ASUM): "Método unificado de soluciones de análisis (ASUM)" tiene cinco fases: analizar, diseñar, configurar y crear, implementar y operar y optimizar. Sin embargo, las tres fases de ASUM de análisis, diseño y configuración y creación se han combinado aquí debido a la naturaleza iterativa de los proyectos de minería de datos/análisis predictivo. (ICESI, 2024).

API: “Interfaz de Programación de Aplicaciones". Se trata de un software intermediario que permite que las aplicaciones se comuniquen entre sí.

AUC: medida útil para comparar el rendimiento de dos modelos diferentes siempre y cuando el conjunto de datos esté equilibrado.

Ciencia de datos: un campo de investigación y práctica que se enfoca en resolver problemas del mundo real, a menudo usando grandes cantidades de datos y combinando habilidades de diferentes áreas del conocimiento: matemáticas, ciencias de computación, estadísticas, ciencias sociales e incluso periodismo de datos o arte.

Cross-Industry Standard Process for Data Mining (CRISP-DM): es un método probado para orientar sus trabajos de minería de datos. Como metodología, incluye descripciones de las fases normales de un proyecto, las tareas necesarias en cada fase y una explicación de las relaciones entre las tareas. (IBM, 2024).

Error: el error de un sistema es la diferencia entre el valor predicho resultado del modelo y el valor real de la variable que se está estimando (IA Responsable, BID, 2020).

Embedded: son sistemas de computación que pueden realizar funciones dedicadas en tiempo real.

Equipo desarrollador: profesionales encargados de crear, diseñar y corregir todo tipo de software, debe realizar las siguientes funciones

Equipo implementador: responsable de desarrollar y verificar los componentes del software que le fueron asignados, de acuerdo con los estándares establecidos en el proyecto

Front-end: "es la parte de una aplicación que interactúa con los usuarios, es conocida como el lado del cliente". (descubrecomunicacion, 2024).

Metodología: estrategia general que sirve de guía para los procesos y actividades que están dentro de un dominio determinado. No depende de tecnologías ni herramientas específicas, ni es un conjunto de técnicas o recetas. La metodología proporciona un marco sobre cómo proceder con los métodos, procesos y argumentos que se utilizarán para obtener respuestas o resultados.

MLOps (Operaciones de Machine Learning): son un conjunto de prácticas que agilizan el desarrollo, despliegue y mantenimiento continuo de modelos y flujos de trabajo de aprendizaje automático. MLOps integra prácticas como la integración continua (CI), el despliegue continuo (CD), pruebas automatizadas y monitoreo de modelos, todo con el objetivo de que los modelos de machine learning puedan ser implementados, escalados y mantenidos de manera más eficiente y confiable.

Scripts: programa insertado dentro del documento html o pequeño código de programación, que es interpretado y ejecutado por el navegador de un usuario.

Sesgo: error sistemático en una dirección o en un subconjunto específico de los datos (IA Responsable, BID, 2020). El sesgo per se no es motivo de prejuicio en la toma de decisiones, ya que puede ser por sí mismo una fuente de mitigación de otro tipo de errores provenientes de la fuente de los datos.

Stakeholders: aquellas personas y colectivos que están interesados, de un modo u otro, en nuestra entidad.

Team Data Science Process (TDSP): "es una metodología de ciencia de datos ágil e iterativa que puede usar para proporcionar soluciones de análisis predictivo y aplicaciones de IA de manera eficiente. El TDSP mejora la colaboración en equipo y el aprendizaje al recomendar formas óptimas de que los roles de equipo funcionen juntos". (MIcrosoft, 2024)

Versionamiento de datos: es la práctica de registrar los cambios que se realizan en los datos a lo largo del tiempo. Es un proceso que se utiliza para gestionar proyectos, especialmente en el desarrollo de software y el diseño digital.

Wrapper: programa que controla el acceso a un segundo programa

4. NORMATIVIDAD ASOCIADA:

- Ley 1266 de 2008, Por la cual se dictan las disposiciones generales del hábeas data y se regula el manejo de la información contenida en bases de datos personales, en especial la financiera, crediticia, comercial, de servicios y la proveniente de terceros países y se dictan otras disposiciones.
- Ley 1581 de 2012. Régimen General de Protección de Datos Personales y sus decretos reglamentarios
- Ley 1712 de 2014. Ley de Transparencia y Acceso a la Información Pública.
- CONPES 3920 de 2018: Política Nacional de Explotación de Datos
- CONPES 3975 de 2019: Política Nacional para la Transformación Digital e Inteligencia Artificial
- Decreto Distrital de Creación de la Agencia: 273 del 2020
- Acuerdo 822 de 2021 - Ciclo virtuoso de la seguridad, uso y aprovechamiento de datos
- Decreto Distrital 575 de 2023 - Gobernanza Infraestructura de Datos

5. DESARROLLO

5.1. REFERENTES METODOLÓGICOS PARA LA IMPLEMENTACIÓN DE EJERCICIOS ANALÍTICOS.

La presente guía de analítica avanzada está basada en marcos metodológicos ampliamente reconocidos como el *Cross-Industry Standard Process for Data Mining (CRISP-DM),* el *Analytics Solutions Unified Method* (ASUM) desarrollado por IBM y el *Team Data Science Process* (TDSP) desarrollado por Microsoft. Además, se han integrado prácticas emergentes de otros enfoques recientes como el *Machine Learning Operations* (MLOps), que promueve la automatización y gestión ágil de modelos analíticos a lo largo de su ciclo de vida.

5.2. ETAPAS DE LA METODOLOGÍA

La metodología de proyectos de analítica sugerida incorpora cinco etapas que son:

I. Comprender el problema

II. Gestionar los datos

III. Modelar

IV. Implementar

V. Retroalimentar.

5.2.1. Etapa I – Comprender el problema

Es quizá, la etapa más crucial, ya que establece la base para todo el ejercicio analítico. Implica definir claramente los objetivos y comprender a fondo la necesidad del usuario final, lo que asegura que el ejercicio analítico esté alineado con los problemas que se desean resolver. Para abordar esta etapa con éxito, es necesario responder las siguientes preguntas:

- ¿Cuál es el objetivo principal del ejercicio?
- ¿El ejercicio tiene como objetivo clasificar, detectar anomalías, hacer predicciones o generar recomendaciones?
- ¿Qué tiempo y recursos están disponibles para completar el ejercicio?
- ¿Cuál es el artefacto más adecuado para consumir los resultados del ejercicio analítico (e interfaz de usuario, reporte dinámico, etc.)?
- ¿Qué actores (stakeholders) estarán involucrados en el proyecto y cómo se gestionarán sus expectativas durante el ciclo del proyecto?
- ¿Cómo se integrarán consideraciones éticas, legales y sociales en el planteamiento del problema?
- ¿Qué requerimientos de producción y operación deben considerarse desde el inicio?

Productos y/o resultados esperados de la etapa de comprensión del problema: como resultado de esta primera etapa, se debe contar con la documentación y los artefactos necesarios para asegurar la correcta planificación y seguimiento del proyecto, que incluirán:

a. **Documento marco del proyecto:** este documento debe contener los objetivos del ejercicio, las preguntas que se pretende resolver, una revisión preliminar de la literatura relacionada y posibles variables a utilizar, así como información de los productos que se describen en los siguientes literales.

b. **Herramienta de planeación del proyecto:** se debe utilizar una herramienta de planificación (por ejemplo, Microsoft Planner, Microsoft Project, Excel o cualquier herramienta similar) para organizar el proyecto. Esta deberá contener como mínimo:

c. **Los miembros del proyecto y sus roles.**

d. **Las fases generales del proyecto con los plazos de ejecución acordados.**

e. **Las actividades preliminares necesarias para cumplir con cada fase.**

f. **Repositorio compartido:** se debe indicar el repositorio compartido en el que se trabajara el proyecto (usando herramientas como Git, GitHub, o similares) para consolidar versiones de código, documentación, análisis preliminares, entre otros recursos clave.

g. **Identificación de los datos disponibles o a gestionar en el proyecto**: se debe realizar una identificación preliminar de los datos disponibles y construir los metadatos y el diccionario de datos que detalle el significado y la estructura de las bases, tablas o archivos de información en caso de que no estén disponibles.

h. **Enfoque metodológico de análisis**: se debe definir el enfoque analítico que se aplicará en el proyecto, como por ejemplo la realización de un análisis descriptivo, la detección de anomalías, la predicción, la clasificación, identificar relaciones de causalidad, entre otros. Esta identificación ayudará a elegir las técnicas de modelado apropiadas.

i. **Diseño inicial de la solución**: se debe contemplar un diseño preliminar de cómo se presentarán y consumirán los resultados del modelo o ejercicio analítico, por ejemplo, a través de un servicio web en tiempo real, un *front-end* interactivo, un tablero de visualización, un reporte, entre otros.

j. **Identificación de requerimientos operacionales**: en este punto se debe determinar qué necesidades operativas de producción deben ser previstas desde el inicio, como escalabilidad, capacidad de integración con otros sistemas y frecuencia de actualizaciones del modelo o análisis.

5.2.2. Etapa II – Gestionar los datos

En esta etapa, se dispone de los datos en un entorno adecuado para el desarrollo de la solución analítica planteada, los explora en busca de inconsistencias o anomalías que deben ser corregidas, y establece un conocimiento profundo de la información disponible. Los pasos en esta etapa incluyen:

a. **Ingestión de datos en un entorno apropiado:** se trata de cargar los datos en un servicio o infraestructura tecnológica que permita la implementación de la solución diseñada. Esto puede implicar su almacenamiento en un servidor en la nube (como Oracle o cualquier otra que cuente con licenciamiento conforme a los lineamientos definidos por la Subgerencia de Tecnologías de la Información y las Comunicaciones), en un entorno local, o en sistemas híbridos, dependiendo de los requerimientos operativos y de escalabilidad.

b. **Auditoría y análisis de datos:** implica la exploración exhaustiva de los datos para identificar valores perdidos, datos anómalos (outliers), inconsistencias o errores que deban ser corregidos antes de la fase de modelado. Este análisis inicial debe incluir una revisión estadística de la distribución de los datos, las relaciones entre variables, y la coherencia con la variable objetivo del ejercicio.

c. **Preprocesamiento y creación de características (Feature Engineering)**: en algunos casos, los hallazgos de la auditoría pueden indicar la necesidad de limpieza o procesamiento previo de los datos. Esto puede incluir la imputación de valores faltantes, la normalización de variables, la eliminación o transformación de outliers, y la fusión o agregación de bases de datos. Estas transformaciones son clave para garantizar que los datos estén listos para la etapa de modelado y que el rendimiento del modelo no se vea afectado por datos de baja calidad.

Productos y/o resultados esperados de la etapa de gestión de datos: al finalizar esta etapa, se debe contar con los siguientes productos:

a. **Reporte de calidad de datos:** este informe debe describir de forma concisa los principales hallazgos del proceso de auditoría de datos, como la cantidad de valores faltantes, los outliers identificados, y cualquier otra anomalía detectada. También debe incluir las acciones tomadas para corregir estos problemas (por ejemplo, imputación de valores, eliminación de duplicados).

b. **Arquitectura de datos:** un diagrama o descripción técnica que detalle las fuentes de datos, los procesos de ingestión, y la disposición de los datos en el entorno en el que se desarrollará la solución. Este documento debe incluir también instrucciones o recomendaciones para facilitar el reentrenamiento del modelo, lo que puede incluir especificaciones sobre cómo manejar nuevas actualizaciones de datos**.**

c. **Notebook con análisis exploratorio de datos (EDA):** un notebook (o documento equivalente) con el código que describe el análisis inicial de los datos. Este análisis puede incluir:

- Estadísticas descriptivas de las variables
- Gráficos con las distribuciones de las variables más importantes.
- Análisis de correlación entre variables, así como pruebas estadísticas de significancia con respecto a la variable objetivo (por ejemplo, pruebas *t-student*, chi-cuadrado, o *F-Fisher)*.

Durante esta etapa es fundamental manejar versionamiento de datos de las diferentes fuentes, así como evaluar la calidad de datos con diferentes métricas, para lo cual es importante revisar la Guía de Calidad de Datos de ATENEA.

5.2.3. Etapa III – Modelar

En esta fase, se consideran los datos preparados en la etapa anterior y se utilizan para generar información relevante y útil que contribuya a la resolución del problema planteado. El modelado no se limita únicamente a la predicción, sino que puede incluir tareas como:

- Clasificar una población en grupos.
- Detectar anomalías en los datos para identificar patrones o características comunes.
- Establecer estrategias de acción mediante la construcción de escenarios artificiales.
- Describir y analizar datos previamente no detectados debido al volumen o complejidad de la información.

Cualquiera que sea el objetivo planteado, el modelaje requiere del desarrollo de los siguientes pasos:

I. **Feature engineering final:** en esta etapa se crean, adicionan o transforman las variables con el objetivo de que se ajusten a los requerimientos temáticos y técnicos de los métodos de

modelación disponibles. Se trata de encontrar un balance entre la experiencia o conocimiento temático de la problemática, la información disponible y la utilidad estadística de las variables.

Este proceso puede implicar transformaciones específicas (como estandarización o codificación de variables categóricas) para mejorar la capacidad predictiva y reducir la complejidad del modelo.

Se recomiendan técnicas de selección de variables como métodos de tipo filtro (correlación, chi-cuadrado), *wrapper (Sequential Forward Selection*, SBS, *Recursive Feature Elimination*) o *embedded* (L1, L2, *elastic net*).

II. **Entrenamiento, selección y evaluación del modelo:** el entrenamiento del modelo es un proceso iterativo que consiste en

- Dividir los datos en conjuntos de entrenamiento y evaluación,
- entrenar diferentes modelos con los datos, y
- compararlos usando métricas de evaluación.

En esta fase, se evalúan los resultados obtenidos y los supuestos del modelo, buscando siempre optimizar el desempeño sin caer en el sobreajuste (*overfitting*).

Cada modelo entrenado debe ser versionado para asegurar trazabilidad y reproducibilidad. Es recomendable usar signos pipe (|) automatizados para que el proceso de entrenamiento y evaluación se ejecute de manera repetitiva y reproducible.

En el desarrollo de los modelos es fundamental abordar las siguientes consideraciones metodológicas:

a. Tipo de modelo: existen distintos tipos de modelos de aprendizaje automático, y es esencial seleccionar el más adecuado para los objetivos del ejercicio analítico. A continuación, se explican las principales categorías de modelos:

- **Aprendizaje supervisado:** cada observación en los datos está asociada a una etiqueta o valor objetivo que se desea predecir. Ejemplos comunes incluyen la predicción del precio de un producto basado en sus características o la clasificación de correos electrónicos como spam o no spam. Este tipo de modelos es útil cuando se cuenta con un conjunto de datos etiquetados y se desea aprender de estos para hacer predicciones sobre datos nuevos. Los modelos de regresión, clasificación, o árboles de decisión pertenecen a esta categoría. Estos modelos deben ser entrenados, validados y monitorizados continuamente para mantener su precisión a lo largo del tiempo.
- **Aprendizaje no supervisado**: en este enfoque, no se tienen etiquetas predefinidas, y el objetivo es descubrir patrones o estructuras subyacentes en los datos. Esto se utiliza principalmente cuando se quiere explorar grandes volúmenes de datos sin un resultado claro predefinido. Los algoritmos de clústeres, como *k-means*, agrupan las observaciones en grupos o categorías con características similares. Estos modelos son útiles para tareas como la segmentación de clientes, reducción de dimensionalidad o detección de anomalías. La implementación y el monitoreo de modelos no supervisados debe incluir mecanismos para

validar continuamente que las agrupaciones o estructuras descubiertas sigan siendo relevantes y útiles.

- **Aprendizaje por refuerzo:** este tipo de modelo se enfoca en la toma de decisiones a través de la interacción con un entorno, donde el agente aprende mediante ensayo y error, recibiendo recompensas o penalizaciones por las acciones que toma. Es comúnmente utilizado en áreas como robótica, juegos o sistemas autónomos (como vehículos sin conductor). El aprendizaje por refuerzo presenta retos adicionales en su implementación y monitoreo, ya que los entornos pueden cambiar dinámicamente, esto requiere un enfoque robusto para la simulación, el testeo automatizado y la actualización continua del agente entrenado.

Al momento de decidir por el modelo a usar, es útil identificar el tipo de problema, la disponibilidad de datos etiquetados y el objetivo del análisis.

a. **Coherencia entre la variable objetivo y el problema diagnosticado:** uno de los principales desafíos en la interpretación de resultados de modelos es asegurar la coherencia entre la definición de la variable objetivo y el problema diagnosticado. Sin embargo, es común que la definición de esta variable involucre supuestos o aproximaciones que se alejan, en mayor o menor medida, de la representación ideal de la problemática cualitativa que el área o equipo busca resolver. Estas discrepancias pueden generar sesgos en la interpretación y desacuerdos tanto teóricos como prácticos sobre la implementación del modelo.

   Por lo tanto, es fundamental que el equipo de la modelación documente de manera precisa y transparente cómo se construyó la variable objetivo, incluyendo los supuestos asumidos en este proceso. Además, debe haber una retroalimentación constante con expertos temáticos internos o externos para asegurar que todas las partes comprendan y acuerden la representación del problema.

   Además, durante la gestión de los datos, se debe realizar un análisis exhaustivo de la calidad de estos, considerando aspectos como la completitud y coherencia de los registros. Esto incluye asegurar que los datos sean lo más representativos posible y que no existan sesgos ocultos. También debe incorporarse un análisis de la composición de la base de datos respecto a variables clave como edad, sexo, grupo étnico y departamento, ya que estas dimensiones pueden influir en los resultados finales del modelo.

b. **Selección adecuada y proporcional de la metodología:** la selección de modelos de Machine *Learning* deberá realizarse siempre considerando otras posibles metodologías que puedan ofrecer mejores resultados para el problema planteado. Se priorizarán aquellos enfoques que ofrezcan un mayor nivel de explicabilidad, interpretabilidad y simplicidad, siguiendo el principio de que la transparencia es crucial para la validación y confianza en los resultados. En este sentido, cuando sea posible, se preferirán modelos con alta capacidad de interpretación estadística, como regresiones y modelos logísticos.

   Si se determina que un modelo con datos agregados no es adecuado, se deberá hacer un análisis exhaustivo de las fuentes de datos disponibles, evaluando sus limitaciones

metodológicas. Es clave identificar sesgos en la recolección de datos, posibles errores de registro en los sistemas administrativos, y cualquier otro factor que pueda distorsionar la distribución de los datos utilizados para el entrenamiento del modelo. Esto implica, además, analizar si existen desbalances en las categorías de la variable objetivo o entre otras variables demográficas, como edad, sexo, grupo étnico, o región.

Si se identifican desbalances, los responsables del modelado, deberá seguir los siguientes pasos:

i. *Recolectar más información:* o complementar la existente mediante cruces con fuentes confiables y oficiales que puedan mejorar la calidad y representatividad de los datos.

ii. *Aplicar mecanismos de balanceo:* cuando persistan diferencias sustanciales en la distribución de las categorías. Entre las técnicas disponibles se encuentran el sobre muestreo de las categorías minoritarias, el submuestreo de las mayoritarias, o la combinación de ambas. También pueden utilizarse algoritmos que asignen pesos diferenciados a las categorías, como una red neuronal ponderada o un modelo logit penalizado, para mitigar el impacto del desbalance.

iii. *Modelos con cortes de probabilidad diferenciada:* es importante no adoptar de manera rígida un umbral estándar de probabilidad (como el 0.5). En su lugar, se deben analizar cortes de probabilidad ajustados a la naturaleza del problema y los desbalances existentes en los datos.

iv. *Descarte de registros incompletos:* se debe establecer un protocolo claro para el manejo de registros incompletos o incoherentes, asegurando que estos cumplan con reglas mínimas de calidad desde el inicio del proceso.

v. *Decisiones frente a desbalances demográficos:* es crucial implementar un mecanismo de decisión que permita abordar desbalances en la variable objetivo relacionados con características demográficas, como edad, grupo étnico o sexo. Este mecanismo debe ser parte integral de las métricas de selección del modelo y alinearse con las variables protegidas, garantizando que el modelo sea justo y no discrimine a ningún grupo.

a. **Definición de métricas adecuadas para la selección de modelos:** la mayoría de los problemas abordados por ATENEA mediante herramientas analíticas corresponden a problemas de clasificación. En este contexto, la selección de la métrica adecuada para evaluar y seleccionar el mejor modelo es fundamental, ya que esta depende directamente de la necesidad planteada por el área misional.

   Para evaluar la efectividad de los modelos de clasificación, se suelen utilizar métricas derivadas de la matriz de confusión, que compara los resultados predichos con los reales. La matriz de confusión es la base para calcular diversas medidas que ayudan a entender el rendimiento del modelo en términos de verdaderos positivos (VP), verdaderos negativos (VN), falsos positivos (FP) y falsos negativos (FN).

	Clasificación Real Positiva	**Clasificación Real Negativa**

Predicción Positiva	Verdadero Positivo (VP)	Falso Positivo (FP)
Predicción Negativa	Falso Negativo (FN)	Verdadero Negativo (VN)

De acuerdo con lo anterior, el objetivo de un modelo de clasificación es maximizar tanto los verdaderos positivos (VP) como los verdaderos negativos (VN), lo que implica optimizar la precisión general del modelo. Sin embargo, en muchos casos, alcanzar un equilibrio perfecto entre ambas categorías no es posible, y será necesario priorizar uno de los dos resultados en función del problema específico.

Para esto, son necesarias algunas métricas clave para la evaluación de modelos:

- Exactitud (Accuracy): es la proporción de predicciones correctas sobre el total de predicciones. Se calcula dividiendo la suma de VP y VN entre el total de observaciones (VP + VN + FP + FN). Es una métrica global que indica qué tan bien el modelo predice tanto positivos como negativos.
- Precisión (Precision): se refiere a la fracción de ejemplos clasificados como positivos que en realidad son positivos. Se calcula dividiendo VP entre la suma de VP y FP. Esta métrica es especialmente útil cuando es importante minimizar los falsos positivos.
- Especificidad (Specificity): es la fracción de negativos correctamente clasificados, calculada como VN dividido entre la suma de VN y FP. Indica la capacidad del modelo para identificar correctamente los ejemplos negativos.

La selección de una métrica específica, o la combinación de varias, dependerá de los siguientes factores:

i. *Importancia de las categorías:* si hay una categoría o clase más importante que otras, la métrica debe reflejar esta prioridad. Por ejemplo, en un modelo de predicción de deserción en educación superior, puede ser más crucial identificar a los estudiantes con alto riesgo de deserción para que se puedan tomar medidas preventivas. En este caso, sería preferible maximizar la recuperación de los verdaderos positivos (es decir, identificar correctamente a los estudiantes que van a desertar), incluso si eso implica aceptar un mayor número de falsos positivos (estudiantes identificados como en riesgo de deserción que en realidad no desertan). Esto permitiría asegurar que la mayoría de los estudiantes en riesgo reciban apoyo, aunque algunos que no estén en riesgo también lo reciban de forma preventiva.

ii. *Tolerancia a falsos negativos:* el nivel de falsos negativos aceptable es una decisión clave que debe ser discutida por el equipo interdisciplinario. En el ejemplo de la deserción en educación superior, un falso negativo representaría un estudiante que estaba en riesgo de desertar, pero no fue identificado por el modelo, lo cual podría tener consecuencias significativas si no recibe la intervención adecuada. Por tanto, la métrica elegida debe minimizar estos errores tanto como sea posible, de acuerdo con las prioridades misionales.

Productos y/o resultados esperados en de la etapa de modelado: *a*l finalizar la etapa de modelado, el equipo debe entregar los siguientes productos:

a. **Códigos implementados:** todo el código utilizado para el feature engineering, entrenamiento y evaluación del modelo debe estar versionado y documentado en un repositorio (como GitHub, GitLab, Azure DevOps, etc.). Este repositorio debe incluir también descripciones claras del proceso de creación de características y los modelos entrenados. Es importante que el código sea modular y siga buenas prácticas de ingeniería de software, como la implementación de pruebas automatizadas para validar la funcionalidad del código (testing).

b. **Reporte del modelo:** el reporte debe contener un resumen de cada modelo entrenado, con métricas clave como precisión, sensibilidad, AUC, precisión ponderada, entre otros, según el tipo de problema. Además, debe incluir detalles sobre los supuestos y limitaciones de cada modelo, de forma que se pueda justificar la elección del modelo final.

5.2.4. Etapa IV – Implementar

La etapa de implementación se centra en dar vida a los modelos desarrollados en la fase de modelado para que sean utilizados de forma efectiva y continua por otras aplicaciones y usuarios finales. Esta etapa implica desplegar los modelos y exponer sus resultados para consumo activo, utilizando interfaces accesibles para otras plataformas y equipos. En este contexto, la integración con prácticas de MLOps asegura que los modelos sean desplegados, monitoreados y gestionados de manera eficiente y reproducible, a lo largo del ciclo de vida del modelo.

En esta etapa son claves los siguientes pasos:

a. **Despliegue del modelo:** para hacer que los modelos sean accesibles, se considera el desarrollo de APIs o cualquier otra herramienta que permitan a otras aplicaciones consumir los resultados de forma sencilla y eficiente o hacer disponible la información para consulta. Esto facilita la integración con:

   - Sitios web
   - Hojas de cálculo
   - Tableros de control
   - Aplicaciones front-end
   - Aplicaciones back-end

   Se debe garantizar que el modelo desplegado esté versionado correctamente, de modo que cualquier cambio en los datos o en el código del modelo pueda ser rastreado y revertido si es necesario.

   Si los recursos o el alcance del proyecto no permiten un consumo activo o en tiempo real de los modelos, es crucial que los resultados se pongan a disposición de las partes interesadas de la forma más amigable posible. Esto puede incluir:

   - Informes periódicos automatizados que incluyan predicciones o resultados clave.
   - Dashboards que presenten visualizaciones interactivas de los resultados.

- Hojas de cálculo pre-procesadas con predicciones.

a. **Reentrenamiento continuo**: se debe planificar el reentrenamiento automático del modelo cuando se detecten cambios en los datos o se degraden las métricas de rendimiento. Esto puede lograrse configurando pipelines que disparen el reentrenamiento cuando se cumplan ciertas condiciones

b. **Cumplimiento de lineamientos internos**: cualquier artefacto de implementación (APIs, dashboards, aplicaciones, etc.) debe cumplir con los lineamientos y normativas establecidas por la Agencia. Esto incluye asegurar que la arquitectura y el desarrollo estén alineados con las políticas de seguridad, tratamiento de datos personales y estándares técnicos o de infraestructura tecnológica que señale la Subgerencia de Tecnologías de la Información y las Comunicaciones.

Productos y/o resultados esperados de la etapa de implementación: Los principales productos de esta etapa (aunque no los únicos, ya que dependerá de la solución) son los siguientes:

a. **Reporte del modelo final:** Este reporte debe incluir una descripción detallada del modelo final seleccionado, sus métricas de evaluación, y su proceso de implementación. La documentación debe estar almacenada en un repositorio accesible (como GitHub, o aquel disponible) para permitir su revisión y mantenimiento por el equipo o por áreas externas.

   La documentación también debe incluir scripts reproducibles y pipelines, asegurando que cualquier miembro del equipo pueda recrear el entorno de entrenamiento y despliegue del modelo. Adicionalmente, se debe documentar el proceso de validación y los supuestos bajo los cuales fue entrenado el modelo

b. **Un documento de solución arquitectónica final**: Este documento describe la arquitectura final utilizada para la implementación del modelo, incluyendo cómo se expone el modelo a través de APIs, la infraestructura utilizada (nube, on-premises, etc.), y las herramientas de monitoreo y gestión implementadas.

5.2.5. Etapa V – Retroalimentar

Esta etapa se enfoca en asegurar que los resultados del modelo sean validados, criticados y mejorados mediante la retroalimentación tanto de expertos como de los usuarios que implementan el modelo en el entorno productivo. Este proceso busca obtener mejoras continuas y garantizar que el modelo siga siendo útil y relevante en el contexto misional. La retroalimentación en MLOps no solo proviene de la interacción con los usuarios, sino también del monitoreo constante del desempeño del modelo en producción, detectando problemas y oportunidades de mejora a medida que los datos y las condiciones cambian.

En esta etapa serán importantes los siguientes pasos:

a. Valid**ación con expertos:** es fundamental comunicar los resultados del modelo y los principales parámetros a un grupo de expertos para recibir críticas constructivas, estos pueden ser internos o externos. Este proceso incluye revisar:

- La metodología y los enfoques utilizados.
- Las variables seleccionadas y su relevancia.
- Las métricas de evaluación y su validez.

b. **Validación de resultados "en campo":** es imperativo mantener una comunicación constante con el equipo o área misional que utiliza los resultados, obteniendo validaciones de primera mano sobre el desempeño del modelo en su contexto de uso. Esto permite identificar problemas prácticos y ajustarse a las necesidades operativas reales.

c. **Evaluación de impacto:** además de las métricas técnicas, es crucial que los resultados del modelo sean medidos en términos de su impacto en los objetivos misionales o de negocio del equipo que lo implementa. Esto asegura que las mejoras en el modelo estén alineadas con los resultados esperados en la realidad operativa.

Productos y/o resultados esperados de la etapa de retroalimentación: el producto de esta etapa es el siguiente:

a. Retroalimentación constante del documento marco del proyecto: este documento debe contener en tiempo real el avance del ejercicio, incluyendo no solo los resultados obtenidos, sino también las observaciones de los expertos y del equipo misional sobre el desempeño del modelo. Además, debe incluir un registro detallado de las métricas de monitoreo en producción, las alertas generadas y cualquier ajuste realizado al modelo.

6. CONSIDERACIONES ÉTICAS Y LEGALES PARA ABORDAR LOS EJERCICIOS ANALÍTICOS

El uso de los algoritmos de modelación utilizados deberá hacerse de manera responsable y considerando los posibles aspectos discriminatorios que podría traer consigo su aplicación, principalmente cuando se trate de personas como lo puede ser la aplicación de algoritmos de beneficiarios de ATENEA. En este sentido, deberá prevalecer para su uso los principios éticos de justicia y beneficio hacia la población objetivo.

En caso de que se identifiquen riesgos de carácter ético, de justicia o beneficio, el equipo que desarrolla el ejercicio analítico deberá acudir a técnicas para su mitigación a través de la entrega agregada de resultados o de la modificación parcial o total del método o artefacto de implementación de los resultados del modelo.

6.1. PRINCIPIOS ÉTICOS PARA CONSIDERAR EN LOS EJERCICIOS DE ANALÍTICA

Los siguientes son los principios éticos que regirán de manera transversal a los ejercicios de analítica desarrollados al interior de ATENEA.

a. **La transparencia es la base de la ética:** Todo ejercicio de analítica descriptiva, predictiva, prospectiva o diagnóstica desarrollada para ATENEA debe ser transparente y suficientemente clara para todo aquel que quisiera hacer una revisión independiente de sus resultados. En este sentido, debe contar con la documentación que dé cuenta clara de cada una de las etapas de su desarrollo, así como de la interpretación y el alcance de sus resultados. Así mismo, todo

modelo implementado debe tener un componente de interpretabilidad y explicabilidad, sea este un modelo de Machine Learning o no. Para los casos de algoritmos de machine learning, se deberán seleccionar metodologías que puedan dar cuenta del aporte de las variables, como por ejemplo SHAP (Shapley Additive Explanations), LIME (Local Interpretable Model-Agnostic Explanations) o Permutation Feature Importance.

Estas herramientas permiten evaluar de forma detallada el impacto que cada variable tiene sobre el resultado del modelo, facilitando su comprensión por parte de los expertos temáticos y misionales, así como de cualquier interesado en revisar el proceso. Asimismo, el uso de estas metodologías ayuda a garantizar que las decisiones derivadas del modelo puedan justificarse de manera clara, lo que refuerza la confianza en su aplicación.

b. **La misionalidad orienta:** Todo ejercicio de analítica desarrollado por ATENEA debe ser inspirado por un problema específico del diseño estratégico o de la operación de sus programas y servicios. De este modo, deben contemplar las necesidades de las dependencias misionales o de apoyo y las soluciones que se planteen, deben ser proporcionales a las problemáticas identificadas.

c. **Vigilancia sobre la protección de los datos personales:** Todos los ejercicios de analítica desarrollados al interior de la agencia deberán ceñirse a las disposiciones reglamentarias y normativas existentes en el país en materia de protección de datos personales y uso ético de la información tomando como marco de referencia la ley 1581 de 2012. También deberá ajustarse a las políticas de seguridad de la información y tratamiento de datos personales de ATENEA.

d. **Igualdad y no discriminación**: El uso de los algoritmos de modelación utilizados deberá hacerse de manera responsable y considerando los posibles aspectos discriminatorios que podría traer consigo su aplicación. En este sentido, deberá prevalecer para su uso los principios éticos de justicia y beneficio hacia la población objetivo de la agencia.

e. **Explicabilidad como herramienta de confianza**: La interpretabilidad de los resultados y de la metodología empleada es crucial para que los usuarios confíen en los resultados de los ejercicios analíticos por lo que es preciso comunicar las capacidades y la finalidad de los algoritmos o herramientas desarrolladas y que las decisiones deban poder explicarse, en la medida de lo posible, a las partes implicadas.

f. **La mitigación de riesgos es obligatoria**: El desarrollo o implementación de los ejercicios analíticos no deberá desencadenar situaciones de discriminación y odio. En caso de que se identifiquen riesgos de este tipo, el equipo que desarrolla el ejercicio analítico deberá acudir a técnicas para su mitigación a través de la entrega agregada de resultados o de la modificación parcial o total del método o artefacto de implementación de los resultados del modelo.

g. **Se deben considerar todas las necesidades**: Los ejercicios analíticos desarrollados por ATENEA deben tener en cuenta las necesidades específicas de cada uno de los grupos poblacionales que son impactados por sus programas y servicios, así como de las dependencias misionales y de apoyo. De esta forma, son igualmente importantes las miradas desde el enfoque diferencial e interseccional como las que resulten del análisis situacional y las necesidades organizacionales.

h. **Responsables de las decisiones que tomamos**: Los sistemas de soporte en la toma de decisiones basados en modelos sean estos de inteligencia artificial o no, deben ser supervisados en cada etapa por un grupo de personas con conocimientos interseccionales relevantes para el proyecto, dentro de las cuales debe considerarse la experiencia del área misional o de apoyo que se impacta o beneficia por el proyecto.

i. **Compartir metodologías y resultados**: La Agencia adoptará siempre una metodología colaborativa para el desarrollo de sus ejercicios de analítica.

j. **Difundir es esencial**: La Agencia tiene el deber social de cooperar en la generación de conocimiento social sobre la población objeto de sus programas y servicios. Toda herramienta de analítica desarrollada al interior de ATENEA podrá ser difundida con la academia y la sociedad civil, de modo que se gestione el conocimiento derivado a nivel metodológico y temático.

k. **Retroalimentación y mejora continua**: Es un deber ético mejorar permanentemente los desarrollos analíticos impulsados por ATENEA. La documentación del proyecto deberá proporcionar pautas que permitan evaluar la herramienta cada cierto tiempo. Además, se deberán prever recursos para diagnosticar y generar cambios que se consideren fundamentales para que la herramienta siga siendo relevante y óptima.

6.2. VIGILANCIA SOBRE VARIABLES QUE PUEDAN GENERAR DISCRIMINACIÓN

El equipo que adelante el procesamiento deberá seleccionar una definición de justicia algorítmica sobre la cual realizará vigilancia. Las definiciones que se adoptan son:

a. Equidad de posibilidades en modelos predictivos y de clasificación: las tasas de error predictivo sobre cada subgrupo definido por una variable protegida o vigilada para el análisis deben ser similares. En modelos de clasificación binaria, las tasas de falsos positivos y falsos negativos también deben estar equilibradas entre estos subgrupos. Por ejemplo, si se crea un modelo para orientar a los estudiantes en la elección de una carrera y se utiliza el sexo como variable, no debe presentarse el caso que, al ser mujer, el modelo sugiera o muestre menos opciones en áreas como ingeniería, mientras que a los hombres se les muestre más. Si el modelo muestra esa tendencia, se estaría incurriendo en un sesgo de género que afecta negativamente la equidad de la orientación. Es esencial que cualquier sesgo implícito en el modelo sea identificado y corregido para garantizar que se ofrezcan opciones educativas justas y equilibradas, independientemente del sexo del estudiante.

b. Justicia contrafactual y su implicancia en la predicción de resultados: implica que los resultados de la predicción de un modelo no cambian cuando se altera alguna de las variables protegidas o vigiladas. En el caso de un modelo de orientación vocacional, si la elección de carrera de un joven se ve alterada al cambiar el sexo del individuo (por ejemplo, si al cambiar el sexo de mujer a hombre, el modelo sugiere más opciones en áreas como ingeniería), el modelo estaría mostrando un sesgo inaceptable. Los resultados de este modelo deben ser consistentes y justificados, independientemente del sexo del estudiante, y los factores que influyan en las recomendaciones deben ser basados en intereses y/o habilidades, no en estereotipos de género.

c. **Consideraciones sobre el sesgo justificable:** es importante reconocer que, dependiendo del contexto y del problema, lo que parece ser un sesgo podría no serlo. Por ejemplo, en situaciones como el riesgo de deserción escolar en estudiantes provenientes de contextos de vulnerabilidad socioeconómica o de pertenencia étnica determinada, la diferencia en las tasas de deserción no necesariamente se considera un sesgo, sino más bien una realidad que refleja una desigualdad estructural. En estos casos, el modelo debe detectar estos riesgos para intervenir de manera proactiva y beneficiar a los estudiantes en situación de vulnerabilidad. Un sesgo en este contexto no sería un error, sino una señal de que se deben tomar acciones correctivas, para mitigar el riesgo y promover la equidad.

Una vez definido el concepto de justicia adecuado, el equipo deberá seleccionar una o más variables protegidas sobre las que realizará seguimiento durante la fase de entrenamiento y selección del modelo.

A continuación, el equipo deberá:

i. Evaluar el grado de alineación de las predicciones con la definición de justicia seleccionada, utilizando las métricas previamente elegidas.
ii. Determinar puntos de corte en la probabilidad que sean adecuados y diferenciados por cada subgrupo de la variable vigilada, con el fin de alcanzar el nivel de justicia deseado.
iii. Corregir cualquier desbalance en la información de las clases o grupos minoritarios, para mejorar la precisión de las predicciones en estos grupos.

6.3. INTERPRETABILIDAD

Cómo se mencionó previamente, los responsables del desarrollo priorizarán siempre aquellos modelos que ofrezcan una mejor forma de explicar e interpretar sus resultados. Además, se evaluará el rendimiento de los modelos utilizando distintos conjuntos de variables, por ejemplo:

- Un conjunto con el 100% de las variables disponibles,
- un conjunto con el 75% de las variables más relevantes,
- un conjunto con el 50% de las variables más significativas, y
- un conjunto con el 25% de las variables clave.

Luego, se prioriza el modelo que demuestre mayor poder explicativo con el menor número de variables, porque este es más fácil de explicar y divulgar.

Asimismo, en la difusión de los resultados, se buscará siempre la manera más sencilla de explicar tanto el proceso de desarrollo como los resultados, con el fin de fomentar una mayor apropiación del ejercicio.

6.4. DOCUMENTACIÓN COMPLETA Y PERFILAMIENTO

Los proyectos de analítica deberán contar con un perfilamiento detallado del modelo entrenado y seleccionado, que permita conocer de manera concisa lo siguiente:

a. Los alcances del modelo, especificando los límites de su aplicación;

b. Las limitaciones metodológicas, destacando posibles sesgos o restricciones en los datos o procesos;

c. Las preocupaciones éticas, identificando posibles riesgos o implicaciones en la implementación del modelo;

d. Las recomendaciones para el reentrenamiento, señalando situaciones que podrían justificar ajustes futuros;

e. Las fuentes de datos utilizadas y las consideraciones sobre su actualización, que puedan afectar el rendimiento del modelo;

f. Los resultados obtenidos, según las métricas elegidas, junto con una justificación clara de la selección de esas métricas;

g. Una descripción general de la metodología empleada y las tecnologías involucradas, explicando cómo se implementaron y su relevancia para el modelo.

6.5. MONITOREO PERMANENTE

Es una obligación del equipo desarrollador o implementador garantizar que se establezca un monitoreo continuo de la optimización de las predicciones del modelo, considerando los siguientes aspectos:

a. Potenciales cambios en las fuentes de datos que puedan afectar la calidad de recolección, las categorías de las variables, la conceptualización o el diseño de estas, y cómo esto influye en la relación entre las variables de entrada y la predicción.

b. Evolución de las distribuciones reales de las variables de interés, tales como el grupo étnico, el sexo, el estado migratorio, entre otros.

c. La necesidad de reentrenamiento del modelo debido a la disponibilidad de una gran cantidad de datos nuevos.

d. Impactos de eventos externos que afecten las hipótesis del modelo, como, por ejemplo, los efectos de la pandemia por Covid-19.

e. Cambios observados en las métricas relacionadas con variables protegidas o vigiladas, que podrían indicar la necesidad de ajustes.

6.6. DIAGNÓSTICO ÉTICO INDEPENDIENTE

La Agencia deberá realizar un diagnóstico ético de la herramienta desarrollada tanto antes como después de su implementación. Este diagnóstico podrá ser realizado de manera independiente, siempre que los recursos estén disponibles. Sin embargo, también podrá apoyarse en herramientas

de autodiagnóstico, como las desarrolladas por instituciones como el Banco Interamericano de Desarrollo (BID) a través de Fair-Lac.

Por otro lado, ATENEA deberá llevar a cabo ejercicios piloto de implementación que permitan evaluar los alcances, las dificultades y las oportunidades de mejora de las herramientas analíticas desarrolladas, además de verificar la veracidad de las hipótesis formuladas.

7. DOCUMENTOS DE REFERENCIA:

- IBM (2015). Metodología Fundamental para la Ciencia de Datos. https://www.ibm.com/downloads/cas/6RZMKDN8
- IBM (2016). Analytics Solutions Unified Method (ASUM). ftp://ftp.software.ibm.com/software/data/sw-library/services/ASUM.pdf
- Comisión Europea (2018) – Grupo Independiente de Expertos de Ato Nivel sobre Inteligencia Artificial. Directrices éticas para IA Fiable. https://op.europa.eu/es/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1
- MICROSOFT (2020). Teams Data Science Process. https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/overview
- Banco Interamericano de Desarrollo-BID (2021). Autoevaluación ética: mitigación de riesgos asociados al uso/aplicación de nuevas tecnologías.
- Departamento Administrativo de la Presidencia (2021). Marco Ético para la Inteligencia Artificial en Colombia.
- ML-Ops (n.d.). MLOps: Machine Learning Operations. https://ml-ops.org/

8. CONTROL DE CAMBIOS:

Fecha	Versión	Descripción del Cambio

VALIDACIÓN	NOMBRE	CARGO	FECHA
Elaboró	Germán Alberto Báquiro Duque	Contratista Profesional Especializado Gerencia de Estrategia	12/12/2024
Revisó	John Alejandro Torres Sichacá	Profesional Especializado SAIGC	12/12/2024
Aprobó	Javier Andrés Rubio Sáenz	Subgerente Análisis de la Información y Gestión del Conocimiento	13/12/2024
Aprobó	Ximena Pardo Peña	Subgerente de Planeación	13/12/2024