

Contexto:

Estamos colaborando con una empresa distribuidora de productos agrícolas que se enfrenta a un problema de exceso de inventario. Recientemente, han modificado su método de proyección de demanda y sospechan que este cambio está relacionado con la situación actual. Actualmente utilizan el modelo cada mes para determinar cuánto inventario deben comprar al proveedor. La orden de compra es enviada al proveedor con un mes de anticipación.

La empresa nos ha contactado con el objetivo de desarrollar un modelo que permita proyectar la demanda de manera más precisa. Para ello, nos han proporcionado los datos de ventas de los últimos 10 años, los cuales están a nuestra disposición. A partir del contexto entregado y de los datos de venta a disposición, responda las siguientes preguntas:

Plan de Trabajo:

¿Cuál sería tu enfoque y plan de trabajo para abordar este desafío?

Mi primer enfoque es entender la necesidad del negocio una vez entendiendo el problema , proceder a realizar el análisis exploratorio y limpieza de datos , después realizaría la transformación de variables, entrenamiento del modelo , medición y monitoreo

¿Qué tareas consideras esenciales para la ejecución de este proyecto?

Análisis exploratorio y limpieza de variables transformación de variables medición del modelo actual y comparación con modelo anterior

¿Qué elementos específicos desarrollarías para resolver el problema planteado? Incluye una estimación del tiempo que dedicarías a cada fase y explique cómo tu solución impactaría positivamente en la situación actual de la empresa.

1. Entender la necesidad del negocio (1 semana) :Tener reuniones con los stakeholders clave para definir el problema a resolver, los objetivos del proyecto y cómo el éxito será medido para alinear expectativas

2. Análisis Exploratorio y Limpieza de Datos (3 semanas)

Revisión y evaluación de la calidad de los datos,Identificación de valores faltantes, duplicados, y outliers,análisis de distribución de variables y relaciones entre ellas (correlaciones),normalización y estandarización si es necesario. Renunir las fuentes de datos y crear las ETLS asociadas para la ingesta transformación y uso dentro del modelo .

4. Entrenamiento del Modelo y Optimización (2 semanas)

Selección del modelo adecuado y prueba de diferentes algoritmos y estrategias para resolver el problema , optimización de hiperparámetros y validación cruzada,comparación con el modelo anterior utilizando métricas como RMSE, MAE para poder enfocar las mejoras del proceso..

5. Medición y Comparación (1 semana)

Evaluar el desempeño del modelo utilizando el set de test y compararlo con el modelo anterior y presentar los resultados a los stakeholders, resaltando mejoras y puntos clave.

6. Implementación y Monitoreo Continuo (2 semanas)

Implementación del modelo en producción.

Establecer un sistema de monitoreo para seguir el desempeño del modelo a lo largo del tiempo y detectar posibles degradaciones o cambios en los datos y las distribuciones de los mismos.

Disponibilizar el modelo y automatizar el proceso.

Tiempo Total: 10 semanas

Construcción del Modelo y Validación:

1. Construya un modelo que permita proyectar la demanda y explique las validaciones que aplicarías para determinar su eficacia. Además, indique la ventana de tiempo que seleccionarías para la proyección y justifica tu elección.
2. Genere un archivo CSV con las predicciones para el año 2024
3. Describa cómo funcionaría este modelo en un entorno productivo, con todo lo que esto implica.

En un entorno productivo, el modelo funcionaría mediante un proceso de inferencia por batch mensual. Esto implicaría la creación de la ETL, cada mes, los datos del mes anterior se extraen de las fuentes de datos correspondientes, ya sean bases de datos relacionales, no relacionales u otros sistemas de almacenamiento. Después de la ETL, los datos almacenarán en una tabla que será accesible para el modelo. Esta tabla contendrá las variables y características que el modelo necesita para realizar sus predicciones. El modelo recibe como entrada los datos disponibilizados para realizar las predicciones correspondientes al mes siguiente. Se puede realizar este proceso de manera automatizada. Para disponibilizar el modelo se puede hacer mediante la creación de un ambiente igual al con el que se entrenó con una máquina virtual o en físico pequeña (2 cpus, 8 gb de ram) en las que se cargará el ambiente y disponibilizará el archivo pickle o a través de una api. Después de la inferencia, las predicciones deben validarse para asegurar que el modelo esté funcionando correctamente y los resultados sean razonables. Finalmente, las predicciones son entregadas al equipo ventas, posiblemente a través de reportes automáticos o por medio de un flujo con el producto id y las ventas predichas para apoyar en la toma de decisiones para el mes siguiente.

Entregue un jupyter notebook con el desarrollo.

Medición del Beneficio en \$:

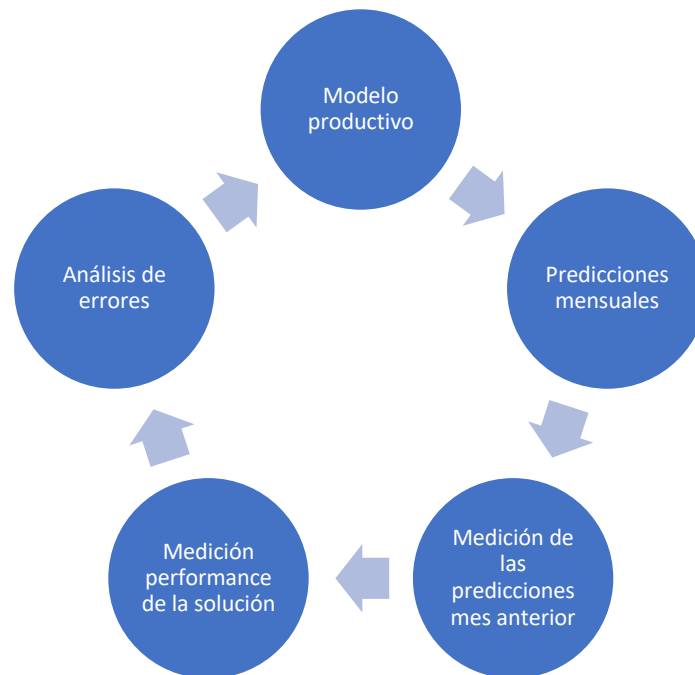
Utilizando solo la información histórica proporcionada, ¿cómo propondrías medir el beneficio económico de tu solución? Detalla los métodos o métricas que emplearías para demostrar el valor generado para el cliente.

Una de las métricas clave es el MAE mean absolute squared error que sería lo más aproximado a que tanto se está equivocando el modelo con respecto a las métricas reales y al modelo actual ,

para dimensionar las capacidades del mismo. Pero esto se puede traducir a que tanto está sobreestimado el inventario y subestimando el inventario , podríamos alinear que tanto le cuesta al negocio y dar métricas monetarias al respecto. Hacer simulaciones a largo plazo de cuánto se estaría equivocando el modelo y generar estrategias , de almacenamiento o redistribución de los productos dependiendo lo que afecte más de esta forma entre más preciso sea el modelo menos perderán.

Metodología para la Puesta en Producción:

Diseña una metodología para asegurar que el modelo implementado en un entorno productivo genera beneficios para la empresa. Explica los pasos clave que seguirías para monitorear y mantener el rendimiento del modelo a lo largo del tiempo.



Una vez que se tiene el modelo en producción se establecerán métricas benchmark o límites de acuerdo a los primeros resultados del mismo que reflejen los principales KPIS del negocio , en este caso podría ser el RMSE o el MAE , así como el performance esperado de las soluciones a nivel hardware como puede ser el tiempo de procesamiento, los recursos consumidos, el tiempo de predicción, latencia etc.. Se realizarán las primeras predicciones y se disponibilizarán dichas predicciones a los usuarios finales o a la solución final.

Para cada predicción subsecuente se monitoreará y comparará con la predicción previa por medio de pruebas A/B para determinar si cambiar dichas métricas establecidas , en caso de que hayan

cambiado se procederá a realizar la si hubo un cambio dentro de los datos es decir que los datos en producción puedan cambiar con el tiempo debido a factores como estacionalidad, cambios en el comportamiento del consumidor, o la aparición de nuevos patrones. Se realizarán pruebas con las distribuciones y patrones de los datos con los datos de entrenamiento y con los datos de predicción que se está evaluando. O si hubo un cambio en el concepto ya sea que la variables explicativas ya no expliquen la variable objetivo o que hubp un cambio de definición de la misma.

En caso de que dichos cambios sean significativos se procederá a realizar el reentrenamiento del modelo con todo lo que implica para ponerlo en producción de nuevo.