

Resumen Ejecutivo

La Promesa

Consultora Focus

Diciembre, 2022

Tabla de Contenidos

Resumen Ejecutivo.....	1
1. Problema por solucionar	3
2. Importancia del problema.....	3
3. Equipo de trabajo y roles	4
4. Hipótesis	4
5. Enfoque del Problema.....	4
6. Limpieza de datos.....	6
6.1. Análisis de datos disponibles	7
7. Modelación.....	8
8. Entrenamiento	9
8.1. Entrenamiento de 5 modelos, con outliers.....	9
8.2. Entrenamiento de 5 modelos, sin outliers	9
9. Resultados	10
10. Implementación	12
11. Conclusiones	13

Lista de Ilustraciones

Ilustración 1: Descripción de datos recibidos y pregunta generada.....	4
Ilustración 2: Posibles incentivos a repartidores.....	5
Ilustración 3: Mockup sobre incentivos a repartidores.	5
Ilustración 4: Fases de la limpieza de datos.	6
Ilustración 5: Modelo Design Thinking	7
Ilustración 6: Cantidad de pedidos vs Horas del día.	8
Ilustración 7: Modelos generados.	8
Ilustración 8: Error del modelo Elastic Net, respecto al modelo base.....	10
Ilustración 9: Error de modelos generados según horarios.....	11
Ilustración 10: Error de modelos generados según horarios según días de la semana.	11

1. Problema por solucionar

Mejora de un modelo de predicción para estimar el tiempo de entrega en domicilio de productos comprados por internet.

Objetivo General: Lograr la mejora de un modelo existente, con objeto de diferenciarse de la competencia y fidelizar al cliente final.

2. Importancia del problema

El comercio electrónico ha aumentado de forma importante en los últimos años, y fue potenciado especialmente durante la pandemia por COVID-19. Los cambios en los hábitos de consumidores desde el año 2020 generaron un explosivo aumento de compras por internet, tendencia que no se ha reducido, sino que continúa aumentando, aun cuando la pandemia ha sido controlada. Esto da cuenta de un nuevo comportamiento del consumidor en la forma de comprar, que muestra características similares tanto a nivel internacional como nacional, y que es también transversal a diversos sectores de la población.

Parte clave del comercio electrónico, tan importante como la mantención de stocks y buen diseño de ventas, es la entrega de los productos, asunto que marca una gran diferencia entre la experiencia de compra y finalmente determina la elección o preferencia entre una tienda u otra. La inmediatez que genera la experiencia de compra exige que el producto llegue a las manos del consumidor lo antes posible, lo que muchas veces no se logra en los tiempos informados, generando frustración, molestia e incluso, puede llevar a disminuir notoriamente la intención de compra para una próxima oportunidad.

Esta es la gran diferencia entre gigantes del comercio electrónico como Amazon y Mercado Libre, versus clásicas multi tiendas que no logran satisfacer las expectativas de sus clientes en la entrega y posible devolución de productos.

Es en este paso clave de la experiencia de compra en donde el rol del proceso de entrega es fundamental. No sólo la fecha de entrega prometida debe cumplirse, sino que el tiempo de entrega también puede marcar una gran diferencia, es especial cuando el consumidor ha debido comprar un producto por una emergencia, apuro u olvido, es decir, momentos claves que permiten ofrecer un servicio óptimo para la solución de un problema. Así, el cumplimiento del tiempo de entrega informado permitirá tomar decisiones y planificar, logrando una óptima satisfacción al cliente.

3. Equipo de trabajo y roles

Consultora Focus cuenta con 14 años en el mercado del asesoramiento integral en informática y negocios. El rol de la consultora va más allá de la solución de un problema específico, analizando las implicancias de la problemática planteada por el cliente, proponiendo un enfoque integral y dinámico, en completa consonancia con las necesidades actuales de la industria.

Para este caso, la consultora desarrolló el trabajo considerando 3 áreas específicas, cuyos coordinadores son:

Mauricio Gómez	Ingeniero de datos
Cristian Pérez	Ingeniero de Modelación
Constanza Gajardo	Aseguramiento de calidad
Jorge Hoffmann	Líder de equipo.

El líder del equipo es quien coordina una respuesta coordinada y se relaciona con el cliente.

4. Hipótesis

Se plantea la siguiente hipótesis: Es posible reducir el tiempo de entrega de un producto, cuya compra se realiza por internet, y prometer al cliente la entrega con un menor nivel de error que el modelo base.

5. Enfoque del Problema

Se reciben 4 archivos .csv por parte del cliente, que contienen características de: los pedidos, los clientes, las tiendas, los *picker* o recolectores de productos.

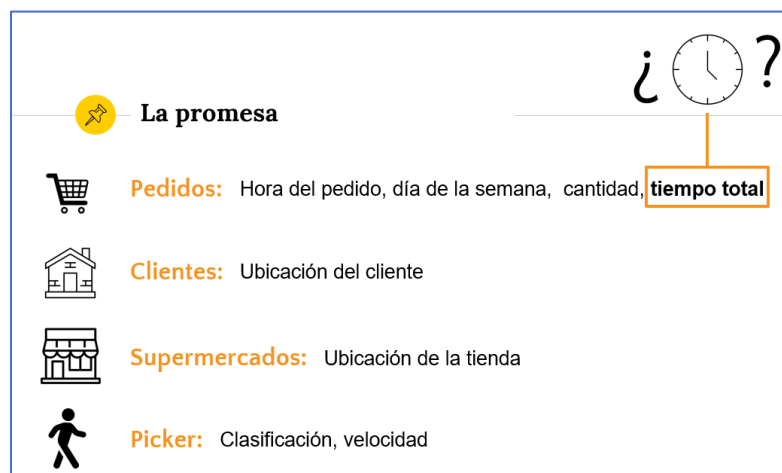


ILUSTRACIÓN 1: DESCRIPCIÓN DE DATOS RECIBIDOS Y PREGUNTA GENERADA.

Con los archivos .csv se realizarán análisis de datos, selección, limpieza y obtención del modelo base. Además, como elemento diferenciador, la propuesta considera la inclusión de un modelo que genere incentivos a los repartidores, como factor clave para el cumplimiento de la venta y la disminución del tiempo de entrega de los productos.

El análisis de este punto es en respuesta a la dinámica actual del negocio online, en donde las empresas que logran ser líderes en el mercado han logrado diferenciarse mediante un servicio de entregas óptimo.

Siguiendo la tendencia mundial en este tema, se creará un sistema de incentivo para los repartidores que, mediante un sistema de puntos, éstos pueden acceder a variados beneficios y descuentos, evitando una retribución económica directa.



ILUSTRACIÓN 2: POSIBLES INCENTIVOS A REPARTIDORES

Los incentivos logran un mayor compromiso del repartidor hacia su trabajo, logrando a la vez satisfacer las necesidades de la empresa y la inclusión de prácticas asociadas a su trabajo como cuidado en el empaque, transporte y trato hacia el cliente.

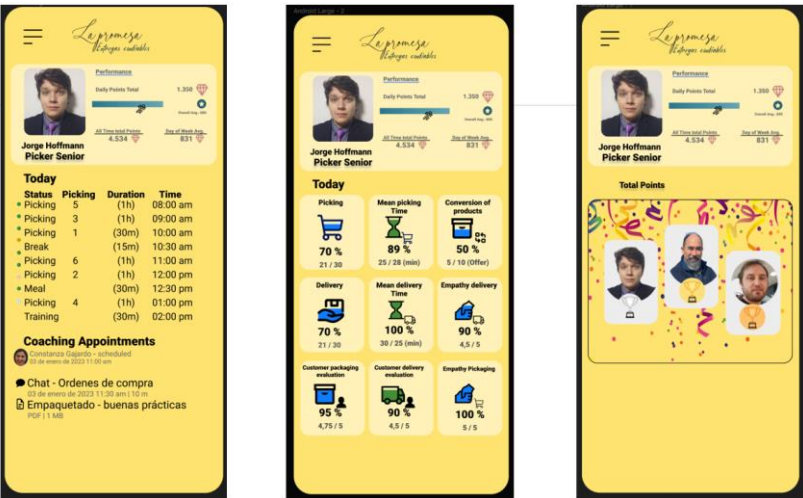


ILUSTRACIÓN 3: MOCKUP SOBRE INCENTIVOS A REPARTIDORES.

6. Limpieza de datos

Sin importar el proceso de recolección de datos puede existir cierto nivel de error. Esto se traduce en datos desordenados que significa es posible tener inconsistencias, estas podrían estar dadas por errores de medición, entrada o contexto. Ambas dadas por sistemas mal diseñados o errores humanos.

La limpieza de datos es básicamente implementar estrategias para la prevención de errores, disminuyéndolos, pero no se eliminan al 100 %. Los errores de los datos se podrán detectar en alguna de las siguientes fases:

- Recopilación o ingreso de datos.
- Transformar, extracción o transferencia de datos.
- Exploración o análisis.
- Validación cruzada.

Fases de la limpieza de datos:

- La detección implica buscar sistemáticamente características sospechosas, por ejemplo, al utilizar:
 - Cuestionarios de evaluación.
 - Bases de datos
 - Conjuntos de datos de análisis.

- El diagnóstico es la identificación de la naturaleza del defecto de los datos

- El tratamiento es borrar, editar o dejar los datos como están.

Las fases de limpieza de datos requieren una comprensión profunda de todos los tipos y fuentes de errores posibles durante los procesos de recopilación e ingreso de datos.

- La documentación de los cambios implica dejar un rastro de auditoría de los errores detectados, alteraciones, adiciones y comprobación de errores y permitirá volver al valor original si es necesario.

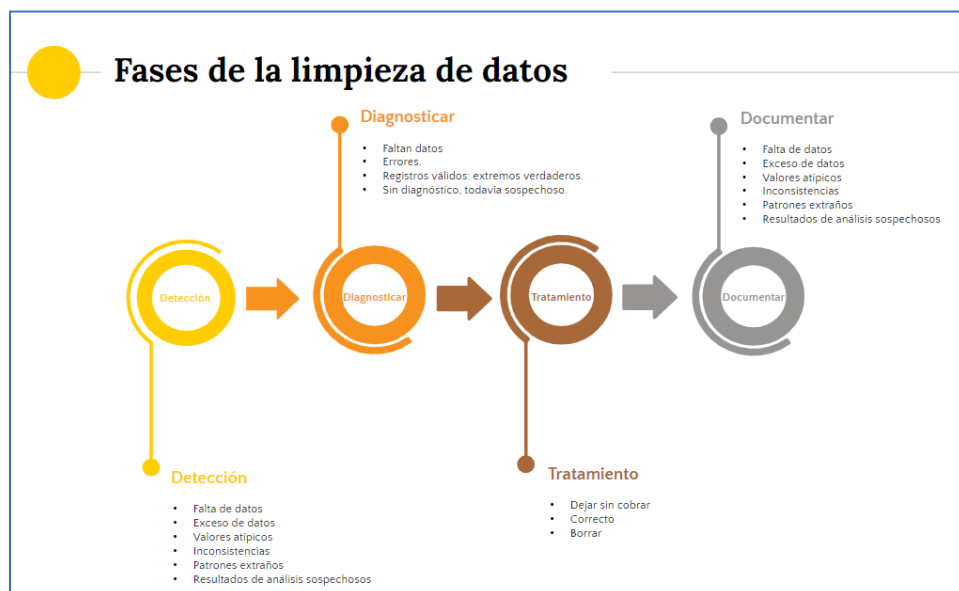


ILUSTRACIÓN 4: FASES DE LA LIMPIEZA DE DATOS.

Design thinking es un enfoque de resolución de problemas que se centra en la comprensión de las necesidades de las personas y en la creación de soluciones innovadoras y sostenibles para satisfacer esas necesidades. El proceso de Design Thinking suele incluir los siguientes pasos:

- Empatizar: se trata de conocer el problema y las necesidades de las personas que se están tratando de resolver.
- Definir: se trata de entender y comprender el problema y objetivos que persigue el cliente.
- Idear: en este paso, se generan muchas ideas diferentes para solucionar el problema.
- Prototipado: se crean prototipos de las ideas más prometedoras para probarlas y validarlas.
- Implementación: se toma una decisión sobre qué solución se implementará y se ponen en marcha los planes necesarios para hacerlo.

El Design Thinking se utiliza a menudo en el diseño de productos y servicios, pero también puede ser aplicado a cualquier tipo de problema que requiera una solución creativa y centrada en el usuario, esto debido a que, está focalizada en el ser humano, en la persona, donde, a través de la empatía, comienza un viaje hacia el conocimiento del cliente, definiendo lo que realmente está buscando, su objetivo, permitiendo generar etapas posteriores de Brainstorming, prototipado como resultado de la congregación de ideas y, finalmente, la puesta en marcha como paso culmine.

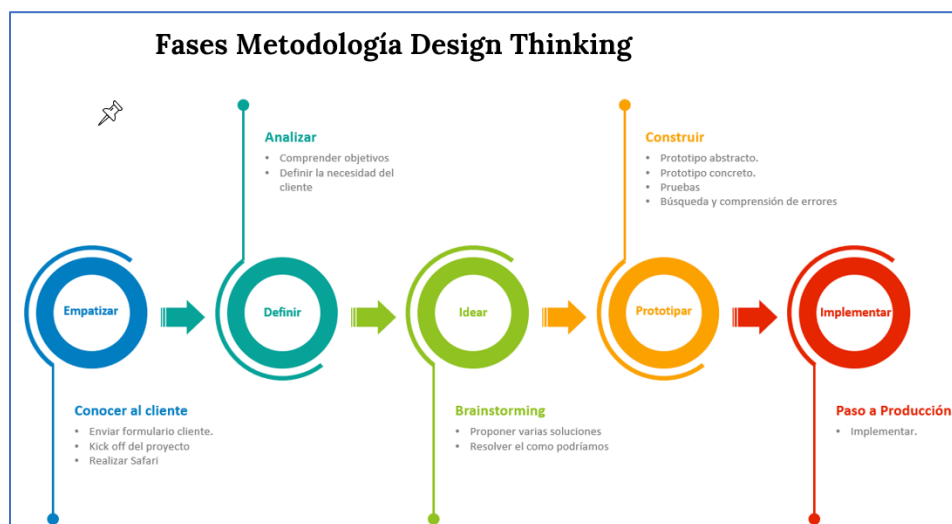


ILUSTRACIÓN 5: MODELO DESIGN THINKING

6.1. Análisis de datos disponibles

Se busca la determinación del vector objetivo y su presentación para lograr la solución al problema. Se logra establecer “Total_minutes” como vector objetivo

Luego de analizar los datos disponibles, se decidió la normalización de datos para unificar formatos en variados atributos. En otros casos, se eliminan datos perdidos.

- Se calculan distancias y analizan opciones de gráficos a utilizar

- Se estudia el vector objetivo con relación a la distancia y otros atributos de interés. Para esto se considera la elaboración de gráficos en Python utilizando variadas librerías.
- Se analiza y estima el error del modelo actual, considerando gráficos para estudiar cambios a nivel de horas del día y de días de la semana.

La inspección de datos considera el comportamiento del vector objetivo respecto a días de la semana, horas del día y frecuencias. Variados gráficos permiten visualizar la información, siendo el siguiente un adecuado resumen de ciertos atributos de interés:

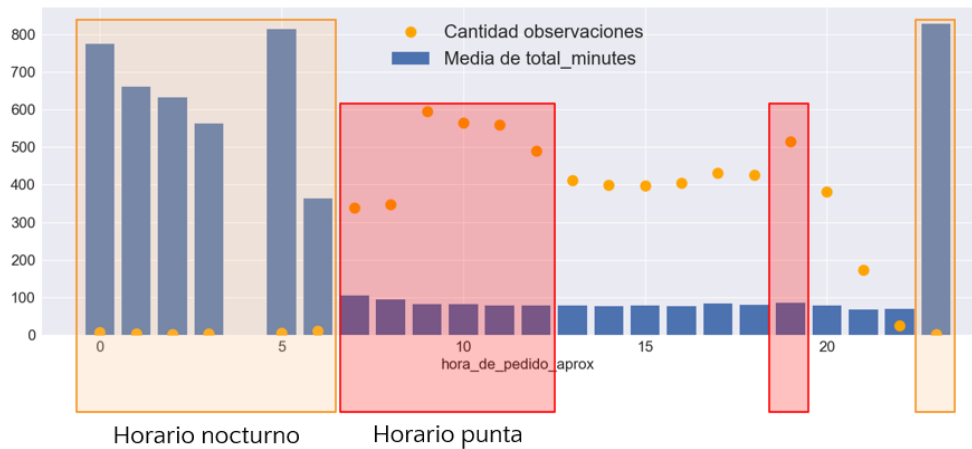


ILUSTRACIÓN 6: CANTIDAD DE PEDIDOS VS HORAS DEL DÍA.

7. Modelación

Se puede apreciar que el modelo inicial presentaba un error cuadrático medio de 1.926 minutos cuadrados y error absoluto mediano es de 35 minutos, lo que permite inferir que existen outliers que tienen mayor error, influyendo en el error cuadrático. Se trabajó en el entrenamiento de 5 modelos mostrados en la siguiente ilustración:

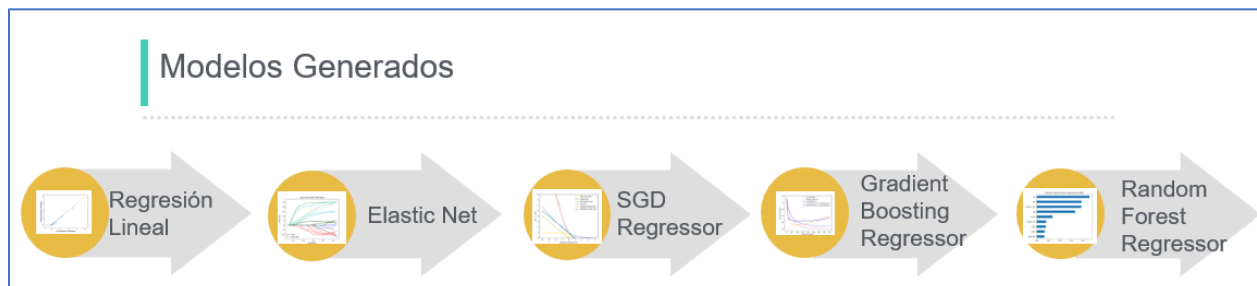


ILUSTRACIÓN 7: MODELOS GENERADOS.

Los cinco modelos fueron considerados incluyendo y excluyendo los valores extremos (outliers). De esta forma, se obtienen dos respuestas para cada modelo.

Además, se clasifican las horas de pedido de un producto y se modela considerando 3 grupos horarios: nocturno, punta y valle.

8. Entrenamiento

8.1. Entrenamiento de 5 modelos, con outliers

Inicialmente, se presentan 5 modelos distintos que consideran todos los datos disponibles. Cada modelo, es entrenado primero en una versión con todos los atributos, para luego pasar a una versión 'depurada'. De cara a este documento, el modelo depurado utiliza 30 variables.

La definición de 30 variables dependerá del modelo. Si es un modelo que cuenta con coeficientes, serán los 30 coeficientes más grandes en términos absolutos. Lo anterior habiendo aplicado un MinMaxScaler, de modo que todos los coeficientes tengan la misma magnitud. Por otro lado, hay modelos que incluyen el feat_importance. Este reemplaza a los coeficientes a la hora de escoger las 30 variables.

Construidos los 10 modelos, se procede a analizar los resultados a nivel agrupado.

- Horario nocturno: Para el error mediano absoluto el modelo campeón tiene una estimación con un error que es fácilmente 10 veces menor. De hecho, la mitad de los pedidos nocturnos tiene un error menor a 41 minutos. En cambio, el mejor modelo desafiante tiene un error de 98 minutos en su versión full y 101 minutos en su versión depurada. Con el error cuadrático la situación es exactamente igual.
- Horario punta: Mientras el modelo campeón tiene un error absoluto mediano de 37.4 minutos, los mejores modelos desafiantes tienen errores bajo los 14 minutos incluso. Eso quiere decir que la mitad de los pedidos se predice con menos de 14 minutos de error. Nuevamente, en el error cuadrático medio se presenta el mismo comportamiento.
- Horario valle: La situación es la misma que en el horario punta, existe una notable mejora respecto al modelo campeón.

En resumen, se observa que el modelo base o campeón tiene un excelente rendimiento en el horario nocturno. Sin embargo, cabe preguntarse si ese es realmente el objetivo. ¿Cuán importante es predecir bien pedidos que se hacen durante la noche? Por otra parte, la mejora de 23 minutos que se produce en el horario punta es de alto impacto.

- Para cada día de la semana el modelo campeón es mejorado por los modelos desafiantes. De forma consistente, Elastic net tiene un error absoluto mediano bastante bueno todos los días.
- En particular, para viernes y sábado Gradient boost tiene un gran rendimiento, llegando incluso a 11.3 minutos de error absoluto mediano.

8.2. Entrenamiento de 5 modelos, sin outliers

Se realizan los mismos entrenamientos, de los mismos cinco modelos, eliminando valores extremos: 2.5% de los valores menores y 2.5% de los valores mayores, quedando una población de datos correspondiente al 95% del total. Se aprecia que los modelos presentan errores menores, sobre todo en error cuadrático medio (que es aquel que castiga los errores extremos).

En estas circunstancias, el modelo GradientBoostingRegressor tiene mejor comportamiento que ElasticNet, con un error mediano absoluto menor a 13 minutos.

Analizando los resultados por horario, se puede apreciar que el modelo base, o “campeón”, es superado por los nuevos modelos propuestos en los tres horarios: Nocturno, Punta y Valle.

Es destacable el rendimiento del modelo RandomForest en horario nocturno, que logra un error de sólo 7 minutos. En horario Punta y Valle, Gradient boosting Regressor y Elastic net presentan muy buenos resultados, con errores absolutos medianos entre 12 y 13 minutos, respectivamente.

9. Resultados

Respecto al error cuadrático medio, existe bastante consistencia entre todos los modelos, con errores entre 700 y 800 minutos cuadrados, lo que es bastante mejor que los 2000 minutos cuadrados del modelo base.

El modelo “ElasticNet” prueba ser el más adecuado a nuestro objetivo, considerando que sus errores son menores al modelo base (Ilustración 8), similares a los otros modelos tanto en lo relativo a horarios y días de la semana (Ilustración 9 e Ilustración 10), pero su carga es más fácil.

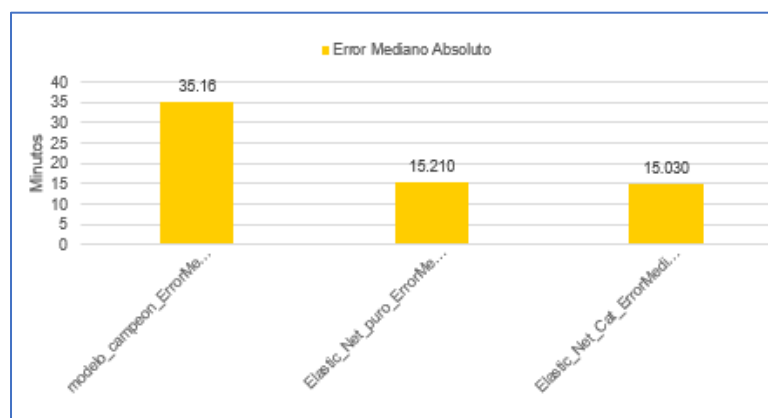


ILUSTRACIÓN 8: ERROR DEL MODELO ELASTIC NET, RESPECTO AL MODELO BASE.

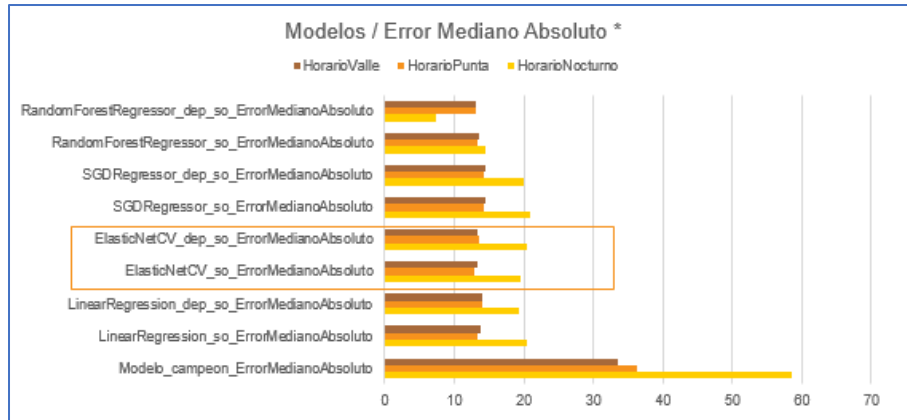


ILUSTRACIÓN 9: ERROR DE MODELOS GENERADOS SEGÚN HORARIOS.

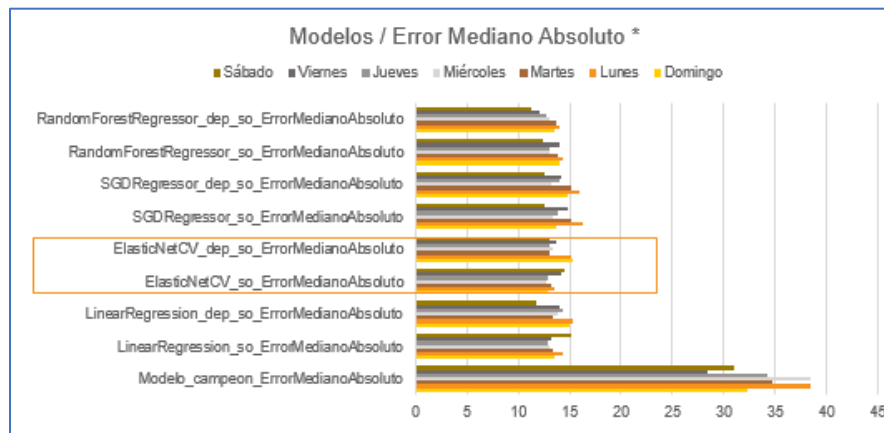


ILUSTRACIÓN 10: ERROR DE MODELOS GENERADOS SEGÚN HORARIOS SEGÚN DÍAS DE LA SEMANA.

Con miras a perfeccionar la propuesta, a decisión del cliente, se propone la profundización de modelos al considerar datos de forma segmentada, así como la posible adición de datos para lograr mayor especificidad.

10.Implementación

No sólo propone un modelo para mejorar el tiempo de entrega, sino que destaca la importancia de los recolectores o “pickers” y los repartidores -quienes finalmente entregan el pedido- pues este factor humano será clave en el éxito del modelo.

Para lo anterior, se propone un sistema de incentivos que logrará que el *picker* se comprometa a entregar un servicio de calidad, pensando en las preferencias del cliente. El sistema se implementa en los teléfonos móviles donde el picker además de ver el pedido en curso, podrá revisar sus estadísticas como las siguientes:

- Total, de ordenes completadas.
- Tiempo medio para completar las ordenes versus el tiempo estimado.
- Conversión de productos sin existencia a productos “insignes”.
- Evaluación de del cliente respecto a:
 - Empaquetado de productos, para evitar empaquetar juntos por ejemplo harina con helados o huevos con tarros conserveros.
 - “sonrisa telefónica” cuando se contacte con el cliente para avisar la no existencia del producto solicitado o el termino de armar el pedido para ver si el cliente requiere un producto adicional en este.

Para el caso del delivery, se mediran estadisticas como las siguientes:

- Total, de ordenes entregadas.
- Tiempo medio para entregar las ordenes versus el tiempo estimado.
- Evaluación de del cliente respecto a:
 - Manipulación del pedido (si golpea o no los productos, si estos son recibidos con algún daño ocasionado en el trayecto, entre otros).
 - Forma de interactuar con el cliente al momento de realizar la entrega del pedido.

Tambien el colaborador tendra a su disposición sesiones de coaching, documentos de buenas practicas, cursos de e-learning para mejorar el trato con los clientes, etcetera.

Todo lo anteriormente señalado será calificado con un sistema de puntos, que permitiran al picker y/o delivery acumular puntos para canjearlos por diversos productos, como por ejemplo, entradas para el cine, accesorios para el vehículo, giftcard para bencina o compras en tiendas suscritas a nuestra aplicación, cenas y mucho mas.

11.Conclusiones

Después de probar cinco modelos diferentes con diferentes horarios de entrega y días de la semana, se encontró que el modelo "ElasticNet" es el más adecuado para predecir de manera más precisa y eficiente el tiempo total de entrega de un pedido ("total_minutes"). Sin embargo, tenemos en cuenta que la entrega también puede depender de las características humanas del picker y el repartidor. Por lo tanto, se sugiere implementar un sistema de incentivos para motivar a estos empleados y aumentar su rendimiento, con el fin de brindar una atención óptima al cliente y fomentar la lealtad a la empresa.