



Tecnológico de Monterrey

CAMPUS MONTERREY

**INTELIGENCIA ARTIFICIAL AVANZADA PARA LA
CIENCIA DE DATOS II**

TC3007C

Momento de Retroalimentación: Reto, Entendimiento del Problema

Prof. Edgar González Fernández

Daniela Jiménez Téllez - A01654798

Lautaro Gabriel Coteja - A01571214

Andrés Villarreal González - A00833915

Héctor Hibran Tapia Fernández - A01661114

Índice

Portada.....	1
Índice.....	2
ARCA CONTINENTAL.....	3
Introducción.....	3
Objetivos.....	3
Antecedentes y Proyectos Relacionados.....	3
Herramientas y Recursos a Usar.....	4
Metodología.....	4
Resultados.....	4
Conclusiones y Trabajo a Futuro.....	5
Referencias.....	5

ARCA CONTINENTAL

Introducción

El problema a resolver, es que Arca Continental busca optimizar la ubicación de productos de lanzamiento en la red de tiendas de Arca Continental, tomando en cuenta proximidad, infraestructuras, y niveles socioeconómicos de la zona, para esto se necesita desarrollar una IA que defina si un producto que será lanzado al mercado será exitoso o no para los clientes que Arca Continental posee.

Objetivos

1. Desarrollar un modelo de análisis
2. Que el modelo permita a la empresa ubicar nuevos productos en las mejores posiciones dentro de su red de tiendas
3. Optimizar las oportunidades de ventas y satisfacción del cliente
4. Analizar características socioeconómicas, y de infraestructura de las zonas de las tiendas para identificar el perfil de cliente más adecuado para un producto

Antecedentes y Proyectos Relacionados

La Inteligencia Artificial (IA) está transformando variedad de industrias como, la salud, finanzas, transporte, y comercio, al permitir decisiones y procesos basados en datos complejos. Su impacto abarca desde la personalización en la atención al cliente hasta la automatización en áreas críticas, como diagnósticos médicos y detección de fraudes. Sin embargo, la adopción de la IA plantea ciertas preocupaciones éticas y normativas:

- Privacidad de los Datos
La recolección masiva de datos personales genera preocupaciones sobre la privacidad, ya que un mal manejo, o fuga, puede poner en peligro los datos personales de los usuarios.
- Transparencia y Sesgo Algorítmico
Los algoritmos de IA, específicamente los que operan como "Cajas Negras", pueden llegar a tener dificultades con la transparencia. No cuidar temas como el sesgo en el manejo de datos, puede llevar a resultados discriminatorios y/o poco precisos, afectando a grupos específicos de personas.
- Responsabilidad y Toma de Decisiones Autónoma
En sectores críticos, donde las decisiones automatizadas pueden tener consecuencias esenciales, se discute sobre la responsabilidad en caso de errores de la IA.
- Impacto Laboral
La automatización a través de la IA reemplaza ciertos tipos de empleos, en especial aquellos que se basan en tareas repetitivas, lo que plantea la necesidad de una transición laboral.

En cuanto al desarrollo de este proyecto / reto, por lo que se busca hacer, hay ciertas consideraciones que hay que tener en cuenta, tales como:

- La Privacidad de los Datos
Al trabajar con datos de ubicación, ingresos, y hábitos de compra, entre otros, es

crucial implementar medidas de seguridad y anonimización para proteger la identidad y privacidad de los clientes, siguiendo regulaciones y normativas locales y/o internacionales.

- **Prevención de Sesgos**
La red neuronal debe entrenarse con datos representativos y equilibrados para evitar cualquier tipo de sesgo que pueda resultar en recomendaciones injustas o erróneas, además, esto ayuda a que la predicción del éxito de un producto de lanzamiento sea lo más imparcial posible.
- **Responsabilidad en la Implementación**
Como los resultados de la red neuronal influyen en variedad de factores, sería prudente considerar que el modelo no sea el único criterio en la toma de decisiones, sino que sea complementado o complementado por evaluaciones humanas, esto para minimizar riesgos.

Herramientas y Recursos a Usar

Las herramientas que se estarán usando para el análisis y desarrollo del modelo, son las siguientes:

- **Python**
Lenguaje base utilizado para el desarrollo del modelo y análisis de datos.
- **Pandas / Numpy**
Herramientas para la manipulación y transformación eficiente de datos.
- **Matplotlib / Seaborn**
Librerías de visualización que facilitan la creación de gráficos y el análisis exploratorio.
- **ScyPy / Scikit - Learn**
Proveen funcionalidades avanzadas para el procesamiento estadístico y modelado.
- **TensorFlow**
Framework para construir y entrenar redes neuronales.
- **Google Colab / Jupyter Notebooks**
Entornos colaborativos en la nube para ejecutar y documentar el código de manera interactiva.

Metodología

Poner algo aquí

Resultados

Después de la exploración de los datos, el análisis de estos, y el procesamiento que se les hizo, a continuación hay una imagen del dataframe final, en el cual se les aplicó merge a los 3 sets de datos entregados por Arca Continental, se dejaron las columnas y datos más importantes para que sea el entendimiento y observación de patrones sea más sencilla para la IA.

	Columna	Tipo de Objeto			
0	CustomerId	int64	24	hospitales	float64
1	Material	int64	25	preescolares	float64
2	successful	int64	26	primarias	float64
3	Material_desc_x	object	27	secundarias	float64
4	pc_agr_300m	float64	28	preparatorias	float64
5	pc_comercial_300m	float64	29	universidades	float64
6	pc_generales_300m	float64	30	gimnasios	float64
7	pc_habitacional_300m	float64	31	gasto_promedio_300m	float64
8	pc_habitacional_mixta_300m	float64	32	gasto_total_300m	float64
9	pc_industrial_300m	float64	33	ingreso_minimo_300m	float64
10	pc_minero_300m	float64	34	ingreso_promedio_300m	float64
11	pc_mixta_300m	float64	35	ingreso_maximo_300m	float64
12	pc_negocios_300m	float64	36	ingreso_remesas_300m	float64
13	pc_turismo_300m	float64	37	ingreso_rentas_300m	float64
14	Peso_manza	float64	38	accesibilidad	float64
15	pob_ab_300m	float64	39	industry_customer_size	int64
16	pob_cmas_300m	float64	40	sub_canal_comercial	object
17	pob_c_300m	float64	41	Productos_Por_Empaque	int64
18	pob_cmen_300m	float64	42	ProductType	object
19	pob_dmas_300m	float64	43	calmonth_x	datetime64[ns]
20	pob_d_300m	float64	44	conteo_instalaciones	float64
21	pob_e_300m	float64	45	categoria_instalaciones	object
22	parques	float64			
23	supermercados	float64			

En las siguientes dos imágenes podemos observar el resultado de los modelos desarrollados hasta ahora, el cual, el que está a la izquierda pertenece a un modelo en base a regresión logística, y el del lado derecho pertenece a un modelo basado en random forest.

Precisión del modelo: 0.79					
Reporte de clasificación:					
	precision	recall	f1-score	support	
0	0.81	0.93	0.87	4141	
1	0.67	0.38	0.49	1485	
accuracy			0.79	5626	
macro avg	0.74	0.66	0.68	5626	
weighted avg	0.77	0.79	0.77	5626	
Matriz de confusión:					
[[3858 283]					
[915 570]]					
Porcentaje de predicciones correctas: 72.44%					

Precisión del modelo: 0.76					
Reporte de clasificación:					
	precision	recall	f1-score	support	
0	0.81	0.87	0.84	4141	
1	0.56	0.45	0.50	1485	
accuracy			0.76	5626	
macro avg	0.69	0.66	0.67	5626	
weighted avg	0.75	0.76	0.75	5626	
Matriz de confusión:					
[[3615 526]					
[823 662]]					
Porcentaje de predicciones correctas: 68.34%					

Conclusiones y Trabajo a Futuro

Los modelos desarrollados hasta el momento, que incluyen regresión logística, y random forest, han permitido analizar factores clave que podrían influir en el éxito de los productos de lanzamiento en las tiendas de Arca Continental. El procesamiento de los datos ha sido fundamental para depurar y consolidar la información, permitiendo que los modelos identifiquen patrones relevantes en la relación entre características de los productos, clientes, y ubicación. Sin embargo, los resultados obtenidos muestran un potencial de mejora en la precisión de las predicciones, lo cual abre oportunidades para optimizar el modelo y mejorar la toma de decisiones.

Para los siguientes pasos del equipo, y la mejora del proyecto, se realizará un análisis más detallado sobre los 27 clientes que no adquieren productos de lanzamiento, con el fin de identificar las razones detrás de este comportamiento y así darle mejor información al modelo y obtener un porcentaje de predicción más alto. De igual forma, se planifica la mejora de los modelos actuales mediante un procesamiento de datos más

específico, que pueda aumentar la precisión de las predicciones. Finalmente, se sugiere implementar nuevos modelos como Redes Neuronales o SVM, que potencialmente ofrecen mejores resultados al capturar patrones más complejos en los datos. Estos siguientes pasos serán fundamentales para refinar la IA y optimizar la recomendación de productos en la red de tiendas de Arca Continental.

Referencias

Poner algo aquí