## **Informe Final: Búsqueda Mejor Modelo para pronóstico de venta, Caso Multinacional Argentina**

**1. Introducción e Hipótesis**

El objetivo de este proyecto fue desarrollar un modelo para predecir las toneladas (tn) de venta de 780 productos para el mes de febrero de 2020, teniendo la información de ventas históricas desde enero 2017 hasta diciembre 2019, utilizando un marco de competencia en Kaggle para la evaluación del rendimiento de los modelos y experimentos. El desafío principal radicó en la Variedad de las series de tiempo, dado que cada producto tenía un comportamiento de venta totalmente diferente a los otros, lo cual es una característica de un entorno económico volátil como el argentino.

Nuestra hipótesis de partida, validada a lo largo de la experimentación, fue que ***ningún modelo tanto clásico como de Machine Learning sería óptimo para todos los productos, dado que cada serie de tiempo tiene un comportamiento único y se debe encontrar el mejor modelo que se ajuste a las características de cada serie****.*

Como segunda instancia en este trabajo experimental queremos validar o eliminar el siguiente sesgo cognitivo: ***Los Modelos modernos de Machine Learning tienen mejor desempeño que los modelos tradicionales de pronóstico.***

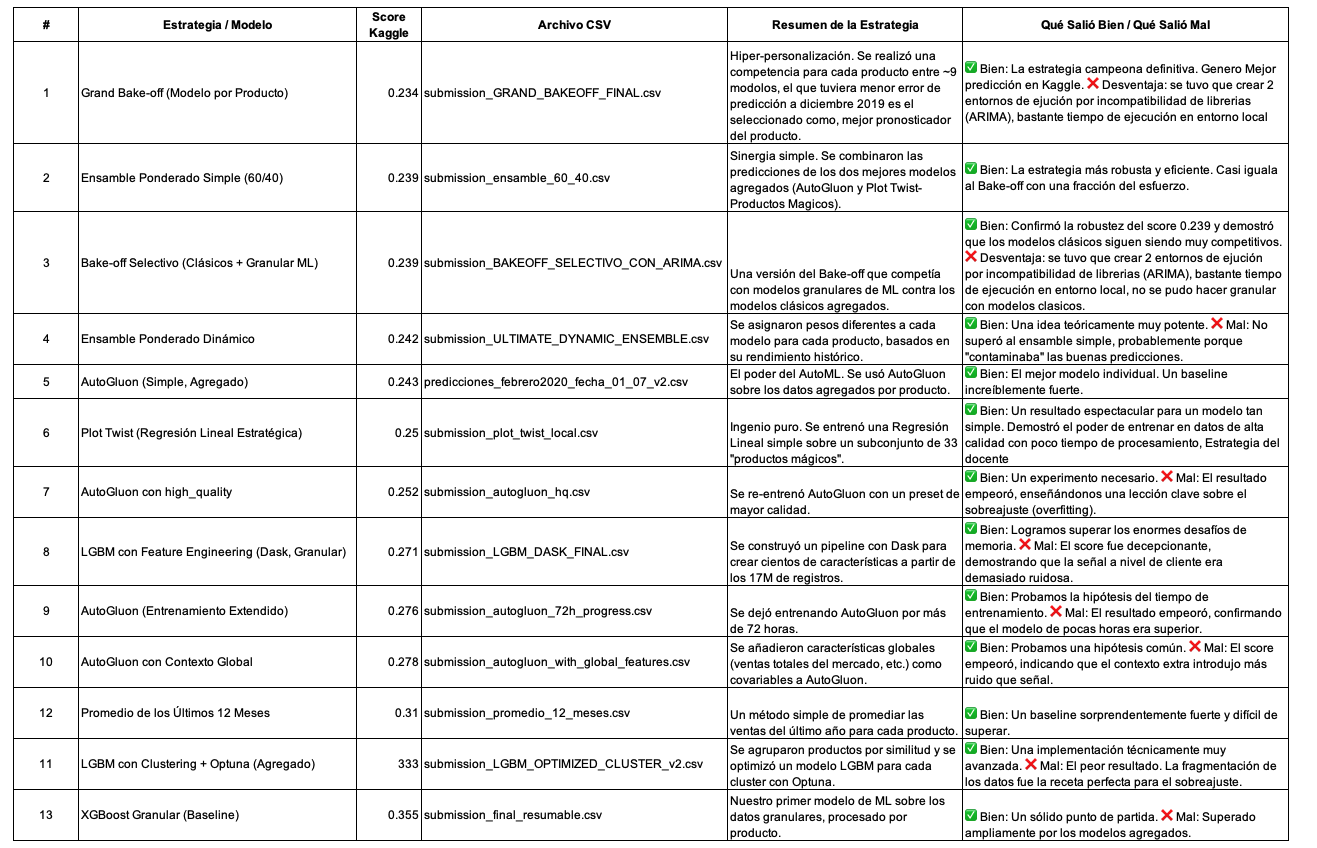
**2. Diseño Experimental y Resumen de Pruebas**

El diseño experimental siguió una lógica progresiva, partiendo de modelos simples y avanzando hacia arquitecturas complejas y ensambles. A continuación, se detallan el pipeline general y los resultados de los experimentos clave.

#### **2.1 Pipeline General**

El proceso se estructuró en las siguientes etapas:

1. **Preprocesamiento y Análisis Exploratorio (EDA)**: Unión de fuentes de datos, limpieza y estudio de tendencias para guiar la creación de variables.
2. **Ingeniería de Características (Feature Engineering)**: Creación de un set masivo de variables temporales (lags, ventanas móviles, etc) y categóricas para alimentar los modelos de Machine Learning.
3. **Modelado y Evaluación**: Entrenamiento iterativo de múltiples familias de modelos (Clásicos, Machine Learning, AutoML, Redes Neuronales) y evaluación constante de su rendimiento en Kaggle.
4. **Ensamble y Selección Final**: Combinación de las mejores predicciones y selección de la estrategia ganadora basada en la evidencia empírica y del resultado en la competencia en el *public de Kaggle.*



#### **3. Estrategia o Modelo Elegido: "Grand Bake-off – Modelo de predicción por producto"**

Resultado en Public: **0.234**.

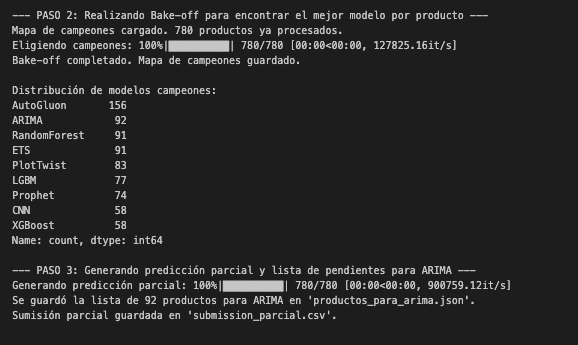
#### **3.1. ¿En qué Consiste?**

Esta estrategia es un enfoque de **hiper-personalización**. En lugar de buscar un único modelo que funcione para todos los productos, se realiza una competencia para cada producto de forma individual tenga el mejor modelo predictivo en función del error de pronóstico a diciembre 2019.

1. **Competición Histórica**: Se utilizaron las predicciones de un arsenal de 9 modelos (LGBM, ARIMA, CNN, XGBoost, Prophet, Holt o ETS, AutoGluon y PlotTwist-Productos Magicos) para un mes conocido (diciembre de 2019). Se compararon estas predicciones con las ventas reales de esa fecha para determinar qué modelo fue el "campeón" para cada uno de los 780 productos.
2. **Predicción con el modelo ganador**: Para la predicción final de febrero de 2020, el sistema re-entrenó para cada producto únicamente a su modelo ganador, utilizando todos los datos disponibles.

#### **3.2. ¿Por qué se Eligió?**

Se eligió porque, experimentalmente, fue la estrategia que produjo el **menor error en la competencia en el public leader**. Su éxito valida la hipótesis inicial del proyecto: la variabilidad de los datos demandaba un enfoque diferenciado. Al aplicar la herramienta correcta para cada caso (un ARIMA para un producto con estacionalidad clara, un modelo de ML para otro más volátil), se logró un mejor resultado que modelos individuales, a continuación la distribución de los modelos ganadores.



#### **4. Conclusiones y Lecciones Aprendidas**

4.1 Conclusiones

La primera conclusión de este trabajo experimental es que un enfoque de modelado diferenciado por producto dio mejor resultado que la utilización de un solo modelo (tanto de machine learning como clásico de predicción) para los 780 productos de la competencia, dado que cada producto tiene una serie temporal con características propias, por lo tanto el desafío fue encontrar que modelo se ajustaba mejor a esas series temporales.

La segunda conclusión elimina el sesgo congnitivo de que los modelos modernos de machine learning tienen mejor capacidad predictiva que los modelos clásicos, lo cual fue desmitificado con experimentos claves como la regresión lineal de los productos mágicos, el promedio de los últimos 12 meses y el ARIMA; si observamos detalladamente la tabla de los modelos ganadores de la estrategia Gran-Bake-Off observamos que 340 Productos de los 780 usaron modelos clásicos.

4.2 Lecciones Claves del Proceso

* **AutoGluon (Simple, Agregado)**: Este modelo se consolidó como el **mejor modelo individual** (score 0.243). Su extraordinario rendimiento se debió a la **simplificación radical del problema** (agregando los datos por producto para eliminar el ruido de los clientes) y a la **fuerza bruta de AutoML**, que probó y ensambló automáticamente modelos de Deep Learning complejos sin necesidad de ingeniería de características manual.
* **Plot Twist (Regresión Lineal Estratégica)**: Con un impresionante score de **0.250**, el éxito de esta estrategia provino de un enfoque quirúrgico. Al entrenar una Regresión Lineal simple sobre un subconjunto curado de 33 "productos mágicos" de alta calidad, el modelo aprendió una relación "pura" y fundamental que se generalizó exitosamente al resto de los productos, demostrando el poder de la **simplicidad para evitar el sobreajuste**.
* **La Sinergia Simple es Robusta**: Nuestro segundo mejor modelo, el **Ensamble Ponderado Simple (score 0.239)**, fue un benchmark increíblemente fuerte, superando a estrategias mucho más complejas y demostrando que la combinación de enfoques diferentes (AutoML + Regresión simple) es una técnica muy poderosa.
* **La Complejidad conllevó al Sobreajuste**: Los intentos de construir un único "super-modelo" a través de ingeniería de características masiva o clustering resultaron en un peor rendimiento, una lección clásica sobre el sobreajuste, esto también pudo generarse por errores en la programación de los modelos al no parametrizar de una manera correcta los modelos.
* **La Importancia de la Experimentación Rigurosa**: Este proyecto validó que un proceso iterativo y metódico es fundamental para encontrar la solución óptima, superando desafíos técnicos como la **gestión de memoria a gran escala** (usando Dask y Parquet) y la **resolución de conflictos de dependencias** (creando entornos de Conda aislados para ARIMA), adicional a lo anterior la experimentación es la única manera de construir conocimiento sólido y de eliminar sesgos cognitivos que limitan la capacidad de pensamiento crítico.

1. **Anexo Explicación del script Ganador.**