



Análisis operativo: Predicción de ingresos por ventas de un Restaurante

Curso: MA4402-1 – Simulación estocástica: Teoría y laboratorio.

Profesor: Joaquín Fontbona.

Ayudantes: Álvaro Márquez, Catalina Lizana, Matías Ortiz.

Ayudante encargado: Álvaro Márquez.

Presentadores: Nicolás Garay, Tomás Vesperinas.





Objetivo

¿Dónde abrir un restaurante?

1. Necesitamos predecir el éxito de las ventas

2. ¿Qué ocurre si los datos tienen un margen de error?



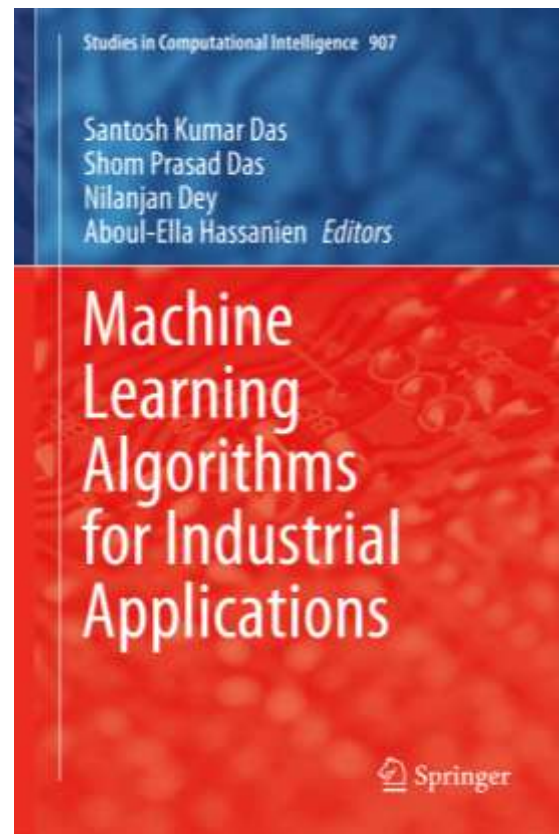
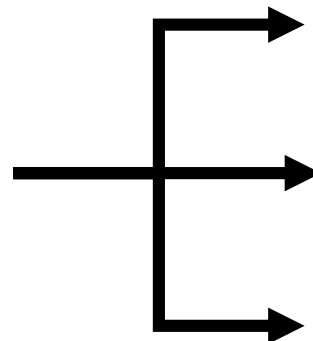


Objetivo

¿Dónde abrir un restaurante?

1. Necesitamos predecir el éxito de las ventas

2. ¿Qué ocurre si los datos tienen un margen de error?



**An Application of Operational Analytics:
For Predicting Sales Revenue
of Restaurant**

Samiran Bera

Abstract Operational analytics improves existing operations of a firm by focusing on process improvement. It acts a business tool for resource management, data streamlining which improves productivity, employee engagement, customer satisfaction and provide investment opportunities. Crucial insights into the problem can be obtained which aids to determine key business strategy through various stages of data analysis and modeling, such as exploratory data analysis, predictive modeling,



Identificando los datos

- Training & Test set
- Campos de la base de datos:

☐ Id

☐ Open date

☐ City

☐ City group

☐ Type

☐ P1, P2, ..., P37

☐ Revenue

kaggle™

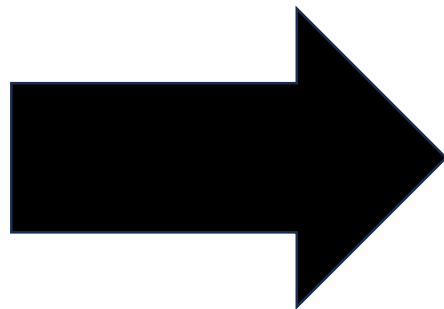




Limpieza de los datos

- **Campos de la base de datos:**

- Id
- Open date
- City
- City group
- Type
- P1, P2, ..., P37
- Revenue



Pipeline (Scikit)

- No habían datos faltantes.
- Se quita Id y Open date.
- Se normalizaron los datos.
- OneHotEncoding para campos de categorías:
City, City Group y Type.



Modelos

- Sea \mathbf{X} la matriz de datos, cuyas columnas son campos, y sus filas son instancias.
- Sea \mathbf{y} el valor objetivo (ventas), e $\hat{\mathbf{y}}$ la predicción del modelo.

Linear Regression

- **Predicción:** $\hat{\mathbf{y}} = \mathbf{X}\mathbf{w} + \mathbf{b}$, donde \mathbf{w} y \mathbf{b} son los pesos y la intersección.

- Se buscan \mathbf{w} y \mathbf{b} que minimicen el MSE: $\|\hat{\mathbf{y}} - \mathbf{y}\|_2^2$.

Ridge Regression

- Ahora se busca minimizar: $\|\hat{\mathbf{y}} - \mathbf{y}\|_2^2 + \alpha \|\mathbf{w}\|_2^2$.

Decision Tree (regression)

- **Predicción:** Particiona el espacio de a pares. Un nodo de data $\mathbf{Q}_m(\boldsymbol{\theta})$ depende de $\boldsymbol{\theta} = (j, t_m)$ una característica j y un umbral t_m .
- Para regresión, en cada nodo se minimiza:

$$G_m(\boldsymbol{\theta}) = \frac{n_m^{\text{left}} \text{MSE}(\mathbf{Q}_m^{\text{left}}) + n_m^{\text{right}} \text{MSE}(\mathbf{Q}_m^{\text{right}})}{n_m}$$



Weighted Models

- Se consideran los tres modelos base implementados (linear regression, ridge regression y decision tree)
- Se entrena cada modelo base
- **Predicción:** Se realiza predicción asignándole una ponderación a las predicciones de los modelos base

Random Forest (regression)

- Se eligen muestras de forma aleatoria mediante Bootstrap (con repetición) del conjunto de datos original.
- Para cada muestra se selecciona de forma aleatoria un subconjunto de variables de la muestra.
- Para cada muestra se realiza una predicción mediante Decision Tree.
- **Predicción:** Promedio de las predicciones de cada Decision Tree



Second Stage Regression

- Se realiza predicción con modelos base (linear regression, ridge regression y decision tree)
- Se realiza regresión lineal a las predicciones de los modelos base.
- **Predicción:** Se realiza un Weighted Model tomando como pesos los coeficientes de la regresión



Comparación Modelos

Coeficiente de determinación R^2 :

Dado un modelo se define el coeficiente de determinación R^2 como:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Donde $(y_i)_{i=1}^n$ son los valores observados, $(\hat{y}_i)_{i=1}^n$ son los valores predichos por el modelo e \bar{y} es la media de los valores observados



Resultados obtenidos:

	R^2 libro (%)	R^2 obtenido (%)
Linear Regression	55.02	51.01
Ridge Regression	51.33	49.84
Decision Tree	83.79	83.86
Random Forest	72.55	75.85
Second Stage Model	86.73	83.97



Coeficientes de determinación para distintas combinaciones de pesos:

Setting	1	2	3	4	5	6	7	8	9	10	11	12
W1	1	1	1	1	1	1	1	1	1	2	3	2
W2	1	1	1	1	1	1	1	1	1	1	1	2
W3	1	2	3	4	5	6	7	8	11	10	10	10
R^2 libro (%)	74.35	80.60	83.10	84.25	84.82	85.11	85.25	85.31	85.29	85.34	85.04	84.82
R^2 obtenido (%)	69.96	76.50	79.45	81.00	81.91	82.48	82.86	83.13	83.56	82.78	81.95	81.91



¿Que pasa si añadimos ruido a los datos?

	R^2 con ruido $N(0,0.1)$ (%)	R^2 sin ruido (%)
Linear Regression	47.73	51.01
Ridge Regression	46.97	49.84
Decision Tree	84.72	83.86
Random Forest	83.81	75.85
Second Stage Model	85.88	83.97



Setting	1	2	3	4	5	6	7	8	9	10	11	12
W1	1	1	1	1	1	1	1	1	1	2	3	2
W2	1	1	1	1	1	1	1	1	1	1	1	2
W3	1	2	3	4	5	6	7	8	11	10	10	10
R^2 con ruido N(0,0.1) (%)	71.94	79.64	82.81	84.33	85.12	85.56	85.80	85.93	86.03	85.77	85.20	85.12
R^2 sin ruido (%)	69.96	76.50	79.45	81.00	81.91	82.48	82.86	83.13	83.56	82.78	81.95	81.91



Conclusiones

- Se estudiaron distintos modelos de regresión en el contexto de predecir las ventas de un restaurante. Se compararon entre sí mediante el coeficiente de determinación R^2 . Obteniéndose resultados similares a los de [2].
- La regression mediante second stage model tiene el mayor coeficiente de determinación. Siendo el más preciso para la predicción de ingresos.
- Al añadirle ruido a los datos no se obtienen grandes diferencias en el coeficiente de determinación.



Referencias

- Géron, A. (2019). Hands-on machine learning with SciKit-Learn, KeRas, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly. [1]
- Das, S. K., Das, S. P., Dey, N., & Hassanien, A. (2020). Machine learning Algorithms for industrial applications. Springer. [2]
- Ekrem Ozer, Meghan O'Connell, Wendy Kan. (2015). Restaurant Revenue Prediction. Kaggle.
<https://kaggle.com/competitions/restaurant-revenue-prediction> [3]



Ingeniería Matemática
FACULTAD DE CIENCIAS
FÍSICAS Y MATEMÁTICAS
UNIVERSIDAD DE CHILE

Gracias!