

Capítulo 2

Marco Teórico

2.1. Economía Colombiana

2.1.1. Introducción

La sección de economía colombiana tiene como finalidad situar el marco de referencia actual y la problemática económica de la fluctuación constante de la tasa de cambios en un país de Latino América [Plazas, 2013]. Aquí es donde introducimos los conceptos de tasa de cambio, TRM, revisamos cuales son los principales productos de exportación del país, y damos un breve repaso a los elementos de reducción de riesgo financiero como los futuros o *forwards*.

2.1.2. La Tasa de Cambio

La moneda de un país tiene una equivalencia en moneda de otro y ese valor se conoce como tasa de cambio. Explicamos el concepto apoyados en los escritos del autor Mauricio Cárdenas [Cárdenas, 2016]. También es importante explicar porqué los productos de exportación tienen un efecto en la canasta de divisas y la balanza de pagos [Carriello, 2010].

2.1.3. La TRM

La TRM (o tasa de mercado representativa) es el nombre oficial de la tasa de cambio del dolar estadounidense en pesos colombianos. A la vez es nuestro objeto de estudio, ya que deseamos predecir su valor modelando el mismo a través de los valores conocidos de materias primas de exportación.

2.1.4. Exportaciones de Colombia

Si bien no buscamos ser expertos en ninguno de los tipos de exportación que hace Colombia, es importante en esta sección describir uno a uno los rubros con mayor contribución, ya que serán nuestras variables independientes para aplicar en el proceso de aprendizaje automatizado y modelar el comportamiento futuro de la TRM.

El Petroleo

El petroleo es el punto de partida de este trabajo de investigación, el cual se basa en el concepto más o menos generalizado de que las exportaciones de petroleo ejercen gran presión sobre la TRM, y que esta última está relacionada con el precio del barril de petroleo Brent y West Texas [Roca et al., 2002].

El Carbón

El carbón es otro de los elementos con mayor importancia en la lista de exportación de Colombia [Cárdenas, 2016].

El Café

El café Colombiano es reconocido a nivel mundial a través de su marca registrada Juan Valdez. Dado que es una de las exportaciones que continua creciendo, es de esperar que sea una fuente de divisas y exista una correlación estrecha entre el precio del café y el valor de la TRM.

2.1.5. Forwards

El siguiente trabajo de investigación no trata sobre opciones de compra de moneda a futuro (conocido como *forwards*). Sin embargo explicamos de forma sucinta qué son y cómo funcionan, ya que el resultado de las predicciones se utilizará muy seguramente para complementar acuerdos de futuros de divisas como medida de control de costos.

2.2. La Ciencia de Datos

La Ciencia de Datos es una disciplina relativamente nueva, inclusive en muchos entornos académicos. El objetivo de este capítulo es el de resumir los aspectos mayores de la ciencia de datos como estudio multidisciplinario

cuyo objetivo es el de hacer sentido de la gran cantidad de datos que surgen de nuestro entorno, con miras a modificar los fenómenos del mundo.

2.2.1. Introducción

El primer punto de entrada es una breve introducción a la ciencia de datos. Quizás la primera persona en hacer un bosquejo de la idea fue el académico Danés Peter Naur en su libro “Concise Survey of Computer Methods”. Naur sin embargo utiliza el término más que nada para sustituir el de ciencia computacional [Naur, 1974]. El investigador de Laboratorios Bell y profesor de la Universidad de Princeton, John Tukey, hace un mejor acercamiento al escribir el primer artículo científico sobre como la disciplina de la estadística cambiaba con el advenimiento de la informática [Tukey, 1962]. Mucho más tarde fue el estadista de la Universidad de Tokio Chikio Hayashi quien definiría de manera sucinta el concepto de Ciencia de Datos como un concepto sintético para unificar la estadística, el análisis de datos y los métodos relacionados con la consecución de resultados [Hayashi et al., 1981].

2.2.2. El Científico de Datos y su Rol como Investigador

Los científicos de datos trabajan en todas las industrias y hacen frente a los grandes proyectos de datos en todos los niveles. La definición mas famosa de las habilidades que componen a un científico de datos se han atribuido al Dr. Nathan Yau, quien precisó lo siguiente:

... el científico de datos es un estadístico que debería aprender interfaces de programación de aplicaciones (API), bases de datos y extracción de datos; es un diseñador que deberá aprender a programar; y es un computólogo que deberá saber analizar y encontrar datos con significado...

2.2.3. La Ciencia de Datos como Herramienta Predictiva

Uno de los enfoques principales de la ciencia de datos es el procesamiento de datos estructurados o no estructurados para la obtención de conocimiento. Es importante destacar que la ciencia de datos trabaja en condiciones especiales que la definen de otras disciplinas (como por ejemplo, la inteligencia de negocios).

2.2.4. Diseño de un Estudio de Ciencia de Datos

El científico de datos es responsable de guiar el proyecto de ciencia de datos de comienzo a fin. El éxito de un proyecto de ciencia de datos no se da por la utilización de alguna herramienta en particular, sino de tener goles cuantificables, buena metodología, interacción interdisciplinaria, y un flujo de trabajo adecuado. Hay seis pasos principales en el diseño de un proyecto de ciencia de datos [Zumel and Mount, 2014].

2.2.5. Tareas Comunes en la Ciencia de Datos

Hemos hablado de la ciencia de datos y su carácter predictivo. Las tareas mas comunes para lo cual se utiliza la ciencia de datos son las siguientes.

- **Clasificación:** Decidir si algo pertenece a una categoría u otra
- **Puntuación:** Predecir o estimar un valor numérico, tal como lo es un precio o la probabilidad de un fenómeno
- **Ranking:** Aprender a ordenar objetos por preferencias
- **Agrupamientos:** Agrupar objetos en grupos de características homogéneas
- **Relaciones:** Encontrar relaciones o causas potenciales de efecto tal cual se ven en la data
- **Caracterizaciones:** Utilización general de visualizaciones y reportes de la data

2.3. Aprendizaje Automatizado

El aprendizaje automatizado juega un rol principal en la metodología de investigación del trabajo. La propuesta del aprendizaje automatizado es sencilla: si existe una fuente estadísticamente grande de datos, o si la misma se puede conseguir, utilicemos esos datos para resolver el problema entrenando a los mismos para buscar la solución.

2.3.1. Introducción al Aprendizaje Automatizado

Es interesante que los métodos de aprendizaje automatizado proliferaron de forma paralela al concepto de ciencia de datos, y solo fueron absorbidos

por esta en los últimos diez años. Alpaydim nos describe el aprendizaje automatizado como la programación de computadoras para optimizar un criterio de desempeño utilizando datos o experiencia pasada (Alpaydim, E., 2010). Tom Mitchell respeta este concepto al describir el aprendizaje automatizado como "... la construcción de programas computacionales que aprenden con la experiencia..." (Mitchell, T., 1997, pág. XV). Solo Peter Harrington utiliza una descripción mucho más simplista al determinar que "El aprendizaje automatizado es la extracción de información de la data." (Harrington, P. 2012, pág. 5).

2.3.2. Métodos Supervisados y No-Supervisados

Para los autores Hastie, Tibshirani, y Friedman el aprendizaje supervisado intenta aprender una función f de predicción a través del uso de uso juegos de datos de entrenamiento en forma de muestras del total de los datos disponibles. El uso de datos de entrenamiento le permite al sistema aprender y minimizar el error del modelo de predicción [Hastie et al., 1997].

Harrington nos da una explicación más sencilla del término, al aclarar que el aprendizaje supervisado es aquel que le pide al computador aprender de los datos utilizando una variable específica como objetivo. Esto reduce la complejidad de algoritmos y patrones que se deben derivar de la muestra de datos [Harrington, 2012].

El profesor Alpaydin agrega que el aprendizaje supervisado tiene como objeto aprender un mapeo de los elementos de entrada a los de salida, teniendo en cuenta que los valores correctos de estos últimos están dados por el supervisor [Alpaydin, 2010].

2.3.3. Error Muestral y Error Fuera de Muestra

El siguiente concepto es fundamental dentro de la teoría de aprendizaje automatizado, y la terminología puede diferir un poco de los términos establecidos en la estadística inferencial.

- **Error dentro de la muestra:** es el margen de error que se obtiene al utilizar el juego de datos de entrenamiento en la construcción del modelo de predicción. También se conoce como error de re-substitución
- **Error fuera de muestra:** es el margen de error que se obtiene cuando se aplica el modelo de predicción a un nuevo juego de datos. También se lo conoce como error de generalización.

2.3.4. Diseño de un Estudio de Aprendizaje Automatizado

El diseño de una investigación de ciencia de datos tiene seis pasos. El diseño del estudio de un problema de aprendizaje automatizado debe verse como el diseño de la fase de modelo (paso tres) mucho más detallado para no confundirlos (Leek, J. 2015).

2.3.5. Tipo de Errores

El concepto de error en estadística es uno que embarca varias dimensiones. En lo que respecta al aprendizaje automatizado, no importa que tan grande sea la muestra ni que tan exacto sea el algoritmo, siempre cabe la probabilidad - aunque pequeña - que una predicción sea falsa a pesar de que arroja un resultado positivo. Podemos entonces dividir los tipos de errores según su predicción y verdadera naturaleza (Yakir, B. 2011).

2.3.6. Sobreajuste

En aprendizaje automatizado, el sobreajuste (también es frecuente emplear el término en inglés *overfitting*) es el efecto de sobre-entrenar un algoritmo de aprendizaje con unos ciertos datos para los que se conoce el resultado deseado. Daroczi define el sobreajuste como la descripción del modelo en conjunto con el ruido aleatorio de la muestra en vez de solo el fenómeno generador de datos [Daroczi, 2015]. El sobreajuste ocurre, por ejemplo, cuando el modelo tiene más predictores de los que puede acomodar la muestra de datos.

Según Zumel y Mount, una de las señales de sobreajuste más sencillas de detectar se da cuando un modelo tiene un excelente desempeño en el juego de datos que se entrenó, pero uno muy malo en un juego de datos nuevo [Zumel and Mount, 2014]. Esto es causa y efecto de memorizar la data de entrenamiento en vez de aprender reglas generales de la generación del patrón.

2.3.7. R y la Biblioteca CARET

La biblioteca *CARET* (nombre extraído de Classification And Regression Training) es una librería de funciones en R para optimizar el proceso de crear modelos predictivos. El paquete contiene herramientas para:

- segmentar juegos de datos
- preproceso de los datos

- seleccion de predictores
- optimizacion del modelo utilizando reconfiguracion de muestras
- estimacion de la importancia de la variable

El paquete esta mantenido en GitHub bajo la administración del Doctor en Estadística Max Kuhn.

2.4. Pronosticando Valores con Regresión

2.4.1. Regresión Lineal

Downey describe la regresión lineal como aquella que está basada en modelos de funciones lineales [Downey, 2014]. Para Mann y Lacke la regresión lineal es aquella que se da como una función lineal entre dos variables, y la cual se puede dibujar en el plano cartesiano como una recta [Mann and Lacke, 2010].

La teoría detrás de la regresión lineal es bastante homogénea a través de todos los autores. Zumel y Mount describen la regresión lineal como el más común de los métodos de aprendizaje automatizado [Zumel and Mount, 2014]. Para los autores hay una probabilidad muy grande que el método funcione bien con el problema, y si no, es muy fácil verificar cual otro método probar como segunda opción. Para Daróczi, el énfasis está en los modelos de regresión multivariable (una extensión de la regresión lineal simple de un solo predictor y resultado) que construyen el camino para la predicción de fenómenos complejos en la naturaleza y negocios (Daróczi, G., 2015). Por su parte, Harrington resume los beneficios de la regresión lineal (Harrington, P., 2012) por la facilidad de interpretar los resultados y lo frugal en el uso de ciclos de computación (aunque puede ser menos útil si el fenómeno no es perfectamente lineal).

La mayor parte de la teoría de esta sección sigue el desarrollo de la fórmula:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

2.4.2. Regresión Multi-Variable

Para Downey (Downey, 2015), la regresión múltiple es aquella en la cual se utilizan múltiples variables independientes, pero una sola variable dependiente.

El Dr. Tattar de la Universidad de Bangalore define que el modelo de regresión línea simple no es realista ni aplicable al mundo práctico (Tattar, 2013). Para aplicaciones más reales, es casi obligatorio el uso de modelos de regresión múltiple, en los cuales varias variables independientes se conjugan como parámetros de regresión.

La regresión multivariable no es un tema mayormente complicado en teoría cómo lo es en llevar a la práctica. No todos los ejemplos de regresiones multivariadas nos van a llevar a funciones lineales, sino que estamos tocando el límite entre regresión lineal y métodos de regresión general con funciones no lineales que pueden necesitar de transformaciones matemáticas para obtener un modelo apropiado [Daroczi, 2015]. Aquí también se explica la selección de un modelo con múltiples variables independientes y cuales conviene seleccionar [Viswanathan and Viswanathan, 2015].

La mayor parte de la teoría de esta sección sigue el desarrollo de la fórmula:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \epsilon_i$$

2.4.3. Presunción del Modelo

Hay cinco factores que deben darse en un modelo cuya presunción es que su muestra sigue una distribución normal. En esta sección revisamos cada uno de esos cinco factores [Daroczi, 2015].

2.4.4. Calce de los Datos

Llegar a un modelo de regresión lineal no significa llegar a una solución óptima, ni mucho menos. Los datos pueden calzar de forma muy elástica dentro del modelo, por lo que debemos recurrir a los coeficientes de correlación y determinación para verificar si el modelo tiene algún poder predictivo de uso científico

R - Coeficiente de Correlación de Pearson

El coeficiente de correlación de Pearson mide la relación lineal entre dos variables aleatorias cuantitativas. La correlación de Pearson es independiente de la escala de medida de las variables, lo que permite tener comparaciones mucho más objetivas independiente del fenómeno estudiado.

De manera menos formal, podemos definir el coeficiente de correlación de Pearson como un índice que puede utilizarse para medir el grado de relación de dos variables siempre y cuando ambas sean cuantitativas.

$$R = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma(x_i - \bar{x})^2 \Sigma(y_i - \bar{y})^2}}$$

R2 - Coeficiente de Determinación]

El coeficiente de determinación - denominado R^2 - es un estadístico usado en el contexto de un modelo estadístico cuyo principal propósito es predecir futuros resultados o probar una hipótesis. El coeficiente determina la calidad del modelo para replicar los resultados, y la proporción de variación de los resultados que puede explicarse por el modelo.

En el caso de regresión lineal, la fórmula del coeficiente de determinación sigue la siguiente forma:

$$R^2 = \frac{\sigma_{XY}^2}{\sigma_X^2 \sigma_Y^2}$$

Valor de p

2.4.5. Seleccionando Variables de Predicción

Un modelo de regresión de múltiples variables independientes puede tener una solución parsimoniosa sin necesidad de incluir todos sus términos. Esto es verdad cuando la cantidad de variables independientes utilizadas en el análisis

2.5. Series de Tiempo

2.5.1. Introducción a las Series de Tiempo

Muchos autores han escrito sobre las series de tiempo, pero es difícil agregar al tema o discutir las ideas del profesor Robert Hyndman, uno de los expertos más respetados en la comunidad de la estadística por su trabajo en las series de tiempo. Hyndman extiende la teoría a las series de tiempo como elementos de pronóstico y su relación con la regresión lineal (Hyndman, R., 2014). Desde el punto de vista técnico, Hyndman es el creador de varias bibliotecas de funciones de pronóstico utilizando series de tiempo y ARIMA en lenguaje R. Dentro de la bibliografía, Daróczi es quien agrega detalles sobre la detección temprana de valores atípicos que pueden dificultar – y mucho – el análisis (Daróczi, G., 2015).

2.5.2. Pronóstico con Series de Tiempo

Los pronósticos con series de tiempo utilizan solamente la información disponible de la variable que se propone pronosticar, sin hacer intento alguno por descubrir los factores adicionales que condicionan su comportamiento. Por lo tanto se extrapolan las tendencias y patrones temporales, pero se ignora toda la información adicional como pueden ser iniciativas de publicidad, actividad de la competencia, cambios en las condiciones económicas y otros [Hyndman and Athanasopoulos, 2014].

2.5.3. Patrones

Las series de tiempo pueden descomponerse según su patrón o tendencia en tres elementos que las componen [Velazco, M., 2017]. A saber:

1. Tendencia Secular: la tendencia secular o tendencia a largo plazo de una serie de tiempo es por lo común el resultado de factores a largo plazo.
2. Variación Estacional: Es el componente de la serie de tiempo que representa la variabilidad de los datos debido a la influencia de las estaciones.
3. Variación Irregular: Esta variación se debe a factores a corto plazo, imprevisibles, y no recurrentes que afectan la serie de tiempo.

2.5.4. Auto Correlación

De igual manera que una correlación mide la extensión de una relación lineal entre dos variables, la autocorrelación mide la relación lineal entre dos valores retrasados de series de tiempo [Hyndman and Athanasopoulos, 2014].

El valor de una autocorrelación para un r_k dado es:

$$r_k = \frac{\sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^T (y_t - \bar{y})^2}$$

donde T es el valor de período temporal de la serie de tiempo.

El autor Daroczi agrega como metodología para la verificación de autocorrelación en un juego de datos (no solo una serie de tiempos, sino cualquier juego de datos espacial) el *Índice I de Moran* [Daroczi, 2015]. Dicho índice esta dado por la formula:

$$I = \frac{N}{W} \frac{\sum i \sum j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum i (x_i - \bar{x})^2}$$

2.5.5. Precisión del Pronostico

Existen dos formas que se utilizan comúnmente para medir la precisión del pronóstico de series de tiempo. Ambas están basadas en el error absoluto o error cuadrático [Hyndman and Athanasopoulos, 2014].

$$ErrorPromedioAbsoluto(MAE) = promedio(|e_i|)$$

$$ErrorPromedioCuadratico(RMSE) = \sqrt{promedio(e_i^2)}$$

La tendencia al comparar precisión en un solo juego de datos es utilizar el MAE ya que es mas sencillo y simple de entender.

2.5.6. Entrenamiento y Evaluación

Al igual que la mayoría de los metodos de aprendizaje automatizado, las series de tiempo se suelen entrenar y evaluar con juegos alternos de muestra de datos. El tamaño de cada uno varía con el investigador, pero en serie de datos el sistema operativo tiende a ochenta por ciento de la muestra para entrenamiento y veinte por ciento para evaluación [Hyndman and Athanasopoulos, 2014].

2.5.7. Descomposición de Series de Tiempo

La descomposición de las series de tiempo facilita el análisis y la investigación exploratoria de los datos. Una de las formas mas sencillas de lograr esto es la aplicación de promedios móviles, lo cual se facilita mucho en R con el uso de la función `decompose()` [Daroczi, 2015].

La descomposición por promedios móviles toma la forma siguiente:

$$s_t = \frac{1}{k} \sum_{n=0}^{k-1} x_{t-n}$$

Muchas series de tiempo no son aditivas sino multiplicativas, y con el paso del tiempo incrementan la amplitud de las fluctuaciones. Para tales series, la biblioteca de R tiene la función `stl()` que aplica descomposicion a base del método Loess. La aplicacion del método Loess y la transformación de la función a una logarítmica tiene el efecto de replicarla como aditiva [Viswanathan and Viswanathan, 2015].

2.5.8. Suavizar Exponencialmente con Holt-Winters

Es posible eliminar todos los efectos de estacionalidad en una serie de tiempo con la aplicación del filtro Holt-Winters. Esto no solo resulta en una lectura más clara de la tendencia secular de la serie, útil para pronósticos, sino que adicionalmente tiene el efecto de eliminar valores atípicos (outliers) en la misma [Daroczi, 2015].

2.5.9. ARIMA

Uno de los métodos más populares para la descomposición de series de tiempo con periodos de doce meses es ARIMA, el cual fue desarrollado por el Buró de Censo de Estados Unidos [Hyndman and Athanasopoulos, 2014]. El método está basado en la descomposición tradicional, pero tiene la ventaja de mantener la tendencia secular en todos los puntos de datos, y de permitir que la tendencia estacional varíe de a poco con el tiempo. Es también un método muy robusto.

2.5.10. Dickey-Fuller

Un componente importante de las series de datos es la detección de si son o no auto-regresivas (lo que determina mucho de su poder predictivo). La fórmula para la detección de series auto-regresivas es el test Dickey-Fuller, y la mejor bibliografía es el artículo científico escrito por ambos profesores en la revista especializada *Econometrica* (Dickey, D., y Fuller, W., 1981). A pesar de ser un artículo contemporáneo, la teoría detrás de la prueba Dickey-Fuller nos permite descartar series de tiempo no-regresivas con poco poder de predicción.

2.6. Modelos Ensamblados

El tema de modelos ensamblados es uno que por lo general se reserva más como técnica de composición que como teoría del aprendizaje automatizado.

2.6.1. Introducción

El uso de modelos ensamblados es en cierta forma la prueba final de la hipótesis de trabajo: la utilización de dos modelos entrecruzados cuyos resultados conforman una tabla temporal de valores esperados de los cuales se genera un nuevo modelo sintético de predicción más general y con mayor

capacidad de predicción en juegos de datos de validación cruzada. Este concepto es novel; Witten y Frank lo describen como combinación de métodos múltiples, y escriben: “... un enfoque obvio para hacer mejores decisiones es tomar el resultado de diferentes métodos y combinarlos...” (Witten, I. y Frank, E., 2005). Zhou nos describe que “... los modelos ensamblados que entrenan múltiples variables y luego las combinan para uso de entrenamiento, con el Boosting y el Bagging como representantes principales, representan lo más novedoso en el estado del arte de la ciencia de datos...” (Zhou, Z., 2012, pg. VII). De una manera un tanto más coloquial, Zhang y Ma describen el uso de modelos ensamblados con una analogía de la vida real, en la cual los pacientes buscan una segunda y hasta tercera opinión de expertos antes de someterse a una operación complicada (Zhang, C. Y Ma, Y., 2012). Curiosamente tanto Zhang, Ma y Zhou hablan de la combinación de métodos de regresión general con clasificadores, y solo Witten y Frank hablan de otras combinaciones (por supuesto, Witten y Frank comenzaban a escribir en los albores del ensamblaje de métodos, cuando los clasificadores no estaban tan de moda porque el análisis era mayoritariamente de números, algo que cambió con el avance de las redes sociales).

2.6.2. Combinando Métodos

La combinación de métodos es el ultimo paso en la estrategia de construcción de un sistema ensamblado de aprendizaje automatizado. La pregunta de que métodos combinar esta estrechamente relacionado con el tipo de juegos de datos y la solución que se busca alcanzar. Por ejemplo, alguno métodos de clasificación como los vectores de soporte solo devuelven valores discretos [Zhang and Ma, 2012]. De tal manera el uso de dos métodos alternos en uno ensamblado estará determinado por la forma final en que se ensamblan y el algoritmo final utilizado para la decisión de predicción. Tanto Polikar [Zhang and Ma, 2012] como Zhou [Zhou, 2012] citan como preferibles las metodologías de voto por mayoría, promedio, promedio ponderado, y ensamblaje infinito.

2.6.3. Diversidad

La diversidad de ensamblaje, o la diferencia entre diferentes métodos de aprendizaje, es un tema fundamental en el ensamblaje de métodos [Zhou, 2012]. Intuitivamente es fácil entender que para obtener una ventaja de la combinación, es necesario que los aprendizajes sean diferentes, de otra manera la ganancia en desempeño no seria marginalmente superior a los métodos por

separado [Zhou, 2012].

2.6.4. Bagging

La idea del *bagging* esta estrechamente ligada al *bootstrapping*, y determinada por la selección de múltiples muestras de datos generadas a través de *bootstrapping*, utilizadas para alimentar clasificadores, sobre cuyos resultados el método ensamblado puede votar [Daume, 2013].

2.6.5. Boosting

El *boosting* es la técnica por la cual se toma un algoritmo de aprendizaje con malos resultados (técnicamente conocido como un clasificador débil) y se lo transforma en un clasificador fuerte. La forma en la cual funciona el *boosting* es que basado en un juego de datos y resultados pasados, va generando nuevas predicciones. Las predicciones con resultados aceptables se les pone menor peso y recursos, mientras que el algoritmo vuelve a iterar en aquellas predicciones con valores lejanos hasta que cobran fuerza [Daume, 2013]. Esta técnica recibe el nombre de **AdaBoost**, del ingles *adaptive boosting algorithm*. Esta fue una de las primeras técnicas practicas en la ciencia de datos.

Bibliografía

- [Alpaydin, 2010] Alpaydin, E. (2010). *Introduction to Machine Learning 2nd Edition*. The MIT Press, Massachusetts, USA.
- [Cárdenas, 2016] Cárdenas, A. O. (2016). *Economía Colombiana 5ta Edición*. ECOE Ediciones, Bogotá, Colombia.
- [Carriello, 2010] Carriello, B. B. (2010). *Crisis Cambiarias en Países Emergentes*. Ediciones Uninorte, Barranquilla, Colombia.
- [Daroczi, 2015] Daroczi, G. (2015). *Mastering Data Science with R*. Packt Publishing, Birmingham, UK.
- [Daume, 2013] Daume, H. (2013). *A Course in Machine Learning*. University of Maryland, Maryland, USA.
- [Downey, 2014] Downey, A. B. (2014). *Think Stats*. Green Tea Press, Massachusetts, USA.
- [Harrington, 2012] Harrington, P. (2012). *Machine Learning in Action*. Manning Publications, Shelter Island, USA.
- [Hastie et al., 1997] Hastie, T., Tibshirani, R., and Friedman, J. (1997). *The Elements of Statistical Learning*. Springer, Stanford, USA.
- [Hayashi et al., 1981] Hayashi, C., Yajima, K., Bock, H. H., Ohsumi, N., Tanaka, Y., and Baba, Y. (1981). *Data Science, Classification, and Related Methods*. Springer, Kobe Japan.
- [Hyndman and Athanasopoulos, 2014] Hyndman, R. and Athanasopoulos, G. (2014). *Forecasting Principles and Practice*. Otexts, Melbourne Australia.
- [Mann and Lacke, 2010] Mann, P. S. and Lacke, C. J. (2010). *Introductory Statistics*. Wiley, Willimantic, USA.

- [Naur, 1974] Naur, P. (1974). *Consice Survey of Computer Methods*. Petrocelli Charter, Lund Sweden.
- [Plazas, 2013] Plazas, J. R. (2013). *Economia Colombiana*. Libreria Ediciones del Profesional Ltda., Bogotá, Colombia.
- [Roca et al., 2002] Roca, A. M., de la Hoz, J. V., Morón, J. B., and Diaz, M. A. (2002). *Experiencias Exportadoras del Caribe Colombiano*. Colección de Economía Regional.
- [Tukey, 1962] Tukey, J. (1962). The future of data analysis. *The Annals of Mathematical Statistics*, 33:67.
- [Viswanathan and Viswanathan, 2015] Viswanathan, V. and Viswanathan, S. (2015). *R Data Analysis Cookbook*. Packt Publishing, Birmingham, UK.
- [Zhang and Ma, 2012] Zhang, C. and Ma, Y. (2012). *Ensemble Machine Learning*. Springer, New York, USA.
- [Zhou, 2012] Zhou, Z. (2012). *Ensemble Methods*. Chapman and Hall - CRC, Boca Raton, USA.
- [Zumel and Mount, 2014] Zumel, N. and Mount, J. (2014). *Practical Data Science with R*. Manning, Shelter Island, USA.