

# Predicción del precio del Bitcoin mediante ventanas móviles y un ensamble de estimadores por Modelos Lineales modelados a través de Bagging con Árboles de Regresión.

Vanessa Alcalde, Pablo Martinez Angerosa

Análisis Multivariado II

Diciembre 17, 2020

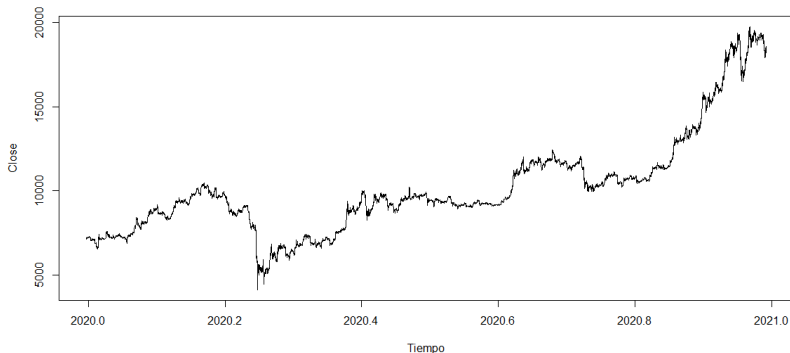
- Bitcoin, una criptomoneda digital esta actualmente siendo integrada y aceptada por los principales actores del mundo de las finanzas.
- Existen diversas investigaciones que analizan la composición del precio del Bitcoin.
- Bitcoin es un paseo aleatorio.
- El algoritmo recomendado para la predicción del precio del Bitcoin es Modelos Lineales.
- Variables explicativas recomendadas están compuestas de rezagos del precio y el volumen del Bitcoin.

- Objetivo proponer alternativas a las investigaciones realizadas actualmente que mejoren el rendimiento.
- Se utilizó un sistema de ventanas móviles el cual genera un ensamble de estimadores mediante Modelos Lineales, que son consideradas como opiniones expertas, y se proponen diversas técnicas para sintetizar estas opiniones en una sola predicción.
- El modelo propuesto logra duplicar el rendimiento de otras técnicas.

# Datos - Descripción de los datos

- Se utilizan los datos por hora del precio de *Cierre* y *Volumen* de transacciones de Bitcoin.
- Los datos se descargan de la API de Binance.
- Registros históricos por hora en dólares americanos (USD) de Bitcoin(BTC).
- La base de datos contiene 8717 registros comenzando en el día 12/12/2019 a las 08:00:00h hasta el 10/12/2020 a la 01:00:00h.

# Datos - Preparación de los datos



Serie de Tiempo del precio del *Cierre* por hora del Bitcoin desde el día 12/12/2019 a las 08:00:00h hasta el 10/12/2020 a la 01:00:00h.

# Datos - Preparación de los datos

- La variable de predicción  $Y$  corresponde al *Cierre* del precio del Bitcoin en el tiempo  $n$ .
- Se crearon variables de rezagos llamadas *lag*, para el precio de *Cierre* y *Volumen* de Bitcoin.
- Cinco primeros rezagos del precio del *Close* y del *Volumen* del Bitcoin.
- Variables son definidas como *closeLag1*, *closeLag2*, *closeLag3*, *closeLag4*, *closeLag5*, *volLag1*, *volLag2*, *volLag3*, *volLag4* y *volLag5*, respectivamente.

- Las 4 técnicas utilizadas fueron Modelos Lineales, Métodos de Ensamble de Modelos, Bagging basado en Árboles de Decisión y Redes Neuronales.
- Un modelo de regresión lineal múltiple (1) es un Modelo Lineal en los parámetros en el cual la variable de respuesta,  $Y$ , es determinada por un conjunto de variables independientes, las variables explicativas (matriz  $X$ ). Se busca el hiperplano que mejor ajuste a los datos. Los parámetros  $\beta_i$  para el modelo se obtienen por mínimos cuadrados.

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i + \cdots + \beta_d x_i + \epsilon_i \quad (1)$$

- Los Métodos de Ensamble de Modelos son útiles para mejorar el rendimiento de los modelos de aprendizaje automático al mejorar su precisión.
- Se construyen varios modelos y se combinan las predicciones resultantes promediando.
- Este promedio de errores produce mejores predicciones generalizando el carácter particular de cada uno de estos modelos.



- Un árbol de decisión es una división recursiva del espacio de variables explicativas en una estructura en forma de árbol, cada nodo interior contiene una pregunta sobre una variable de entrada y cada nodo terminal una decisión.
- Pueden ser utilizados en problemas de regresión y clasificación.
- Los árboles CART (Classification And Regression Tree) se construyen dividiendo el conjunto de valores posibles de  $X_1, X_2, \dots, X_p$  en  $J$  regiones disjuntas  $R_1, R_2, \dots, R_J$ . En el caso de un árbol de regresión, para cada observación en la región  $R_j$  se predice el valor medio de las respuestas.

- Para llevar a cabo la construcción del árbol se comienza con un conjunto de datos de entrenamiento, el cual es segmentado mediante particiones binarias. Se crean regiones  $R_1, R_2, \dots, R_J$  de manera que se minimice la Ecuación (2).

$$\sum_{j=1}^J \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2 \quad (2)$$

Donde  $\hat{y}_{R_j}$  es la respuesta media para las observaciones del conjunto de entrenamiento en la región  $j$ -ésima.

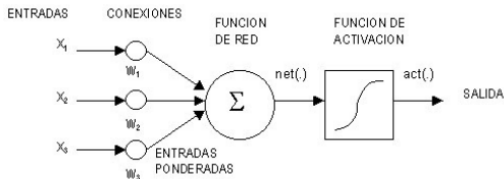
- Una vez que se encuentra la mejor partición, se separan los datos en las regiones resultantes y se repite el proceso. Este proceso termina cuando se satisface algún criterio de parada.

# Metodología - Bagging (Bootstrap Aggregating)

- Método para reducir la varianza de un modelo de aprendizaje automático.
- Se divide el conjunto de datos en entrenamiento  $L$  y testeo  $T$ , se toma una muestra bootstrap  $L_b$  de  $L$  y se construye un estimador usando  $L_b$ . Se repite el procedimiento  $B$  veces. Luego a cada dato de  $T$  se le asigna el promedio de las respuestas de los estimadores construidos en el paso anterior (para el caso de un modelo de regresión). La proporción de veces que la clase estimada difiere de la verdadera es el error Bagging.
- Bagging para Árboles de Decisión: se construyen  $B$  árboles con conjunto de entrenamiento obtenido mediante una muestra bootstrap del conjunto de entrenamiento original, luego se hace un promedio de las predicciones resultantes. Esto reduce la varianza, lo que mejora la precisión de las predicciones.

# Metodología - Redes Neuronales

- Las Redes Neuronales son un modelado matemático que homologa el comportamiento de una neurona biológica. El objetivo de su diseño es emular al cerebro humano.
- Se construyen combinaciones lineales de las entradas y se obtiene la salida como una función no lineal de estas.



# Metodología - Redes Neuronales

Se muestra el modelado de una red neuronal con una capa de entrada, una capa oculta y una salida. Donde  $i, j$  son los índices correspondientes a la capa de entrada y oculta respectivamente,  $w$  corresponde a los pesos de cada neurona y  $f(.)$  a la función de activación.

$$g_k(x) = f \left( \underbrace{\sum_{j=1}^{n_H} w_{kj} \underbrace{f \left( \sum_{i=1}^d w_{ji} x_i + w_{j0} \right)}_{\text{net}_j}}_{\text{net}_k} + w_{k0} \right)$$

- El procedimiento para encontrar los pesos que configuren el mejor modelo consiste en primero multiplicar cada dato de entrada por un peso y los valores ponderados se combinan linealmente.
- Posteriormente se aplica una función de activación no lineal. El valor de salida es comparado con el valor objetivo. La diferencia de error que se produce es utilizada para actualizar los pesos y se itera hasta obtener el criterio deseado de parada.
- Actualmente el algoritmo más utilizado para esto es el de Backpropagation.

Para evaluar el rendimiento de los distintos modelos de predicción se utilizó la medida de error RMSE.

## RMSE

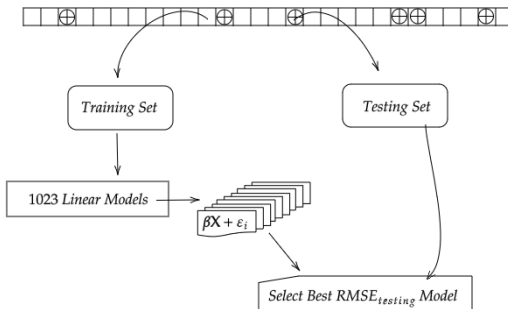
$$(\min)RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - f_i)^2}$$

Siendo  $y_i$  las predicciones y  $f_i$  los valores reales.

- 1 El conjunto de datos se dividió aleatoriamente, en una muestra de entrenamiento del 70% y una muestra de testeo del 30%.
  - 2 Se recrea la predicción del precio del Bitcoin mediante la estrategia de Modelos Lineales propuestas en la literatura.
- Se construyó un algoritmo de fuerza bruta que evalúa secuencialmente una a una todas las posibles combinaciones de las variables seleccionadas.
  - Se compara el RMSE generado por cada una de estas combinaciones con la muestra de testeo.
  - El modelo seleccionado es aquel con el menor RMSE.
  - Se cuenta con 1023 posibles modelos.



Diagrama del proceso de selección del Modelo Lineal con menor RMSE con respecto a la muestra de testeo a partir de un ensamble de 1023 posibles modelos generados desde las combinaciones obtenidas.



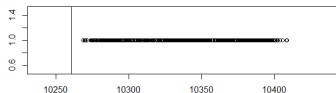
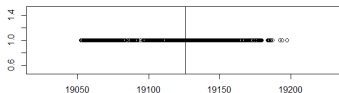
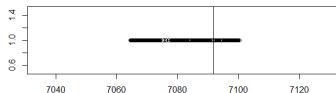
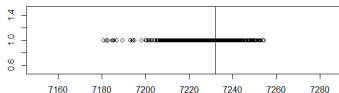
# Proceso - Ventanas móviles

- Buscando nuevas alternativas de predicción se introduce el concepto de ventanas móviles.
- Generan un enfoque dinámico de muestras de entrenamiento y testeo.
- Para predecir el precio en el momento  $n$  se genera una ventana de  $w$  rezagos desde el momento  $n - 1$  hasta el momento  $n - w - 1$  (ancho de  $w$ ).
- Se utiliza una ventana móvil de 25 rezagos.
- Para cada momento  $n$  que se quiere predecir se genera un ensamble de 1023 Modelos Lineales, a partir de todas las combinaciones de las variables seleccionadas con una muestra de entrenamiento de ancho  $w$ .

# Proceso - Linealidad y precisión del ensamble

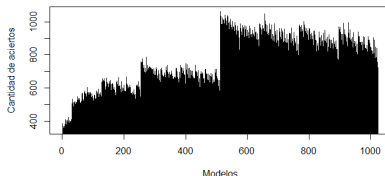
- La linealidad en cada una de las ventanas móviles es extremadamente fuerte ya que el RMSE obtenido con las mismas muestras de entrenamiento es considerablemente bajo.
- Utilizar Modelos Lineales sobre el concepto de ventanas móviles es una buena opción.
- Además entre las 1023 predicciones obtenidas por los modelos lineales en cada ventana en un 76.15% de los casos al menos uno de los modelos realizaba una estimación casi perfecta en un intervalo  $\pm 5$  dólares.

Muestra aleatoria de cuatro momentos de la base y las 1023 predicciones del precio de *Close*.



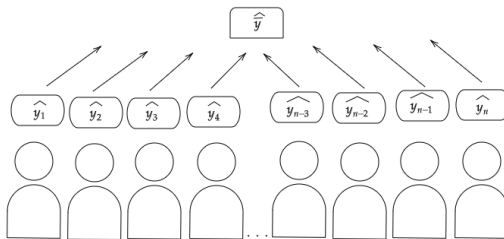
- La estrategia a seguir es encontrar posibles patrones entre los precios estimados por cada uno de los 1023 modelos y el precio real de *Close* del Bitcoin en cada ventana móvil.
- Los posibles patrones a encontrar entre este ensamble de predicciones y el precio real son muchos.
- La pregunta fundamental a responder es ¿existe algún patrón entre las 1023 predicciones por ventana de rezagos y el precio real del Bitcoin?
- Se crea una nueva base de datos donde se mantiene la temporalidad de la base anterior pero se cambian las variables explicativas de cada fila.
- Cada una de las filas está compuesta por un ensamble de 1023 variables que mantienen en orden estricto las predicciones generadas por los 1023 Modelos Lineales ajustados en la ventana.

- Histograma de frecuencia absoluta de aciertos de los 1023 modelos en toda la base.
- Intervalo de acierto es de  $\pm 5$  dólares.
- Umbral mínimo que logran todos los modelos de 351 (4.03%) aciertos.
- Primer umbral tiene un promedio de acierto de 545.904 (6.26%).
- Segundo umbral tiene un acierto promedio de 713.1167 (8.18%).
- Tercer umbral tiene un acierto promedio de 887.797 (10.19%).



- Siguiendo pasos fueron aplicados algoritmos de aprendizaje automático para buscar la relación entre el ensamble de estimaciones de precios y el precio correcto.
- Utilizamos Métodos de Ensamble de Modelos, Bagging basado en Árboles de Decisión y Redes Neuronales.
- Se considera el ensamble de estimaciones (1023) por Modelos Lineales como un conjunto de opiniones de expertos.

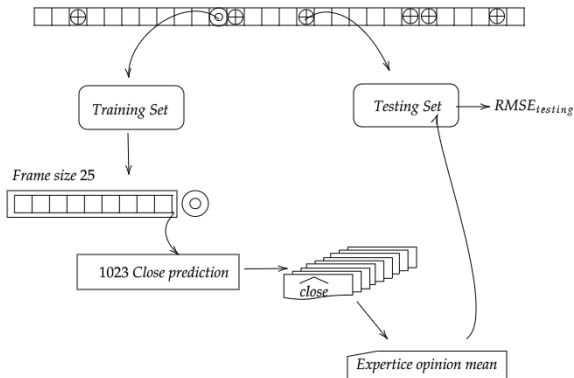
Un primer criterio para sintetizar estas opiniones es el algoritmo de Ensamble de Modelos, donde se promedian las opiniones buscando mediar las más extremas y lograr un consenso.



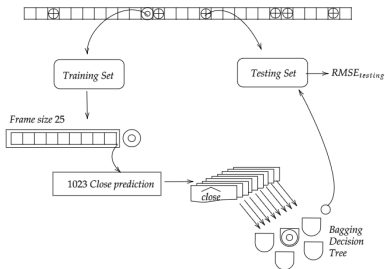


# Proceso

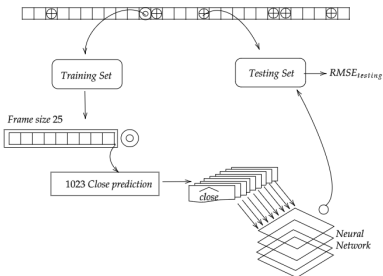
Para cada momento  $n$  a predecir de la muestra de testeo se realiza una ventana de 25 rezagos, se obtienen las 1023 predicciones para el precio del momento  $n$  (opiniones expertas) y se promedian para obtener el precio estimado.



- El segundo criterio fue utilizar Bagging con Árboles de Regresión
- Se ajusta exactamente con la misma muestra de entrenamiento con la que se realizó la primera estimación por Modelos Lineales.
- Con el modelo ajustado, se predice cada una de las muestras de testeo y se calcula el RMSE.



- El tercer criterio fue aplicar Redes Neuronales.
- Primero se entrena una red donde sus variables de entrada son el ensamble de 1023 opiniones de expertos generados por una ventana de 25 rezagos para cada momento  $n$  de la muestra de entrenamiento original
- Se calcula el RMSE con el modelo obtenido sobre la muestra de testeo.

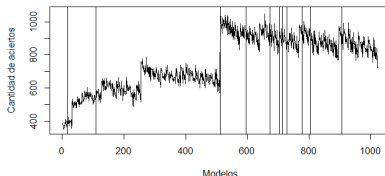


# Resultados - Modelos Lineales y Ensamble de Modelos

- Objetivo de esta investigación es encontrar metodologías alternativas a investigaciones previas y la literatura y lograr mejorar el rendimiento propuesto.
- El modelo referencia, mediante Regresión Lineal, logra un RMSE de predicción de 231.45.
- Este modelo está compuesto por dos variables explicativas, *closeLag1* y *closeLag3*.
- El Ensamble de Modelos logra un RMSE de 91.33 con respecto a las observaciones de la muestra de testeo original. Logra duplicar el rendimiento obtenido por el RMSE de referencia.

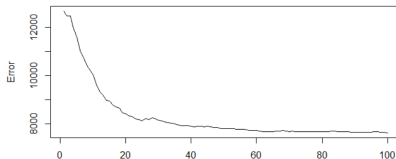
# Resultados - Bagging

- El algoritmo de Bagging con Árboles de Regresión logra un RMSE resultante de 78.87 con respecto a las observaciones de testeo de la muestra original.
- Refleja una mayor flexibilidad y complejidad de decisión.
- Se utilizaron 100 árboles y las 1023 variables de entrada.
- Se utilizó el paquete de R randomForest.
- Se contrastan las 12 variables explicativas que tienen más peso en la media de precisión en las predicciones en la muestra cuando cada una se excluye del modelo con el histograma de frecuencia absoluta de aciertos.



# Resultados - Bagging

- En la Figura se muestra como disminuye el error MSE a medida que se aumenta la cantidad de árboles utilizada para Bagging.
- Este algoritmo es el que logra el mejor RMSE de todas las propuestas.



# Resultados - Redes Neuronales

- Se utilizó una Red Neuronal con 1023 variables de entrada. Tres capas ocultas.
- La primera capa oculta es utilizada para la estandarización de los valores de las variables de entrada.
- Las siguientes dos capas están compuestas por 16 y 8 nodos respectivamente.
- Una capa de salida.
- Todas las funciones de activación son relu.
- Se realizó un entrenamiento de 1000 épocas.
- El RMSE con respecto a la muestra original de testeo es de 136.8908.
- Se utilizó el paquete Keras de R que es una capa de alto nivel de Tensorflow.

# Resultados - Redes Neuronales

- Redes Neuronales es el algoritmo con mayor flexibilidad y complejidad de decisión utilizado.
- El rendimiento con menor cantidad de nodos en las capas afecta enormemente de forma negativa el rendimiento de la red.
- Este algoritmo obtuvo el peor rendimiento, pero logra superar el rendimiento alcanzado por el modelo de referencia.

Este Cuadro muestra las distintas configuraciones de las redes neuronales empleadas y el RMSE.

Capa 2	Capa 3	Activación	RMSE
16	8	relu	136.89
8	8	relu	10640.92



# Resultados

- En el Cuadro se muestra el desempeño de cada uno de los modelos seleccionados, Ensemble de Modelos, Bagging y Redes Neuronales, comparados con los resultados obtenidos por el modelo de referencia basado en Modelos Lineales.
- Bagging es el que logra mejores resultados.
- Los tres modelos propuestos superan el rendimiento del modelo de referencia original.

Algoritmo de predicción	RMSE
Modelos Lineales	231.45
Ensamble de Modelos	91.33
Bagging (Árbol de Regresión)	78.87
Redes Neuronales	136.89

# Conclusiones

- El objetivo de esta investigación es encontrar técnicas alternativas para la predicción del precio del Bitcoin.
- El sistema de ventanas móviles presentado que genera un ensamble de estimadores mediante Modelos Lineales aplicado a la técnica de Bagging obtiene un rendimiento ampliamente mayor que los modelos propuestos originalmente.
- Aunque Redes Neuronales no obtuvo los mejores resultados, si mostró una tendencia a la mejora en predicción a medida que se aumentaba la complejidad de las redes.
- El nuevo enfoque presentado en esta investigación aplicado al algoritmo de Bagging sobre Árboles de Regresión se convierte en una opción candidata para la predicción del precio del Bitcoin.

# Referencias

- Gareth, J., Witten, D., Hastie, T., Tibshirani, R. (2013). An introduction to statistical learning: with applications in R. New York: Springer.
- Alcalde, V., Martinez Angerosa, P. (2020). Predicción del precio del Bitcoin mediante ventanas móviles y un ensamble de estimadores por Modelos Lineales modelados a través de Bagging con Árboles de Regresión.