



UNIVERSIDAD AUTONOMA DE AGUASCALIENTES

CENTRO DE CIENCIAS BÁSICAS

DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN

---

**Tesina**

“Análisis de bolsa de valores para predecir el valor de una acción en un determinado mercado, mediante sistemas predictores de inteligencia artificial, aprendizaje supervisado y minería de datos.”

**Presenta**

“Francisco Javier Hermosillo Alonso”

Para Obtener el Grado de Ingeniería en Computación Inteligente

Director:

Dr. Alejandro Padilla Díaz

Comité Tutorial:

Dr. Julio Cesar Ponce Gallegos

Dr. Francisco Javier Álvarez Rodríguez.

Aguascalientes, Ags. ,24 mayo 2021

Aguascalientes, Ags., 24 de mayo de 2021

A quien corresponda:

Por medio de la presente me permito informar que el alumno(a) **Francisco Javier Hermosillo Alonso** de la carrera de Ingeniería en Computación Inteligente con **ID: 174744**, ha terminado satisfactoriamente su tesina titulada: "Análisis de bolsa de valores para predecir el valor de una acción en un determinado mercado, mediante sistemas predictores de inteligencia artificial, aprendizaje supervisado y minería de datos.", correspondiente a la materia de Seminario de Investigación II.

Para los fines que al interesado convengan.

**ATENTAMENTE**



---

**DR. Alejandro Padilla Díaz**  
Director

Aguascalientes, Ags., 24 de mayo de 2021

A quien corresponda:

Por medio de la presente me permito informar que el alumno(a) **Francisco Javier Hermosillo Alonso** de la carrera de Ingeniería en Computación Inteligente con **ID: 174744**, ha terminado satisfactoriamente su tesina titulada: "Análisis de bolsa de valores para predecir el valor de una acción en un determinado mercado, mediante sistemas predictores de inteligencia artificial, aprendizaje supervisado y minería de datos" correspondiente a la materia de Seminario de Investigación II.

Para los fines que al interesado convengan.

ATENTAMENTE



---

Dr. Julio Cesar Ponce Gallegos

---

Dr. Francisco Javier Álvarez Rodríguez

Miembro del Comité Tutorial.

## Resumen

Con esta tesina se encontró un método eficiente en el análisis de la bolsa de valores para predecir el valor de acciones en determinadas empresas o mercados, esto mediante el uso de sistemas predictores de inteligencia artificial que hicieron uso de diversas técnicas de aprendizaje supervisado y minería de datos. Con lo anterior se crearon sistemas que resolvieron la problemática de saber cuándo es conveniente invertir en un mercado, a partir de predicciones y análisis del comportamiento del mercado que es en sí la problemática que se buscó resolver.

Se hizo uso de distintos enfoques de investigación con el propósito de encontrar el método más eficiente para encontrar los valores futuros y el comportamiento de las acciones de una empresa, mediante métodos de correlación lineal y redes neuronales. La investigación se abordó por distintos frentes con el objetivo de obtener una mejor calidad en los datos con los que se trabaja, así como de un conjunto de métodos eficientes que se aproximen a una buena descripción del comportamiento de cada uno de los mercados seleccionados.

Todo se abordó desde una perspectiva científica y matemática en la que se fueron contrastando resultados derivados de un análisis, desarrollo y comparativa de cuatro herramientas de software (Dos de redes neuronales y dos de correlación lineal) que se aproximan a una predicción real de los valores que tomó una acción en el mercado, así como de su comportamiento en el tiempo.

Estos programas hicieron uso de tratamiento en las bases de datos del valor histórico de las acciones de las 10 empresas con mejor cotización en bolsa del mercado que fueron las seleccionadas, para que posteriormente se hicieran las predicciones de posibles valores que pueden tomar las acciones de las empresas en el futuro.

Al final se hizo un análisis de los resultados obtenidos y una comparación con los cuatro modelos desarrollados en el que se verificó cuáles fueron las variables más influyentes y sobre todo si es conveniente hacer este tipo de predicciones para saber cuándo es conveniente invertir o no en un determinado mercado.

Para alcanzar lo anterior el trabajo se dividió en 5 etapas de desarrollo que permitió alcanzar todos los objetivos anteriormente mencionados, a continuación, cada fase:

1. Investigación de generalidades del mercado de valores y sobre métodos de aprendizaje supervisado.
2. Recolección de datos, generación de bases de datos y análisis de variable.
3. Definición y creación de sistemas de predicción.
4. Pruebas y fiabilidad de datos obtenidos, en el que se hicieron 120 pruebas (3 por cada empresa por modelo)
5. Comparación de resultados obtenidos, en el que se evaluó cuál fue el mejor modelo para predecir los valores que toman acciones o divisas en el futuro.

En cuanto a los resultados obtenidos el método que tuvo un mejor comportamiento para la mayoría de los escenarios fue el de correlación lineal pues fue el que obtuvo algunas ventajas con respecto al resto, ventajas que van desde una mejor precisión media en sus datos debido o que sus resultados tenían menos datos atípicos.

Este modelo fue el que tuvo una mayor estabilidad para la mayoría de las empresas analizadas, además fue el método que logró acercarse más al valor real de la acción en el mercado, pues como se dijo anteriormente este logró hacer una buena descripción del curso que tomaran el precio de las acciones para los mercados.

Además de lo anterior se concluyó que ambos métodos (correlación lineal y redes neuronales) son capaces de realizar predicciones al valor real de una acción en el futuro, además pueden llegar a describir el comportamiento real de una acción en el tiempo. Esto mediante un trabajo exhaustivo de tratamiento de datos, recolección de datos, análisis de datos y un correcto uso de parámetros que pueden variar de método a método y que sirven para ajustar al modelo y alcanzar una precisión más alta.

## Tabla de Contenido

1.	Introducción .....	15
2.	Planteamiento del problema de investigación .....	16
3.	Justificación .....	16
4.	Objetivos .....	17
4.1	Objetivo General .....	17
4.2	Objetivos Específicos .....	17
5.	Hipótesis .....	18
6.	Preguntas De Investigación.....	18
7.	Marco Teórico .....	19
7.1	Generalidades de la predicción de datos .....	19
7.1.1	Análisis Predictivo.....	19
7.1.2	Definición he Importancia del Data Mining en la predicción de datos .....	20
7.1.3	Importancia del tratamiento de datos para la predicción de variables .....	21
7.1.4	Ciencia de datos .....	22
7.1.4.1	Científico de datos .....	23
7.2	Acciones .....	26
7.3	Correlación lineal .....	28
7.3.1	Definiciones formales dentro de la correlación.....	29
7.3.2	Datos clave para entender la correlación lineal.....	31
7.4	Redes Neuronales .....	32
7.4.1	Definiciones de una red neuronal .....	32
7.4.2	Ventajas de las redes neuronales .....	32
7.4.3	Elementos básicos de una red neuronal .....	33
7.5	Predicción de valor de acciones .....	36
7.5.1	Obtención de datos.....	37
7.5.2	Preprocesamiento de datos.....	38
7.5.3	Modelización de datos .....	38
7.5.4	Interpretación y evaluación de resultados .....	38
7.5.5	Fases en el proceso de minería de datos .....	38
7.6.6	Predicción dentro de la minería de datos .....	39
7.6.7	Consideraciones al predecir en bolsa .....	40
8.	Metodología .....	43

8.1	Tipo de investigación .....	43
8.2	Nivel de la investigación .....	44
8.3	Procedimientos para la investigación .....	45
8.4	Etapas para alcanzar los objetivos de la investigación .....	46
9.	Desarrollo.....	48
9.1	Recolección de datos, generación de bases de datos y análisis de variables .....	48
9.1.1	Selección de mercados o empresas a predecir .....	48
9.1.2	Descripción de datos a obtener .....	49
9.1.3	Preparación de los datos .....	49
9.2	Definición y creación de sistemas de predicción.....	50
9.2.1	Definición de esquema general de desarrollo .....	50
9.2.2	Uso y tratamiento de las bases de datos mediante librería de Python .....	51
9.2.3	Definición de los métodos principales para la predicción de datos usando correlación lineal.....	52
9.2.4	Definición de los método principales para la predicción de datos usando redes neuronales.....	59
10.	Experimentación y pruebas .....	70
10.1	Pruebas y fiabilidad de datos obtenidos.....	71
10.2	Comparación de resultados obtenidos .....	77
11.	Análisis e interpretación de resultados .....	80
12.	Conclusiones .....	95
13.	Referencias.....	98
14.	Anexos .....	101
14.1	Anexo A. Pruebas con empresa Apple.....	101
14.2	Anexo B. Pruebas con empresa Amazon.....	106
14.3	Anexo C. Pruebas con empresa Berkshire Hathaway .....	112
14.4	Anexo D. Pruebas con empresa Facebook .....	117
14.5	Anexo E. Pruebas con empresa Alphabet.....	122
14.6	Anexo F. Pruebas con empresa Johnson & Johnson .....	127
14.7	Anexo G. Pruebas con empresa Microsoft.....	133
14.8	Anexo H. Pruebas con empresa Procter & Gamble .....	138
14.9	Anexo I. Pruebas con empresa Visa Inc. .....	143
14.10	Anexo J. Pruebas con empresa Walmart.....	149
14.11	Anexo K. Ejemplo de Base de datos (Apple) .....	154
14.12	Anexo L. Resultados general de modelo de correlación lineal 1 .....	154
14.13.	Anexo M. Resultados general de modelo de correlación lineal 2 .....	156
14.14.	Anexo N. Resultados general de modelo de redes neuronales 1 .....	158
14.15.	Anexo Ñ. Resultados general de modelo de redes neuronales 2.....	160

## **Índice de Figuras.**

Figura 1 Clasificación del Data Mining [2] .....	21
Figura 2 Diagrama de ciencia de datos de Drew Conray (AllDataScience) [4].....	22
Figura 3 representación del Impacto de la ciencia de datos en Analytics por Capgemini (AllDataScience, 2016) [4].....	23
Figura 4 Flujo de trabajo de un Científico de Datos [5].....	24
Figura 5 Funcionamiento de la bolsa de valores [6].....	25
Figura 6 Mercado de valores [8] .....	26
Figura 7 Esquema sobre tipos de valuación [10] .....	28
Figura 8 Interpretación de valores de r [14].....	29
Figura 9. correlación parcial [14] .....	31
Figura 10. Correlación y su interpretación [14] .....	31
Figura 11. Elementos básicos que componen una red neuronal [15].....	33
Figura 12. comparación entre una neurona biológica (izquierda) y una artificial [15].....	33
Figura 13. ejemplo de una neurona con 2 entradas y 1 salida [15] .....	34
Figura 14. Clasificación de las minería de datos [19].....	39
Figura 15. Etapas metodológicas .....	43
Figura 16. Niveles de Investigación.....	45

## Índice de Gráficas

Gráfica 1. Predicción de acciones de Alphabet.....	55
Gráfica 2. Predicción de acciones de Alphabet modelo 2.....	59
Gráfica 3. Gráfica de resultados de modelo de redes neuronales 1 .....	64
Gráfica 4. Resultados de modelo de redes neuronales 2. ....	65
Gráfica 5. Datos Originales Alphabet .....	70
Gráfica 6. Redes Neuronales mediante Keras y LSTM .....	70
Gráfica 7. Gráfica de dispersión de modelos de correlación lineal I .....	78
Gráfica 8. Gráfica de dispersión de modelos de correlación lineal II .....	78
Gráfica 9. Gráfica de dispersión de modelos de redes neuronales I .....	79
Gráfica 10. Gráfica de dispersión de modelos de redes neuronales II .....	79
Gráfica 11. Comparación de los 4 modelos de predicción.....	80
Gráfica 12. Mejor escenario de Correlación lineal 1 .....	81
Gráfica 13. Peor escenario de Correlación lineal 1 .....	82
Gráfica 14. Mejor escenario de Correlación lineal 2 .....	84
Gráfica 15. Peor escenario de Correlación lineal 2 .....	85
Gráfica 16. Mejor escenario de redes neuronales 1 .....	86
Gráfica 17. Peor escenario de redes neuronales 1 .....	88
Gráfica 18. Berkshire Hathaway Acciones .....	89
Gráfica 19. Mejor escenario de redes neuronales 2 .....	90
Gráfica 20. Peor escenario de redes neuronales 2 .....	91
Gráfica 21. Test 1-Correlación lineal 1-Apple.....	102
Gráfica 22. Test 2-Correlación lineal 1-Apple.....	102
Gráfica 23. Test 3-Correlación lineal 1-Apple.....	103
Gráfica 24. Test 1-Correlación lineal 2-Apple.....	103
Gráfica 25. Test 2-Correlación lineal 2-Apple.....	103
Gráfica 26. Test 3-Correlación lineal 2-Apple.....	104
Gráfica 27. Test 1-Redes Neuronales 1-Apple.....	104
Gráfica 28. Test 2-Redes Neuronales 1-Apple.....	104
Gráfica 29. Test 3-Redes Neuronales 1-Apple.....	105
Gráfica 30. Test 1-Redes Neuronales 2-Apple.....	105
Gráfica 31. Test 2-Redes Neuronales 2-Apple.....	106
Gráfica 32. Test 3-Redes Neuronales 2-Apple.....	106
Gráfica 33. Test 1-Correlación lineal 1-Amazon .....	107
Gráfica 34. Test 2-Correlación lineal 1-Amazon .....	107

Gráfica 35. Test 3-Correlación lineal 1-Amazon .....	108
Gráfica 36. Test 1-Correlación lineal 2-Amazon .....	108
Gráfica 37. Test 2-Correlación lineal 2-Amazon .....	109
Gráfica 38. Test 3-Correlación lineal 2-Amazon .....	109
Gráfica 39. Test 1-Redes Neuronales 1-Amazon .....	110
Gráfica 40. Test 2-Redes Neuronales 1-Amazon .....	110
Gráfica 41. Test 3-Redes Neuronales 1-Amazon .....	110
Gráfica 42. Test 1-Redes Neuronales 2-Amazon .....	111
Gráfica 43. Test 2-Redes Neuronales 2-Amazon .....	112
Gráfica 44. Test 3-Redes Neuronales 2-Amazon .....	112
Gráfica 45. Test 1-Correlación Lineal 1-Berkshire Hathaway .....	113
Gráfica 46. Test 2-Correlación Lineal 1-Berkshire Hathaway .....	113
Gráfica 47. Test 3-Correlación Lineal 1-Berkshire Hathaway .....	113
Gráfica 48. Test 1-Correlación Lineal 2-Berkshire Hathaway .....	114
Gráfica 49. Test 2-Correlación Lineal 2-Berkshire Hathaway .....	114
Gráfica 50. Test 3-Correlación Lineal 2-Berkshire Hathaway .....	114
Gráfica 51. Test 1-Redes Neuronales 1-Berkshire Hathaway .....	115
Gráfica 52. Test 2-Redes Neuronales 1-Berkshire Hathaway .....	115
Gráfica 53. Test 3-Redes Neuronales 1-Berkshire Hathaway .....	116
Gráfica 54. Test 1-Redes Neuronales 2-Berkshire Hathaway .....	116
Gráfica 55. Test 2-Redes Neuronales 2-Berkshire Hathaway .....	117
Gráfica 56. Test 3-Redes Neuronales 2-Berkshire Hathaway .....	117
Gráfica 57. Test 1-Correlación lineal 1-Facebook .....	118
Gráfica 58. Test 2-Correlación lineal 1-Facebook .....	118
Gráfica 59. Test 3-Correlación lineal 1-Facebook .....	119
Gráfica 60. Test 1-Correlación lineal 2-Facebook .....	119
Gráfica 61. Test 2-Correlación lineal 2-Facebook .....	119
Gráfica 62. Test 3-Correlación lineal 2-Facebook .....	120
Gráfica 63. Test 1-Redes Neuronales 1 -Facebook .....	120
Gráfica 64. Test 2-Redes Neuronales 1 -Facebook .....	120
Gráfica 65. Test 3-Redes Neuronales 1 -Facebook .....	121
Gráfica 66. Test 1-Redes Neuronales 2 -Facebook .....	121
Gráfica 67. Test 2-Redes Neuronales 2 -Facebook .....	122
Gráfica 68. Test 3-Redes Neuronales 2 -Facebook .....	122
Gráfica 69. Test 1-Correlación Lineal 1 -Alphabet .....	123
Gráfica 70. Test 2-Correlación Lineal 1 -Alphabet .....	123
Gráfica 71. Test 3-Correlación Lineal 1 -Alphabet .....	123
Gráfica 72. Test 1-Correlación Lineal 2 -Alphabet .....	124
Gráfica 73. Test 2-Correlación Lineal 2 -Alphabet .....	124

Gráfica 74.Test 3-Correlación Lineal 2 -Alphabet .....	125
Gráfica 75.Test 1-Redes Neuronales 1 -Alphabet .....	125
Gráfica 76.Test 2-Redes Neuronales 1 -Alphabet .....	125
Gráfica 77.Test 3-Redes Neuronales 1 -Alphabet .....	126
Gráfica 78.Test 1-Redes Neuronales 2 -Alphabet .....	126
Gráfica 79.Test 2-Redes Neuronales 2 -Alphabet .....	127
Gráfica 80.Test 3-Redes Neuronales 2 -Alphabet .....	127
Gráfica 81.Test 1-Correlación lineal 1 -J&J .....	128
Gráfica 82.Test 2-Correlación lineal 1 -J&J .....	128
Gráfica 83.Test 3-Correlación lineal 1 -J&J .....	129
Gráfica 84.Test 1-Correlación lineal 2 -J&J .....	129
Gráfica 85.Test 2-Correlación lineal 2 -J&J .....	130
Gráfica 86.Test 3-Correlación lineal 2 -J&J .....	130
Gráfica 87.Test 1-Redes Neuronales 1 -J&J .....	131
Gráfica 88.Test 2-Redes Neuronales 1 -J&J .....	131
Gráfica 89.Test 3-Redes Neuronales 1 -J&J .....	131
Gráfica 90.Test 1-Redes Neuronales 2 -J&J .....	132
Gráfica 91.Test 2-Redes Neuronales 2 -J&J .....	132
Gráfica 92.Test 3-Redes Neuronales 2 -J&J .....	133
Gráfica 93.Test 1-Correlación Lineal 1 -Microsoft .....	133
Gráfica 94.Test 2-Correlación Lineal 1 -Microsoft .....	134
Gráfica 95.Test 3-Correlación Lineal 1 -Microsoft .....	134
Gráfica 96.Test 1-Correlación Lineal 2 -Microsoft .....	135
Gráfica 97.Test 2-Correlación Lineal 2 -Microsoft .....	135
Gráfica 98.Test 3-Correlación Lineal 2 -Microsoft .....	135
Gráfica 99.Test 1-Redes Neuronales 1 -Microsoft .....	136
Gráfica 100.Test 2-Redes Neuronales 1 -Microsoft .....	136
Gráfica 101.Test 3-Redes Neuronales 1 -Microsoft .....	137
Gráfica 102.Test 1-Redes Neuronales 2 -Microsoft .....	137
Gráfica 103.Test 2-Redes Neuronales 2 -Microsoft .....	138
Gráfica 104.Test 3-Redes Neuronales 2 -Microsoft .....	138
Gráfica 105.Test 1-Correlación Lineal 1 -P&G .....	139
Gráfica 106.Test 2-Correlación Lineal 1 -P&G .....	139
Gráfica 107.Test 3-Correlación Lineal 1 -P&G .....	140
Gráfica 108.Test 1-Correlación Lineal 2 -P&G .....	140
Gráfica 109.Test 2-Correlación Lineal 2 -P&G .....	140
Gráfica 110.Test 3-Correlación Lineal 2 -P&G .....	141
Gráfica 111.Test 1-Redes Neuronales 1 -P&G .....	141
Gráfica 112.Test 2-Redes Neuronales 1 -P&G .....	141

Gráfica 113.Test 3-Redes Neuronales 1 -P&G .....	142
Gráfica 114.Test 1-Redes Neuronales 2 -P&G .....	142
Gráfica 115.Test 2-Redes Neuronales 2 -P&G .....	143
Gráfica 116.Test 3-Redes Neuronales 2 -P&G .....	143
Gráfica 117.Test 1-Correlación Lineal 1 -Visa .....	144
Gráfica 118.Test 2-Correlación Lineal 1 -Visa .....	144
Gráfica 119.Test 3-Correlación Lineal 1 -Visa .....	145
Gráfica 120.Test 1-Correlación Lineal 2 -Visa .....	145
Gráfica 121.Test 2-Correlación Lineal 2 -Visa .....	145
Gráfica 122.Test 3-Correlación Lineal 2 -Visa .....	146
Gráfica 123.Test 1-Redes Neuronales 1 -Visa.....	146
Gráfica 124.Test 2-Redes Neuronales 1 -Visa.....	147
Gráfica 125.Test 3-Redes Neuronales 1 -Visa.....	147
Gráfica 126.Test 1-Redes Neuronales 2 -Visa.....	147
Gráfica 127.Test 2-Redes Neuronales 2 -Visa.....	148
Gráfica 128.Test 3-Redes Neuronales 2 -Visa.....	148
Gráfica 129.Test 1-Correlación Lineal 1 -Walmart .....	149
Gráfica 130.Test 2-Correlación Lineal 1 -Walmart .....	149
Gráfica 131.Test 3-Correlación Lineal 1 -Walmart .....	150
Gráfica 132.Test 1-Correlación Lineal 2 -Walmart .....	150
Gráfica 133.Test 2-Correlación Lineal 2 -Walmart .....	151
Gráfica 134.Test 3-Correlación Lineal 2 -Walmart .....	151
Gráfica 135.Test 1-Redes Neuronales 1 -Walmart .....	151
Gráfica 136.Test 2-Redes Neuronales 1 -Walmart .....	152
Gráfica 137.Test 3-Redes Neuronales 1 -Walmart .....	152
Gráfica 138.Test 1-Redes Neuronales 2 -Walmart .....	153
Gráfica 139.Test 2-Redes Neuronales 2 -Walmart .....	153
Gráfica 140.Test 3-Redes Neuronales 2 -Walmart .....	154

## Índice de Tablas

Tabla 1. Tabla descriptiva índices [21] .....	41
Tabla 2. Clasificación de las compañías en el NASDAQ [43].....	48
Tabla 3. Descripción de los atributos [43] .....	49
Tabla 4. Resultados Correlación lineal 1 .....	72
Tabla 5. Resultados de Correlación lineal 2.....	73
Tabla 6. Resultados Redes Neuronales 1 .....	74
Tabla 7. Resultados Redes Neuronales 2.....	76
Tabla 8. Comparación de modelos.....	77
Tabla 9. Comparación de resultados Correlación lineal 2 (Mejor) .....	84
Tabla 10. Comparación de resultados Correlación lineal 2 (Peor) .....	85
Tabla 11.Comparación de resultados Redes Neuronales I (Mejor) .....	87
Tabla 12.Comparación de resultados Redes Neuronales I (Peor) .....	88
Tabla 13.Comparación de resultados Redes Neuronales II (Mejor) .....	90
Tabla 14.Comparación de resultados Redes Neuronales II (Peor) .....	91
Tabla 15. Ejemplo de Base de datos (Apple) .....	154
Tabla 16. Correlación lineal 1-Apple .....	155
Tabla 17. Correlación lineal 1-Amazon .....	155
Tabla 18.Correlación lineal 1-Berkshire Hathaway .....	155
Tabla 19. Correlación lineal 1-Facebook.....	155
Tabla 20. Correlación lineal 1-Alphabet .....	155
Tabla 21. Correlación lineal 1-J&J.....	155
Tabla 22.Correlación lineal 1-Microsoft .....	156
Tabla 23.Correlación lineal 1-P&G .....	156
Tabla 24. Correlación lineal 1-Visa.....	156
Tabla 25.Correlación lineal 1-Walmart .....	156
Tabla 26. Correlación lineal 2-Apple .....	156
Tabla 27. Correlación lineal 2-Amazon .....	157
Tabla 28. Correlación lineal 2-Berkshire Hathaway .....	157
Tabla 29.Correlación lineal 2- Facebook.....	157
Tabla 30. Correlación lineal 2-Alphabet .....	157
Tabla 31. Correlación lineal 2- J&J.....	157
Tabla 32. Correlación lineal 2-Microsoft .....	158
Tabla 33. Correlación lineal 2-P&G .....	158

Tabla 34. Correlación lineal 2-Visa.....	158
Tabla 35. Correlación lineal 2- Walmart .....	158
Tabla 36. Redes Neuronales 1-Apple .....	158
Tabla 37. Redes Neuronales 1-Amazon .....	159
Tabla 38. Redes Neuronales 1- Berkshire Hathaway .....	159
Tabla 39. Redes Neuronales 1- Facebook.....	159
Tabla 40. Redes Neuronales 1- Alphabet .....	159
Tabla 41. Redes Neuronales 1- J&J.....	159
Tabla 42. Redes Neuronales 1- Microsoft.....	159
Tabla 43. Redes Neuronales 1- P&G .....	160
Tabla 44. Redes Neuronales 1- Visa.....	160
Tabla 45. Redes Neuronales 1- Walmart .....	160
Tabla 46. Redes Neuronales 2- Apple .....	160
Tabla 47. Redes Neuronales 2- Amazon .....	160
Tabla 48. Redes Neuronales 2- Berkshire Hathaway .....	161
Tabla 49. Redes Neuronales 2- Facebook.....	161
Tabla 50. Redes Neuronales 2- Alphabet .....	161
Tabla 51. Redes Neuronales 2- J&J.....	161
Tabla 52. Redes Neuronales 2- Microsoft.....	161
Tabla 53. Redes Neuronales 2- J&J.....	162
Tabla 54. Redes Neuronales 2- Visa.....	162
Tabla 55. Redes Neuronales 2- Walmart .....	162

## Índice de Ecuaciones

Ecuación 1. Formula de varianza .....	30
Ecuación 2. Formula de covarianza .....	30
Ecuación 3. Formula del coeficiente de correlación de Pearson de covariación estandarizado .....	30
Ecuación 4. Función de entrada [15].....	34
Ecuación 5. Sumatoria de las entradas pesadas [15] .....	34
Ecuación 6. Productoria de las entradas pesadas [15] .....	35
Ecuación 7. Máximo de las entradas pesadas [15] .....	35
Ecuación 8. Función Lineal [15] .....	35
Ecuación 9. Función sigmoidea [15].....	35
Ecuación 10. Función tangente hiperbólica [15] .....	36
Ecuación 11. Función de salida Binaria [15] .....	36

## **1. Introducción**

El presente trabajo de investigación y desarrollo tiene como objeto el obtener un método de predicción eficiente para el valor que tomen las acciones de un mercado en el futuro, esto mediante técnicas de minería de datos y de aprendizaje supervisado que permitan hacer un análisis valioso de cuáles son las variables más recomendables para predecir una acción y además proporcionar un método que sirva como herramienta para hacer movimientos de inversión.

La tesis se va a estructurar primero en una investigación donde se abordarán los principales temas teóricos para el desarrollo de un sistema capaz de lograr los objetivos que plantean más adelante en la investigación, para después abordar la metodología y con ello el desarrollo de sistema de predicción.

El sistema y la investigación desarrolladas en el presente documento buscan principalmente el encontrar un método de predicción lo más eficiente y preciso posible para lograr predecir el valor que tome una acción en el futuro en un determinado mercado, esto con el propósito de saber dónde es conveniente invertir y así proporcionar un buen instrumento financiero a la hora de invertir. El sistema hará uso de las técnicas de predicción más actuales de aprendizaje supervisado y minería de datos en particular redes neuronales y correlación lineal, pues según la literatura son de los métodos de predicción más eficientes y aplicables a más áreas de desarrollo. El sistema será desarrollo al completo en Python usando librerías de machine learning y de tratamiento de datos más recientes.

Otro de los puntos importantes del cual partirá el desarrollo de este sistema, es por el cual existen muchas organizaciones y empresas que buscan en la bolsa de valores herramientas que les doten de la información necesaria sobre posibles movimientos en el futuro de las bolsas para realizar los procesos de emisión, intercambio de valores y de otras herramientas financieras lo más precisa posible y sin asumir perdidas.

## **2. Planteamiento del problema de investigación**

El problema que se plantea es el de encontrar un método eficiente para el análisis de bolsa de valores para predecir el valor que tome una acción en un determinado mercado, mediante sistemas predictores de inteligencia artificial usando diversas técnicas de aprendizaje supervisado y minería de datos. Con lo anterior se realizará un sistema que busque resolver la problemática de saber cuándo es conveniente invertir en una criptomoneda o mercado, a partir de predicciones y análisis de datos.

Como menciona Huertas López en [16] esta problemática tiene su origen en la enorme complejidad que tiene el determinar cuándo es conveniente invertir, pues a la hora de hacer inversiones en cualquier mercado se tienen que tomar en cuenta una gran cantidad de variables que hace que sea cada vez más complejo el predecir el comportamiento de un mercado y si retornara o no la inversión inicial.

Además de lo anterior Calvo González en [21] deja en claro que la mayoría de los sistemas desarrollados para predecir el valor y el comportamiento de las acciones de un mercado tienen un solo método para hacer una predicción, por lo que se hace necesario tener un enfoque con el uso de distintas herramientas para hacer una mejor predicción.

## **3. Justificación**

Existen un gran número de sistemas que desempeñan un papel similar al que se presenta en este trabajo, como por ejemplo los presentes en [16], [18], [20] y [21] pero pocos son los que se sirven de diversas técnicas de aprendizaje, análisis de datos, análisis de mercados y uso de métodos avanzados de tratamiento de datos para realizar un análisis correcto del comportamiento de determinados mercados. A continuación, se en listan los puntos en los que aportaría esta investigación para dar solución a la problemática, tal y como recomienda Espinosa Muñoz en [34].

1. Un sistema predictor eficiente al usar diversas técnicas.
2. La investigación será multidisciplinaria, por lo que no se va a enfocar solamente al área tecnológica (Incluirá análisis de mercados, uso de métodos estadísticos, matemáticos y económicos).
3. Hará uso de los métodos más vanguardistas de aprendizaje no supervisado.
4. Un análisis más puntual de cuál es la mejor técnica de predicción de datos, en este caso de valores de mercado.
5. Tratar de reducir al máximo el error a la hora de predecir los datos de la bolsa.
6. Obtener porcentajes aceptables de asertividad en la predicción.

7. Reducir el sesgo de incertidumbre que trae el trabajar con datos a futuro con múltiples variables a predecir. Los puntos anteriores en conjunción van a garantizar resolver la problemática planteada al inicio, pues abarcan los puntos más importantes para resolver tanto de forma específica como general la problemática.

## **4. Objetivos**

### **4.1 Objetivo General**

Encontrar métodos de predicción precisos para encontrar valores de acciones en el futuro en distintos mercados, esto para saber si es conveniente invertir en ellos.

### **4.2 Objetivos Específicos**

1. Crear sistemas que hagan estimaciones del valor que puede tomar una acción en un periodo de tiempo futuro.

Se crearán dos sistemas de predicción por cada uno de los siguientes métodos:

- Redes neuronales
- Correlación lineal

2. Evaluar cuál es el mejor sistema para predecir los valores que toman las acciones en el futuro con respecto a datos reales y simulaciones.

3. Establecer cuáles fueron los mercados donde se obtuvieron mejores resultados en las predicciones obtenidas de los modelos desarrollados.

## **5. Hipótesis**

Las redes neuronales y los métodos de correlación lineal son métodos que ayudan a predecir el valor de acciones de un determinado mercado, mediante el uso de minería de datos y una selección correcta de variables.

## **6. Preguntas De Investigación**

### **Pregunta general de la Investigación:**

¿Cuáles son los métodos de predicción más precisos para encontrar valores de acciones en el futuro en distintos mercados y con ello conocer si es conveniente o no invertir en ellos?

### **Preguntas específicas de la Investigación:**

1. ¿Cómo funciona la bolsa de valores?
2. ¿Qué es una acción?
3. ¿Qué datos son necesarios para alimentar las bases de datos para predecir valores de acciones?
4. ¿Cuáles variables son mejores para predecir valores de acciones?
5. ¿Cómo se puede predecir un acción?
6. ¿Cuál sistema de predicción es mejor para predecir valores futuros de acciones?
7. ¿Cuándo es correcto invertir o no en un determinado mercado?

## 7. Marco Teórico

En esta sección de abordaran las consideraciones teóricas para entender el funcionamiento del mercado, de los métodos de predicción y del análisis de variables para entender cómo realizar una predicción en el valor de una empresa y a partir de ello entender si es conveniente o no invertir en ella. A continuación, los temas que se abordaran:

- a. Generalidades de la predicción de datos: En esta sección se analizará la importancia que tiene el análisis predictivo para hacer una buena elección de variables y un correcto trabajo de minería de datos para realizar una buena predicción. Se definirán varios conceptos tales como (minería de datos, ciencia de datos, análisis predictivo y científico de datos)
- b. Acciones: En esta sección de definirán los principales temas financieros asociados a la inversión en acciones tales como la definición de acción, tipos de derechos adquiridos en una acción, consideraciones a la hora de hacer la compra de una acción entre otros.
- c. Correlación lineal: Aquí se definirán las definiciones formales dentro de la correlación y los datos clave para entenderla y con ello construir un modelo de esta naturaleza para realizar predicciones.
- d. Redes Neuronales: Aquí se definen las definiciones de una red neuronal, sus ventajas con respecto a otros modelos a la hora de hacer una predicción y los elementos básicos que la componen. Todo esto para construir un modelo de redes neuronales y generar una predicción a partir de ello.
- e. Predicción de valor de acciones: En esta sección se abordarán todas las generalidades del concepto de “predicción”, que van desde sus antecedentes, la estimación de su fiabilidad, manejo de datos, modelización de datos y finalmente la interpretación y evaluación de resultados en la predicción.  
De igual forma se definirán algunas definiciones de la minería de datos dentro de la predicción, mencionando la importancia que tiene la importancia dentro de la generación de un modelo predictivo. Posteriormente se incluirán algunas consideraciones al predecir en bolsa y los errores principales a la hora de hacer un modelo de predicción.

### 7.1 Generalidades de la predicción de datos

#### 7.1.1 Análisis Predictivo

Según Carlos Espino Timón [1] el área del análisis predictivo es un área dentro de la minería de datos que ayuda a la extracción de información de grandes volúmenes de datos para lograr predecir ciertos patrones o tendencias en el comportamiento de los datos con el único objetivo de predecir eventos en el pasado o futuro. Este según Espino se fundamenta en lograr identificar las relaciones que existen entre las variables de eventos pasados que

nos servirá para luego explotar esas relaciones y predecir el comportamiento y los datos que puede llegar a tener un evento. Lo anterior tendrá resultados solo si el análisis elaborado tiene una las suposiciones correctas que logren tener los resultados esperables.

Para Espino el análisis predictivo puede asemejarse a hacer un pronóstico, pero deja en claro que esto es muy distinto, lo describe mediante la siguiente analogía “*Mientras que un pronóstico puede predecir cuántos helados se van a vender el mes que viene, el análisis predictivo puede indicar qué individuos es más probable que se coman un helado*” [1] Esto a palabras de Espino refleja que si la información se utiliza de una manera óptima habrá un cambio fundamental en los valores del análisis para así dirigir los esfuerzos a un enfoque más productivo y funcional.

Espino [1] deja claro que para realizar un análisis predictivo correcto es indispensable trabajar con grandes volúmenes de información, tanto del pasado como del futuro, pues para lograr entender comportamientos, patrones y tener un mayor conocimiento de los datos es fundamental contar con la mayor cantidad de información.

Este proceso de recopilación y análisis de la información se puede realizar mediante técnicas de aprendizaje computacional, pues las computadoras son capaces de trabajar y aprender de forma autónoma el trabajo con grandes volúmenes de información.

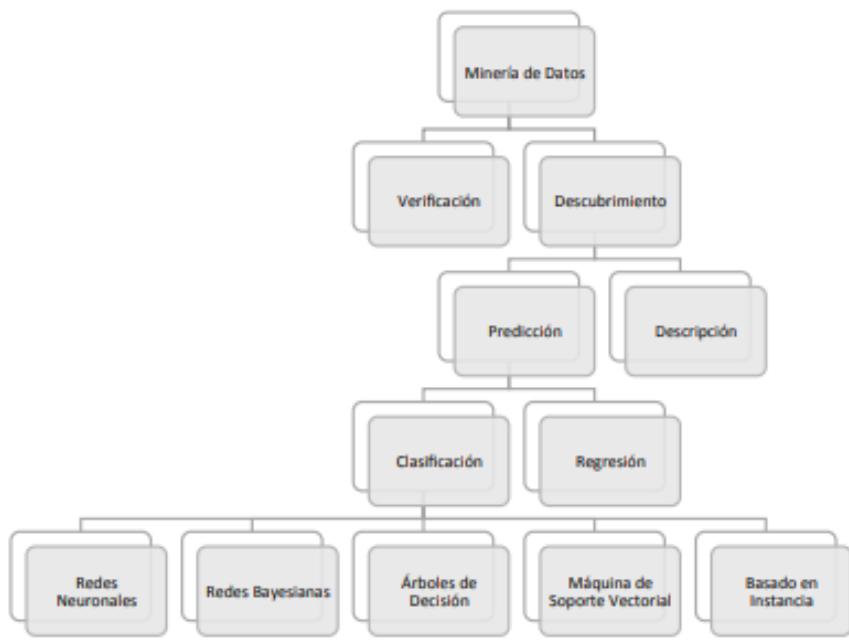
### **7.1.2 Definición he Importancia del Data Mining en la predicción de datos**

La minería de datos según Hernández, Ferrari & Ramírez, 2004 [2] es un proceso que se basa en extraer información (que sea útil y comprensible) de grandes volúmenes de información de datos previamente almacenadas en distintos formatos, en la que la minería de datos es crucial para encontrar modelos inteligibles a partir de la información recopilada.

Mientras que para Kamber 2006 [2] se refiere solamente a la extracción de conocimiento de grandes volúmenes de información de distinta naturaleza. De igual forma menciona que las ideas de vinculación entre extracción y minería de datos se han convertido en técnicas muy populares en años recientes.

Hoy en día existen una gran variedad de modelos matemático-computacionales de data mining entre los que destacan los métodos de correlación normal para la predicción de datos o la logística para clasificar los datos. Otras técnicas como las redes neuronales y los árboles de decisión permiten tener un mejor tratamiento de los datos.

Según Maimon y Rokach 2010 [2] el data mining se puede clasificar como se puede ver en la Figura 1.



*Figura 1 Clasificación del Data Mining [2]*

La importancia del data mining en la predicción de datos radica en que permite determinar la influencia de distintos factores a la hora de predecir o estimar variables en su comportamiento en el futuro. Así como la habilidad de clasificar la información en grupos con características similares, obtener secuencias que provoquen un determinado comportamiento, estimar valores de datos entre otras herramientas que en conjunción ayudan para desarrollar predicciones sobre los mismos datos sin la necesidad de conocer patrones o reglas que hayan sido elaboradas con anterioridad. [2]

El data mining proporciona las herramientas necesarias para que se logren hacer predicciones sobre grandes volúmenes de información de la forma más precisa posible, pues se basa en sofisticados modelos tanto computacionales como matemáticos para lograr hacer una buena predicción de la información.

### 7.1.3 Importancia del tratamiento de datos para la predicción de variables

Según Eduardo Núñez [3] el manejo de datos para predecir variables es crucial para lograr un buen modelo de predicción. Esto se refleja en la gran cantidad de modelos que hay (en su mayoría multivariados) que depende en su totalidad de los datos con los que trabajan. En [3] se hace una recomendación de distintos puntos que reflejan la importancia del tratamiento de la información a la hora de hacer una predicción. A continuación, se enlistan estos puntos:

1. Incorporar siempre que se pueda datos exactos
2. Tratar de trabajar siempre con la mayor cantidad de información (A más información, más calidad en la predicción)
3. Especificar en todo momento la complejidad o la no linealidad de los datos en cada predictor.
4. Seguir la regla de 10-15 eventos por la variable dependiente.

5. Tener en cuenta los procesos asociados a la selección de variables y a la selección escalonada de datos.
6. Utilizar la experiencia y la información que se tenga del tema para decidir qué sistema se adapta mejor a el tratamiento de datos.
7. Validar el modelo resultante en relación con parámetros de calibración y discriminación de datos (mediante técnicas de remuestreo).
8. Utilizar métodos de simplificación o reducción de datos si las predicciones son excesivamente optimistas.

#### 7.1.4 Ciencia de datos

Para Pérez Ramis [4] la ciencia de datos es un área de la investigación que es difícil de definir, pues en ella se conectan una gran diversidad de métodos científicos, de habilidades computacionales, modelos matemáticos, estadísticos que permiten extraer volúmenes de información grandes. Esto mediante conocimiento adquirido, un análisis estadístico sofisticado, técnicas de minería de datos y aprendizaje automático permite obtener un modelo predictivo.

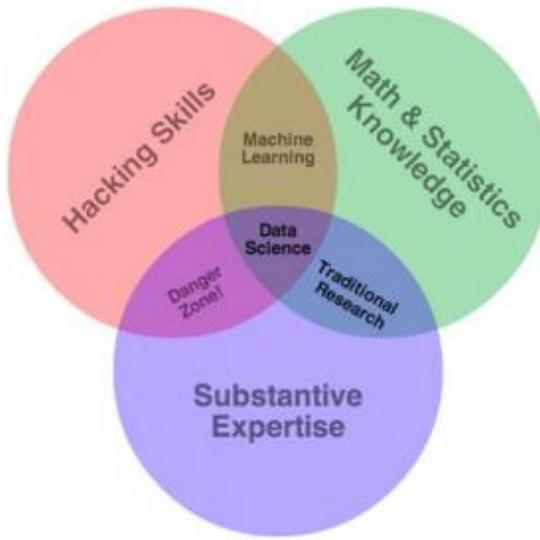


Figura 2 Diagrama de ciencia de datos de Drew Conray (AllDataScience) [4]

La ciencia de datos tuvo su primer definición en el año de 1997, en una conferencia que tuvo lugar en la Universidad de Michigan, donde se definió como una serie de datos estadísticos, pues todos los procesos y métodos de la ciencia de datos se basan en esto, pero no fue hasta 2001 que el estadístico William S. Cleveland los definió como un área que trabaja de forma independiente en la que se trata a la estadística como un área de aplicación para la ciencia de datos [4].

Fue entre los años 2002 a 2003 que empezaron a surgir distintas publicaciones en las que se empezó a ver un crecimiento en el número de publicaciones sobre ciencias de datos, que es donde precisamente se empezó a abordar su importancia para el desarrollo de modelos de predicción. [4] La figura 3 muestra a manera de resumen la evolución histórica de la ciencia de datos en los últimos años.

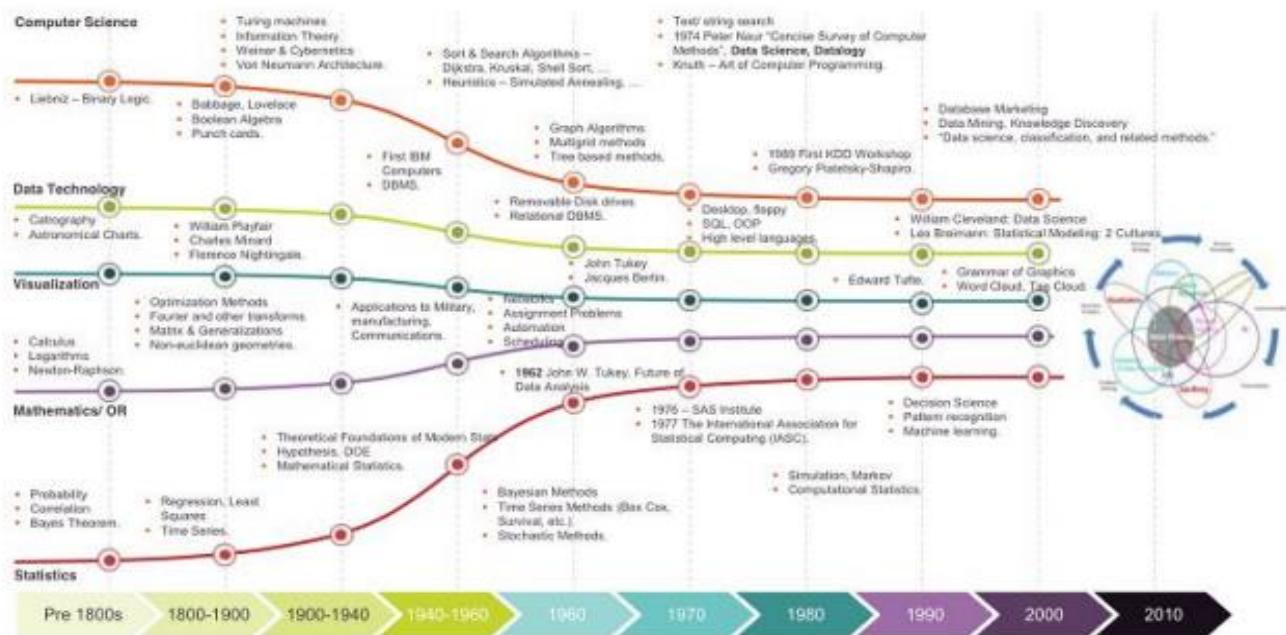


Figura 3 representación del Impacto de la ciencia de datos en Analytics por Capgemini (AllDataScience, 2016) [4]

Según Mendoza Jurado [5] que dentro de las ciencias exactas la ciencia de datos es la que se encarga de lo relacionado con la modelación, análisis, visualización, extracción de conocimiento a partir de una serie de datos. Lo anterior partiendo de un análisis de una serie de tareas secuenciales relacionados con los datos, que van desde la recopilación, preparación, abstracción y almacenamiento, que permiten tomar decisiones más efectivas a la hora de tomar decisiones con respecto al manejo de la información.

#### 7.1.4.1 Científico de datos

Dentro de la ciencia de datos un científico de datos es el que se encarga de construir los sistemas necesarios para encontrar una serie de patrones reconocibles en datos. Este es un experto que intenta hacer trabajar con eficientes modelos matemáticas que permitan reconocer patrones de información de la mano con sistemas informáticos. [5]

Un científico de datos no solo debe recolectar y analizar la información con la que trabaje, estos deben seguir una serie de pasos bien estructurados que van desde la definición de objetivos, el establecimiento de prioridades de medición (Qué es importante obtener de los datos), la propia recolección de los datos, su limpieza, análisis y interpretación. Todo esto mediante una secuencia bien definida como se logra apreciar en la figura 4.



*Figura 4 Flujo de trabajo de un Científico de Datos [5]*

## Bolsa de valores y mercado

### ¿Qué es la bolsa de valores?

Para Isaías Mireles Vázquez 2012 [6] La bolsa de valores es una organización de índole privada que proporciona las facilidades a corredores o intermediarios a diversos clientes, sobre las posibles negociaciones de compra y venta de valores.

Estas acciones van desde acciones de empresas o sociedades de distinta índole, bonos públicos y privados, además de una amplia gama de herramientas para la inversión. Todas las decisiones que se toman en la bolsa de valores se basan en precios que se conocen en tiempo real, donde los sistemas de transacciones están regulados por sistemas legales y de seguridad. Las distintas negociaciones que se realizan en la bolsa se realizan a través de corredores (sociedades de corretaje de valores), comisionistas y casa de bolsa que realizan sus operaciones a partir de una comisión. En determinados mercados estos entes logran tener acceso parcial al mercado bursátil [6].

### Objetivos de la bolsa

Según Isaías Mireles Vázquez [6] los objetivos de la bolsa de valores se pueden enlistar en los siguientes puntos:

1. Hacer una canalización del ahorro hacia la inversión
2. Ser la institución intermediaria entre empresas y las distintas organizaciones gubernamentales que requieren recursos de inversión con respecto a los ahorradores.
3. Dar acceso de liquidez a la inversión.
4. Dar la certificación necesaria a los precios del mercado.
5. Ser promotor de una correcta asignación de recursos.
6. Promover y contribuir a la valoración de activos financieros.
7. Ser proveedor de índices de la evolución de la economía.

8. Proveer de la protección necesaria de problemas como la inflación, para que no se pierda dinero con el paso del tiempo.

### Funcionamiento de la bolsa de valores

La figura 5 muestra un esquema que sintetiza el funcionamiento básico de la bolsa de valores:

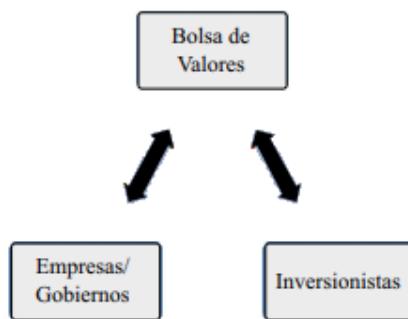


Figura 5 Funcionamiento de la bolsa de valores [6]

Muchas organizaciones y empresas buscan que la bolsa de valores proporcione la infraestructura necesaria para que haya una supervisión y se generen los servicios necesarios para que se logre el objetivo de realizar los procesos de emisión, intercambio de valores y otras herramientas financieras basadas en la regulación vigente de los países. Estas buscan que se realicen todos los manejos necesarios para que exista un correcto manejo administrativo de todas las operaciones realizadas dentro de la bolsa y que además se logre transmitir con eficacia la información a las autoridades correspondientes. Estas organizaciones tratan de hacer que exista una instancia que permita que se logre una supervisión de todas las actividades generadas por empresas y casas de bolsa, además de fomentar la expansión y competitividad del mercado de valores. [6]

Producto de lo anterior es que surgen los inversionistas que se encargan de ser los intermediarios entre las casas de bolsa, fondos de ahorro, bancos, instituciones financieras, etc. Con la finalidad de generar y obtener rendimiento a largo plazo o de ser los propietarios proporcionales de alguna empresa. Un inversionista puede llegar a participar en una gran diversidad de mercados, cada uno con sus ventajas y desventajas. [6]

### ¿Qué es un mercado?

Para Mireles Vázquez [6] el mercado es un ambiente que puede ser tanto social como virtual que da origen a las condiciones necesarias para que se genere el intercambio de algún bien o servicio. Además, la entiende como una institución mediante la cual se puede lograr el proceso mediante el cual oferentes y demandantes logren establecer una relación de naturaleza comercial para lograr alguna transacción, acuerdo o intercambio. [6]

### Investigación del mercado

Malhotra define a la investigación de mercado en [7] como la función que se encarga de conectar el consumidor, cliente y al público en general con un vendedor en particular, mediante mecanismos de información que dote de

las herramientas necesarias para identificar y definir todas las áreas de oportunidad, así como de los problemas que se generen en el mismo mercado.

Otro punto por destacar mencionado en [7] es el que menciona cuando es conveniente realizar o no una investigación de mercado, que se tiene que sustentar en consideraciones como un análisis del costo-beneficio y de los recursos con los que se cuenta para realizar una investigación. La necesidad que genera la investigación de mercado es muy similar al proceso que se da en el concepto de mercadeo; y se da principalmente en el proceso de recopilación de información, en la especialización en un determinado mercado, en el área comercial bajo la cual se hace la inversión, en posibles utilidades y en reducir el riesgo lo más posible.

### Mercado de Valores

Martínez Mendoza en [8] Describe al mercado de valores como una serie de herramientas y mecanismos que permiten la emisión y distribución de valores. Menciona que como en todo mercado existe un componente de oferta y otro que corresponde a la demanda. Para el mercado de valores la oferta representa los títulos que se pueden emitir tanto por el sector público como por el privado, mientras que por otro lado la demanda representa los fondos disponibles para la inversión que tienen como origen personas morales o físicas. En el siguiente diagrama se puede apreciar las entidades que hacen que funcione el mercado de valores que según Mendoza se pueden clasificar en la siguientes áreas (reguladoras, operativas y de apoyo).



Figura 6 Mercado de valores [8]

### 7.2 Acciones

Para Díaz en [9] una acción es el instrumento financiero que representa una renta variable que se entiende como un título que representa el capital asignado en una empresa o corporativo. Por lo tanto, los propietarios de

acciones de empresas son copropietarios y asumen parte de la responsabilidad de la empresa. De manera que si una empresa tiene un correcto desempeño provoca que el valor de las acciones se aprecie y por tanto se reciban mayores dividendos producto de la compra/venta de acciones.

### **Definición de acción**

En términos financieros se define en [9] a una acción se define como el capital social de una empresa o sociedad anónima que representa lo que tiene en propiedad, en general y bajo ciertas excepciones una acción puede ser transmisible, otorga derechos tanto económicos y políticos al dueño de la acción.

Las acciones están sujetas a negociación al público inversionista, esto a través de instituciones financieras como la Bolsa Mexicana de Valores (BMV) que tiene como objeto el financiamiento y gestión de estos propósitos. Al ser herramientas de inversión las acciones son un activo de renta variable, pues esta no representa un retorno fijo de la inversión baja la cual fue adquirida, si no que más bien depende de las variaciones y comportamiento de la empresa. Una acción otorga derechos y obligaciones a los socios que las adquieren, pero estas están sujetas a leyes vigentes para cada país.

### **Tipos de derechos adquiridos en una acción**

#### 1. Derecho Corporativo

Para [2] son todos aquellos relacionados con la participación en asambleas, representaciones, voz y voto bajo ciertas circunstancias que limiten este derecho. Estos derechos entran en juego cuando se tienen circunstancias especiales que tienden a ser estratégicas para el porvenir de una empresa o sociedad.

#### 2. Derecho patrimonial

Representan el derecho que tienen los accionistas de obtener utilidades mediante dividendos o otras acciones. Estos tienen derecho vender o transferir acciones a fin de modificar el monto nominal con el que se cuente.

### **Consideraciones a la hora de hacer la compra de una acción**

Para [10] y [11] los siguientes elementos son fundamentales a la hora de adquirir acciones de una determinada empresa:

1. Ventas que tenga la empresa.
2. Utilidades.
3. Flujo de activos (En particular efectivo).
4. Valor de la empresa frente al mercado.
5. Activo circulante (Es decir flujo constante de activos).
6. Pasivo a corto y largo plazo (se tiene que vigilar el comportamiento de la empresa a diferentes plazos)
7. Capital social y contable
8. Utilidad de la empresa

El siguiente esquema de Gitman 2012 [10] muestra las herramientas de evaluación financiera que permite hacer una correcta valuación de las acciones mediante un análisis fundamental.

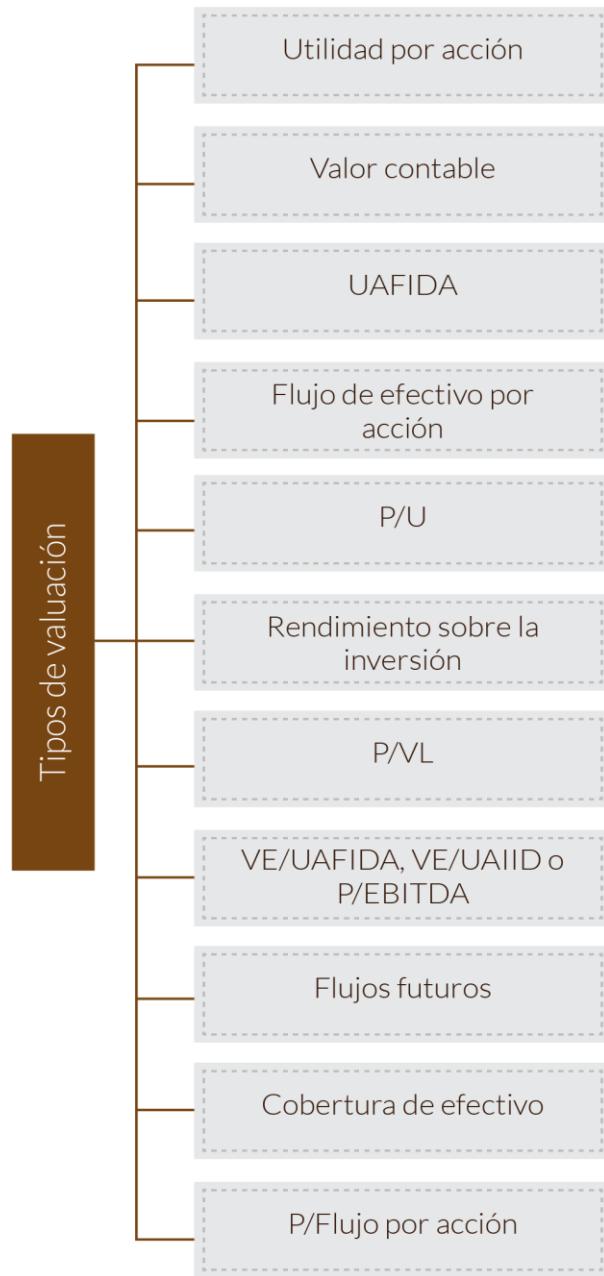


Figura 7 Esquema sobre tipos de valuación [10]

### 7.3 Correlación lineal

#### Correlación simple

Para Guillen en [13] La correlación Lineal mide el grado de precisión en una predicción que depende de una relación entre dos variables, que se le conoce como grado de correlación o incluso por asociación de dos variables. La correlación lineal corresponde a un modelo matemático que estudia la dependencia existente entre dos variables cuantitativas (En el existen tanto variables independientes y dependientes).

Dentro de la correlación simple existen dos formas:

- 1- Correlación positiva (Se da cuando el valor de  $r$  esta entre 0 y +1), en el que los valores de X y Y tienden a moverse hacia una misma dirección.
- 2- Correlación negativa (esta se da cuando el valor asignado a  $r$  va de -1 a 0), en el que los valores de X y Y tienden a moverse en dirección opuesta.

Para Vinuesa en [14] La forma más sencilla de saber si dos variables están o no correlacionadas es el determinar si estas co-varian (es decir que varían conjuntamente). Vinuesa aclara que muy importante hacer notar que una covariación no implica estrictamente causalidad, pues una correlación se puede dar por causa aleatoria. Para exemplificar el hecho anterior se pone el ejemplo entre la relación por venta de helados en un mes y el número de incendios en ese mismo periodo de tiempo provocados por una tercer variable que para este caso sería la temperatura alta.

La correlación es pues básicamente una medida de naturaleza normalizada de la covariancia lineal existente entre dos variables. Este modelo o índice puede variar en el rango de -1 a +1, donde se puede indicar un tipo de correlación perfecta al extremo del rango ya sea positiva o negativa. Mientras que el valor de  $r=0$  indica que no existe ningún tipo de correlación lineal entre las dos variables analizadas. Mientras que si nos encontramos con un valor negativo de  $r$  estaríamos ante una relación que va en sentido opuesto [14]. La siguiente figura muestra cómo se pueden interpretar los variables de  $r$ .

- **correlación despreciable:**  $r < |0.1|$
- **correlación baja:**  $|0.1| < r \leq |0.3|$
- **correlación mediana :**  $|0.3| < r \leq |0.5|$
- **correlación fuerte o alta:**  $r > |0.5|$

Figura 8 Interpretación de valores de  $r$  [14]

### 7.3.1 Definiciones formales dentro de la correlación

### 7.3.1.1 Correlación

La correlación está definida en términos de la varianza ( $s^2$ ) de las variables "x" y "y", además de la covarianza de "x", "y". Por tanto, es una medida que se encarga de la variación conjunta de las dos variables ( $\text{cov}(x, y)$ ). [14]

### 7.3.1.2 Varianza

Es la representación del promedio de la desviación de los datos con respecto a la media. [14]

Ecuación 1. Formula de varianza

$$\text{Varianza}(s^2) = \frac{\sum(x_i - \bar{x})^2}{N - 1} = \frac{\sum(x_i - \bar{x})(xi - \bar{x})}{N - 1}$$

### 7.3.1.3 Covarianza

Medida que se encarga de la relación promedio entre dos variables. Esta es la desviación promedio que se produce por el producto cruzado de estas. [14]

Ecuación 2. Formula de covarianza

$$\text{cov}(x, y) = \frac{\sum(x_i - \bar{x})(yi - \bar{y})}{N - 1}$$

### 7.3.1.4 El coeficiente de correlación de Pearson de covariación estandarizado

Unidad que resuelve la dependencia de escala de valores. Esta se encarga de medir el promedio de los datos con respecto a la media. [14]

Ecuación 3. Formula del coeficiente de correlación de Pearson de covariación estandarizado

$$\text{Coef. de correlación de Pearson}(r) = \frac{\text{cov}(x, y)}{s_x s_y} = \frac{\sum(x_i - \bar{x})(yi - \bar{y})}{(N - 1)s_x s_y}$$

### 7.3.1.5 Correlaciones parciales

Esta variable permite hacer una evaluación entre dos variables considerando el efecto de varianza de una tercera o un mayor número de variables [14]. La siguiente figura muestra el cómo se comporta este parámetro:

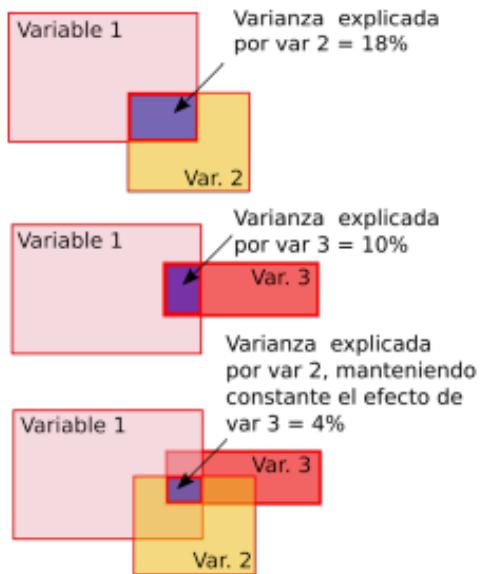


Figura 9. correlación parcial [14]

### 7.3.2 Datos clave para entender la correlación lineal

Los siguientes datos son cruciales para entender la correlación lineal, que están definidas por Vinuesa en [14]:

- Al momento de estandarizar la covarianza, el valor de  $r$  oscilará entre  $\pm 1$
- La correlación es positiva solo si ambas variables covarián hacia el mismo sentido
- La correlación es negativa solo si ambas variables covarián en sentidos opuestos
- Cuando el valor de  $r$  es de  $\pm 1$  se tiene una correlación perfecta (ajuste perfecto a modelo lineal)
- Si el valor que toma  $r$  es 0 se dice que no existe correlación entre la variable de dependiente y la variable independiente.
- Al asumir que  $r$  es una medida que está estandarizada, esta se usa en muchas ocasiones para obtener el tamaño de un determinado efecto, que se puede interpretar de la siguiente manera:

- **efecto despreciable:**  $|r| < 0.1$
- **efecto pequeño:**  $0.1 < |r| \leq 0.3$
- **efecto mediano:**  $0.3 < |r| \leq 0.5$
- **efecto grande:**  $|r| > 0.5$

Figura 10. Correlación y su interpretación [14]

-Es crucial para entender la correlación que los datos con los que trabaje se encuentren visibles mediante gráficos de dispersión, pues el valor de  $r$  solo se puede aplicar a variables que covarién linealmente.

- Se pueden utilizar correlaciones parciales para que se obtengan variables de r más de acuerdo a la realidad de las variables (x,y).

## 7.4 Redes Neuronales

### 7.4.1 Definiciones de una red neuronal

Para Matich en [15] hay diversas formas de definir a una red neuronal desde el campo de las ciencias de la computación, pues estas abarcan desde las definiciones simples y cortas hasta las que son más complejas. Matich explica cuatro posibles definiciones de una red neuronal:

1. Nuevo paradigma de la computación inspirado en el modelo biológico de las neuronas.
2. Modelo matemático constituido de un gran número de elementos organizados en distintos niveles.
3. Sistema computacional constituido por un gran número de elementos que se consideran simples, de elementos de proceso interconectados que se encargan de procesar información mediante su estado dinámico producto de entradas externas.
4. Son redes artificiales interconectadas de forma paralela entre elementos simples (que suelen ser adaptativos) estos cuentan con una organización jerárquica he intenta interactuar con objetos del mundo real, de forma similar a la que un sistema nervioso biológico lo haría.

### 7.4.2 Ventajas de las redes neuronales

En [15] se menciona que debido a los fundamentos de las redes neuronales artificiales estas tienen un elevado parecido al funcionamiento de nuestro cerebro, pues estas son capaces de adquirir experiencia y aprender de ello, de pasar de un caso particular a uno general, de abstraer distinta número de características a partir de una experiencia generada. Esto refleja las numerosas ventajas que trae consigo el trabajar con este tipo de tecnología en una amplia gama de áreas. Matich en [15] define las siguientes ventajas de las redes neuronales:

1. Aprendizaje adaptativo: Es describe la capacidad que tiene de aprender a realizar tareas que se basan en un entrenamiento o experiencia.
2. Auto-organización: Esta es la capacidad que tiene una red neuronal para crear su propia organización y de representar la información mediante etapas de aprendizaje.
3. Tolerancia a fallos: Una destrucción parcial en una red neuronal conduce indudablemente a una degradación de su estructura, pero algunas de las capacidades de la red se pueden conservar, aunque el dano haya sido considerable.
4. Operación en tiempo real: Las operaciones que hace una red neuronal se pueden llevar a cabo de forma paralela, lo que permite que tiempos de ejecución y otras variables tengan mejores números.
5. Fácil inserción dentro de la tecnología existente: Son tecnologías que se pueden integrar de forma rápida y eficiente a una gran cantidad de sistemas.

### 7.4.3 Elementos básicos de una red neuronal

A continuación, se muestra una figura con la estructura básica de una red neuronal y los elementos que la componen:

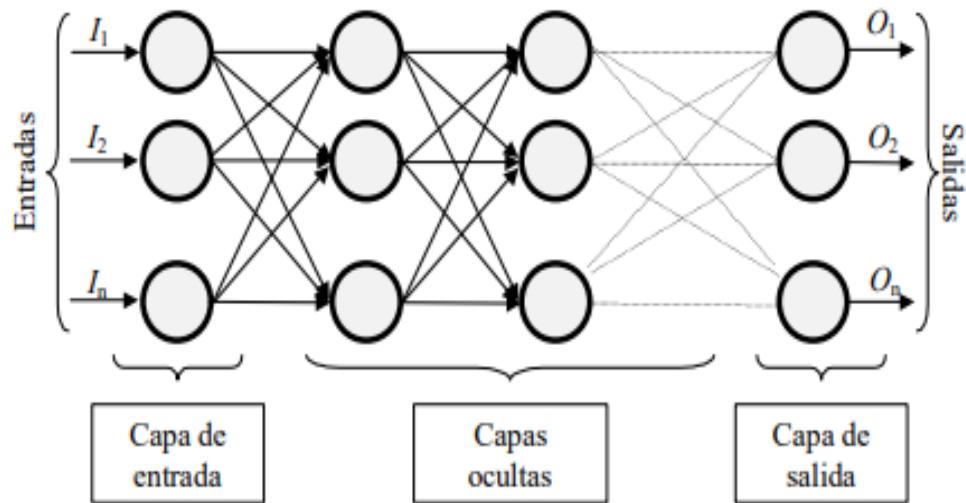


Figura 11. Elementos básicos que componen una red neuronal [15]

Una red neuronal según Matich en [15] está compuesta por una serie de neuronas que están conectadas y que se configuran en tres capas (siendo esta última opción algo que puede variar). Todos los datos que ingresen mediante la capa de entrada tienen que pasar a través de la capa de oculta y a su vez salir por la de salida. Matich destaca que la capa oculta puede llegar a estar constituida a su vez por varias capas. La siguiente figura muestra una comparación entre una neurona biológica y una neurona artificial:

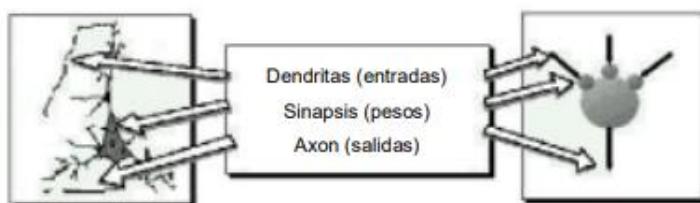


Figura 12. comparación entre una neurona biológica (izquierda) y una artificial [15]

La neurona por si misma es un sistema muy pequeño, pero cuando se junta con cientos incluso miles de neuronas más, es capaz de resolver problemas de una alta complejidad tal y como sucede en el cerebro humano que se llega a componer de billones de neuronas. [15]

#### Función de entrada

La neurona al trabaja con un gran número de datos de entrada como si se tratara de uno solo que se denomina entrada global. La función de entrada es la que se encarga de combinar un gran número de entradas simples dentro de una entrada más general, la cual se da a partir de un vector de entrada [15]. La siguiente ecuación describe la función de entrada:

Ecuación 4. Función de entrada [15]

$$input_i = (in_{i1} \bullet w_{i1}) * (in_{i2} \bullet w_{i2}) * \dots * (in_{in} \bullet w_{in})$$

donde: \* representa al operador apropiado (por ejemplo: máximo, sumatoria, productoria, etc.),  $n$  al número de entradas a la neurona  $N_i$  y  $w_i$  al peso.

En la función de entrada los datos se multiplican por pesos que son anteriormente ingresados en la neurona. Generalmente en esta función los pesos no son constantes y pueden cambiar su nivel de influencia con respecto al valor de los datos de entrada. En pocas palabras se quiere decir que un valor grande en la entrada puede tener poca influencia si estos llegan a ser lo suficientemente pequeños [15]. A continuación, se ve un ejemplo de una neurona con 2 entradas y una salida:

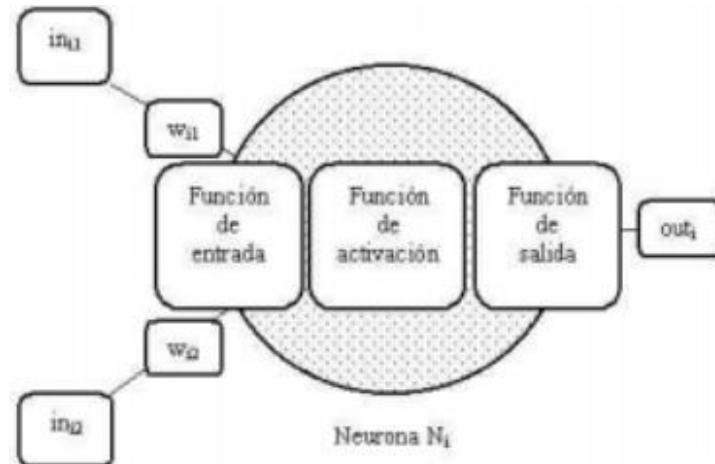


Figura 13. ejemplo de una neurona con 2 entradas y 1 salida [15]

Según Matich en [15] las funciones de entrada más utilizadas son las siguientes:

- 1) Sumatoria de las entradas pesadas: esta es un tipo de función donde la suma de todos los valores de entrada a la neurona es multiplicada por los pesos que les corresponden. Esta dada por la siguiente ecuación:

Ecuación 5. Sumatoria de las entradas pesadas [15]

$$\sum_j (n_{ij} w_{ij}), \quad \text{con } j = 1, 2, \dots, n$$

- 2) Productoria de las entradas pesadas: Esta función corresponde al producto de cada uno de los valores de entrada con respecto a la neurona, multiplicados por cada uno de sus pesos.

*Ecuación 6. Productoria de las entradas pesadas [15]*

$$\prod_j (n_{ij} w_{ij}), \text{ con } j = 1, 2, \dots, n$$

- 3) Máximo de las entradas pesadas: Esta función solo considera el valor de entrada que sea más fuerte, ya que previamente será multiplicado por su peso correspondiente.

*Ecuación 7. Máximo de las entradas pesadas [15]*

$$\max_j(n_{ij} w_{ij}) \text{ con } j = 1, 2, \dots, n$$

## Función de activación

En [15] se describe a la función de activación como la que se encarga de calcular el estado que tiene una actividad dentro de una neurona. Esta transforma la entrada global en un valor de activación, que puede tomar un rango que va de 0 a 1 o de -1 a 1. Lo anterior explica que una neurona se encuentra activa si toma 1 como valor o inactiva si toma un 0 o -1.

Para Matich en [15] las funciones de activación más utilizadas son las siguientes:

1. Función lineal: Se expresa en los siguientes términos:

*Ecuación 8. Función Lineal [15]*

$$f(x) = \begin{cases} -1 & x \leq -1/a \\ a * x & -1/a < x < 1/a \\ 1 & x \geq 1/a \end{cases}$$

con  $x = g \cdot n_i - \Theta_i$ , y  $a > 0$ .

2. Función sigmoidea: Aquí todos los valores de salida que se llegan a proporcionar por esta función están acotados dentro del rango que va de 0 a 1. Al realizar modificaciones del valor de  $g$  se va a ver afectada la pendiente en la función de activación.

*Ecuación 9. Función sigmoidea [15]*

$$f(x) = \frac{1}{1 + e^{-g \cdot x}}, \text{ con } x = g \cdot n_i - \Theta_i.$$

3. Función tangente hiperbólica: En esta función los valores correspondientes de salida de la función tangente hiperbólica están acotados en un rango que va de -1 a 1. Al realizar modificaciones en el valor de  $g$  la pendiente de la función de activación se ve afectada.

*Ecuación 10. Función tangente hiperbólica [15]*

$$f(x) = \frac{e^{gx} - e^{-gx}}{e^{gx} + e^{-gx}}, \text{ con } x = gin_i - \Theta_i.$$

## Función de salida

Matich en [15] la función de salida es la que se encarga de dar como resultado la salida de la neurona, por lo tanto, esta determina el valor que se llega a transferir a las neuronas que estén vinculadas. Se dice que, si la función de activación está por debajo de cierto umbral, no existe ninguna salida que pase a la neuro subsiguiente. En la función de salida no cualquier valor está permitido como entrada de la neurona, por ellos los valores de salida siempre tienden a estar comprendidos entre [0,1] o [-1,1]. Estos también pueden tomar valores binarios.

Para Matich en [15] las funciones de salida más utilizadas son las siguientes:

1. Ninguna: En esta función la salida es la misma que la entrada. También se le conoce como función de identidad.
2. Binaria:

*Ecuación 11. Función de salida Binaria [15]*

$$\text{Binaria: } \begin{cases} 1 & \text{si } act_i \geq \xi_i \\ 0 & \text{de lo contrario} \end{cases}, \text{ donde } \xi_i \text{ es el umbral.}$$

## 7.5 Predicción de valor de acciones

### Antecedentes

Huertas López [16] menciona la necesidad de contar con una herramienta que proporcione la información necesaria para conocer el futuro, esto bajo diversas condiciones siempre ha sido uno de los objetivos de la ciencia. Para esto es necesario explotar las distintas herramientas de análisis y desarrollo que permitan obtener buenos resultados.

López en [16] menciona que para un correcto análisis técnico de los mercados financieros es necesario estudiar con especial detalle los movimientos en el mercado, pues el prestar atención a esta variable permite hacer predicciones más precisas. En particular aconseja los siguientes puntos para realizar un correcto análisis del mercado.

1. Los movimientos que existen dentro del mercado lo son todo. Básicamente cualquier cosa o variable que llegue a afectar el precio de un determinado objeto financiero puede llegar a reflejar el valor de su precio en el mercado.
2. Los precios se mueven normalmente por tendencias.
3. El mercado tiene memoria y esta a su vez corresponde a distintos plazos.

### **Estimación de fiabilidad**

Para Deco en [17] el estudio de fiabilidad es uno de los puntos cruciales para hacer un buen modelo de predicción, en el que logra identificar principalmente dos variables:

1. Análisis local de sensibilidad. Es utilizado para determinar la estabilidad de un modelo.
2. Realizar perturbaciones a el conjunto de aprendizaje para así mejorar el rendimiento de los modelos.

Según Deco [17] se pueden clasificar en dos grupos los modelos de predicción basados en su fiabilidad, estos son:

- a. Simples: regresión lineal y regresión ponderada localmente
- b. Complejos: arboles de regresión, redes neuronales artificiales y máquinas de vectores de soporte.

Los modelos de predicción simples tratan de modelar todo el espacio de entrada al mismo tiempo, a partir de las ecuaciones de los modelos simples se puede determinar con mucha exactitud predicciones de sensibilidad, una vez conocidas la predicción inicial y el de la nueva instancia. A su vez los modelos complejos tienen una inclusión de la nueva instancia que llega a afectar la estructura del modelo, haciendo inviable o complicado para la determinación de una fórmula para predicciones de sensibilidad. [17]

Esto se resume a que un modelo complejo sirve para: Experimentos con modelos complejos.

Mientras que los modelos simples se recomiendan cuando: Ya se ha resuelto analíticamente el problema.

#### **7.5.1 Obtención de datos**

Para hacer una buena predicción a partir de un análisis de datos, lo fundamental es conseguir la mayor cantidad de datos mediante distintas técnicas que se identifican por Fuentes en [18]. La clasificación y por tanto obtención que hace de los datos es la siguiente:

1. Batch (Históricos)
  - 1.1 Estructurados: Que pueden ser bases de datos, hojas electrónicas, formato CSV.
  - 1.2 Semi estructurados: Archivos que están en formato JSON o XML.
  - 1.3 No estructurados: Texto plano
2. Stream
  - 2.1 Realtime: Este tipo de dato se da cuando cada dato que llega a la base de donde se toma la información es procesado inmediatamente.
  - 2.2 Near realtime: Este tipo de dato se da cuando periódicamente se logran procesar los datos que llegan en un determinado intervalo de tiempo.

### **7.5.2 Preprocesamiento de datos**

En [18] se explica que en esta etapa se analiza si existen datos que estén aislados o no balanceados que pueden llegar a ser causantes de errores a la hora de la obtención de datos. Esta es una de las etapas que más tiempo consume en el análisis de datos pues puede llegar a tener valores omisos o necesitar que se les someta a varias transformaciones.

### **7.5.3 Modelización de datos**

Esta etapa aplica varias técnicas de minería de datos para encontrar posibles patrones. Se llegan utilizar en su mayoría herramientas de machine learning. [18]

### **7.5.4 Interpretación y evaluación de resultados**

Esta es una de las etapas más importantes a la hora de trabajar con los datos pues implica determinar si la información con la que se trabajó en la etapa de modelización es lo suficientemente importante para los objetivos del análisis de datos. Si esta llega a no ser relevante se debe iniciar de nueva cuenta el proceso, elegir otros atributos y a aplicar otros criterios de preprocesamiento o de otros métodos de modelización. [18]

Deco en [18] recomienda los siguientes puntos cuando los resultados no han sido satisfactorios:

1. Datos de entrenamiento: Pueden que estos no logren representar de forma adecuada a la población de muestra pues probablemente no han sido seleccionados de forma aleatoria o incluso el número de la muestra es demasiado pequeño como para poder hacer generalizaciones a partir de ellos. Por lo anterior es que se recomienda que se cuente con datos de entrenamiento.
2. Datos de prueba: Este tipo de datos son bastante funcionales para probar el modelo con información nueva que no haya sido utilizada en la etapa de entrenamiento. Se recomienda que estos sean escogidos también de forma aleatoria y en una proporción dependiendo de caso del modelo.
3. Método de modelización: Se recomienda intentar con todas las que la herramientas computacionales que dispongan y que se adecúen a la naturaleza de los datos.
4. Atributos: Una correcta selección, extracción y transformación de cada uno de los atributos puede dar como resultado un modelo más eficiente.
5. Parámetros: Se pueden ajustar parámetros hasta obtener mejores resultados.

### **7.5.5 Fases en el proceso de minería de datos**

Bermúdez en [19] menciona que es necesario realizar una serie de actividades previas que van encaminadas a preparar los datos de entrada debido a que, en muchas ocasiones los datos proceden de fuentes con distinto origen, pues pueden llegar a tener un formato no adecuado o contener datos erróneos o redundantes. Por otro lado, es necesario que se interpreten y evalúen los resultados obtenidos. En [19] se identifican las siguientes fases en la minería de datos:

1. Definición de los objetivos.
2. Preparación de datos.
3. Análisis exploratorio de los datos.
4. Especificación de los métodos.
5. Análisis de los datos.
6. Evaluación de los métodos.
7. Implementación de los métodos.

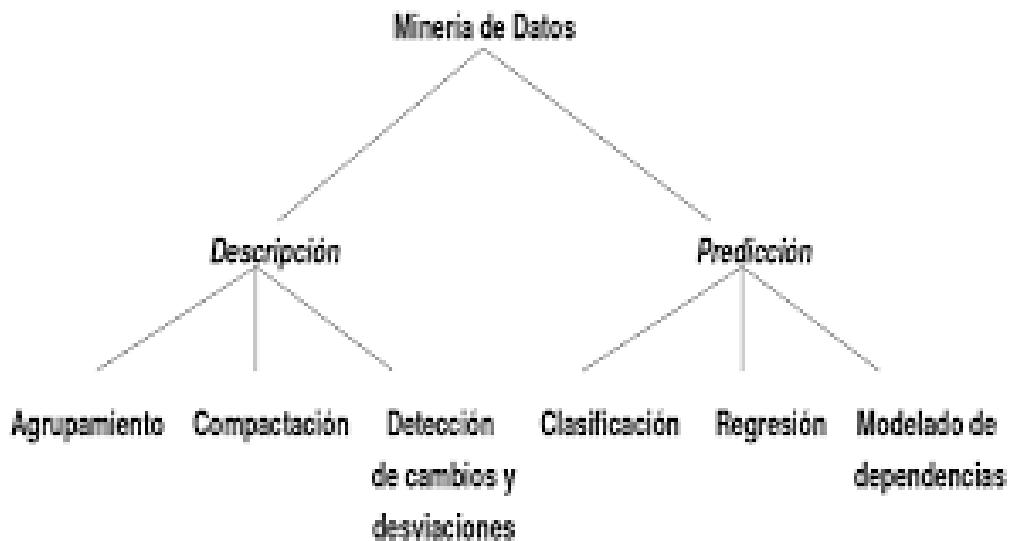
#### **7.6.6 Predicción dentro de la minería de datos**

La predicción en [19] se define como un proceso similar a la clasificación y a la estimación, donde los registros de tienen que clasificar de acuerdo con el comportamiento que pueden llegar a tener en el futuro, o incluso a un valor estimado en el futuro. La predicción es una de las ramas de la minería de datos más importante.

Una de las razones principales para que la predicción se ha tratada como una actividad separada de la clasificación o incluso de estimación es que en un modelo de predicción hay otras cuestiones referentes a la relación temporal entre las variables de entrada o de predictores de la variable objetivo.

Los datos históricos se usan básicamente para construir un modelo que intente explicar el comportamiento que se observó en los datos. Al aplicarse nuevas entradas de datos, el resultado puede llegar a ser una predicción del comportamiento futuro de los mismos.

Muchas de las técnicas de Minería de Datos son pertinentes para poder hacer una predicción a partir de los datos históricos que vayan surgiendo.



*Figura 14. Clasificación de las minería de datos [19]*

### 7.6.7 Consideraciones al predecir en bolsa

Para Londoño en [20] el realizar pronósticos de cada una de las tendencias en variables tales como el precio y rendimiento de una acción en el mercado de valores es un tema de gran envergadura y importancia, pero a su vez resulta difícil a la hora de realizar predicciones confiables que se ajusten a la realidad. Esto producto de la enorme complejidad del mercado de valores.

Pero para Londoño existen una gran cantidad de trabajos que han permitido el desarrollo de técnicas para la predicción que pueden incluso explicar el comportamiento del mercado y dar una buena predicción del precio de una acción en el futuro.

#### Indicadores bursátiles

Para las personas dedicadas a la compra/ venta de acciones según Calvo en [21] es crucial que se elabore una evaluación del estado y de la predicción del posible comportamiento a distintos plazos en el futuro de los índices bursátiles. Constantemente, los indicadores que llegan a ser utilizados para la realización de las técnicas de open (precio con el que se pueden iniciar las transacciones en una sesión bursátil) y la de close (que es el precio con el que finalizan las transacciones en una sesión).

Se sabe que, para un día concreto en el mercado, el valor de open puede llegar a ser similar al valor que corresponde a la variable close del día anterior, si existe una gran diferencia puede ser producto de algún fenómeno poco concurrente que justifique el cambio. Otros de los indicadores es el high (que es el que describe el valor más alto que ha llegado en una sesión) y otro es el low (que describe el valor más bajo que ha alcanzado), estos índices no son tan representativos cuando se quiere entender el comportamiento de los índices en el mercado, por lo tanto, en la mayoría de los casos no son utilizados [21].

Por otro lado, el indicador de volume (nos indica la cantidad de instrumentos financieros o incluso las acciones intercambiadas en una sesión bursátil), para Calvo este es un índice bursátil que no es tan utilizado en la predicción del comportamiento futuro de acciones en el mercado, pero recomienda que sí sea incluido en los métodos de predicción, pues explica que esconde ciertas propiedades relacionadas directamente con el precio real de los índices que se deben de aprovechar.

Pone como ejemplo el hecho de un volumen alto con respecto a un precio que puede seguir cierta tendencia a la alta que indica la enorme fiabilidad del movimiento y a su vez confirma una tendencia clara. Por el otro lado un volumen alto y un precio bajo, indica que los movimientos no son nada fiables y que existe una probable reversión del valor del índice a la baja. En pocas palabras que el volumen puede llegar a ser un precursor de los cambios en cada uno de los precios, ya que permite llegar a medir la fuerza de una tendencia [21]. A continuación, una tabla descriptiva de índices bursátiles:

Tabla 1. Tabla descriptiva índices [21]

Índice	Descripción
INDU Index	Dow Jones Industrial Average (30 mayores sociedades anónimas que cotizan en el mercado bursátil de EEUU)
SPX Index	Standard & Poor's 500 (500 grandes empresas que cotizan en las bolsas NYSE o NASDAQ)
CCMP Index	NASDAQ Composite (más de 5000 empresas que cotizan en las bolsas NASDAQ)
SPTSX Index	Renta variable de Canadá
MEXBOL Index	Bolsa Mexicana de Valores
UKX Index	Bolsa de Valores de Londres (100 compañías de mayor capitalización bursátil del Reino Unido)
CAC Index	CAC 40 (40 valores más significativos de las empresas que cotizan en la Bolsa de París)
DAX Index	DAX PERFORMANCE-INDEX (30 compañías más grandes de Alemania que cotizan en la Bolsa de Fráncfort)
IBEX Index	Ibex 35 (35 empresas con más liquidez que cotizan en el SIBE)
FTSEMIB Index	FTSE MIB (350 compañías que cotizan en la Bolsa Italiana)
OMX Index	OMX Stockholm 30 (30 valores que cotizan en la Bolsa de Estocolmo)
NKY Index	Nikkei 225 (225 valores más líquidos que cotizan en la Bolsa de Tokio)
HSI Index	Hang Seng (225 valores más líquidos que cotizan en la Bolsa de Hong Kong)
SHSZ300 Index	Shanghai Shenzhen CSI 300 (300 que cotizan en las bolsas de Shanghai y Shenzhen)
AS51 Index	ASX 200 (200 compañías que cotizan en la Bolsa de Australia)
GSPG2YR Index	Spanish Govt Generic Bonds 2Yr Note (primer futuro sobre bono español a 2 años)
GSPG5YR Index	Spanish Govt Generic Bonds 5Yr Note (primer futuro sobre bono español a 5 años)
GSPG10YR Index	Spanish Govt Generic Bonds 10Yr Note (primer futuro sobre bono a español 10 años)
EUR003M Index	Euribor 3 meses
EUR006M Index	Euribor 6 meses
EUR012M Index	Euribor 12 meses
CB3 Govt	Libor 3 meses
CB6 Govt Index	Libor 6 meses
CB12 Govt Index	Libor 12 meses
BDIY Index	Baltic dry (media del precio del transporte por mar de las principales materias primas sólidas y a granel)
ECCPEMUY Index	IPC armonizado europa (índice de precios al consumo armonizado europa)
CRY Index	Thomson Reuters/CoreCommodity CRB Commodity Index

## Errores de predicción

Para Calvo [21] el que debe de ser el objetivo primordial a la hora de comparar distintos métodos de predicción es el de establecer cuál es el que ofrece el mejor valor de predicción y además el qué ofrezca los parámetros más adecuados. A continuación, se enlistan una serie de errores que según Calvo son a los que se les debe dar una mayor prioridad:

1. El permitir que se comparan entre distintas alternativas (elección de columnas, etc), esto no es más que dejar que el error no dependa de un rango en la columna elegida o, si esto no es posible lo anterior, reescalar cada uno de los datos para que así todas las columnas que tengan el mismo rango permitan una comparación.

2. El permitir comparar distintos métodos entre sí. Calvo pone como ejemplo el hacer una comparación entre coeficiente de correlación R<sup>2</sup> que es bastante útil a la hora de comparar entre distintas regresiones, pero no así a la hora de comparar con otros métodos de predicción.

3. El error siempre nos proporciona una idea de la ganancia o pérdida que se obtiene con la compra de valores reales.

## 8. Metodología

En esta sección se van a especificar los distintos tipos de investigación, su nivel, procedimientos y etapas por medio de las cuales se van a alcanzar los objetivos específicos y el general de la investigación.

La metodología de la investigación se desarrollará en 5 etapas que buscarán en primera instancia buscar las generalidades del mercado de valores sobre métodos de aprendizaje supervisado, la recolección de datos, generación de bases de datos, análisis de variables, la definición y creación de los sistemas de predicción, las pruebas, obtención de fiabilidad de datos obtenidos y finalmente la comparación de resultados obtenidos, todo ello contenido en el siguiente diagrama:

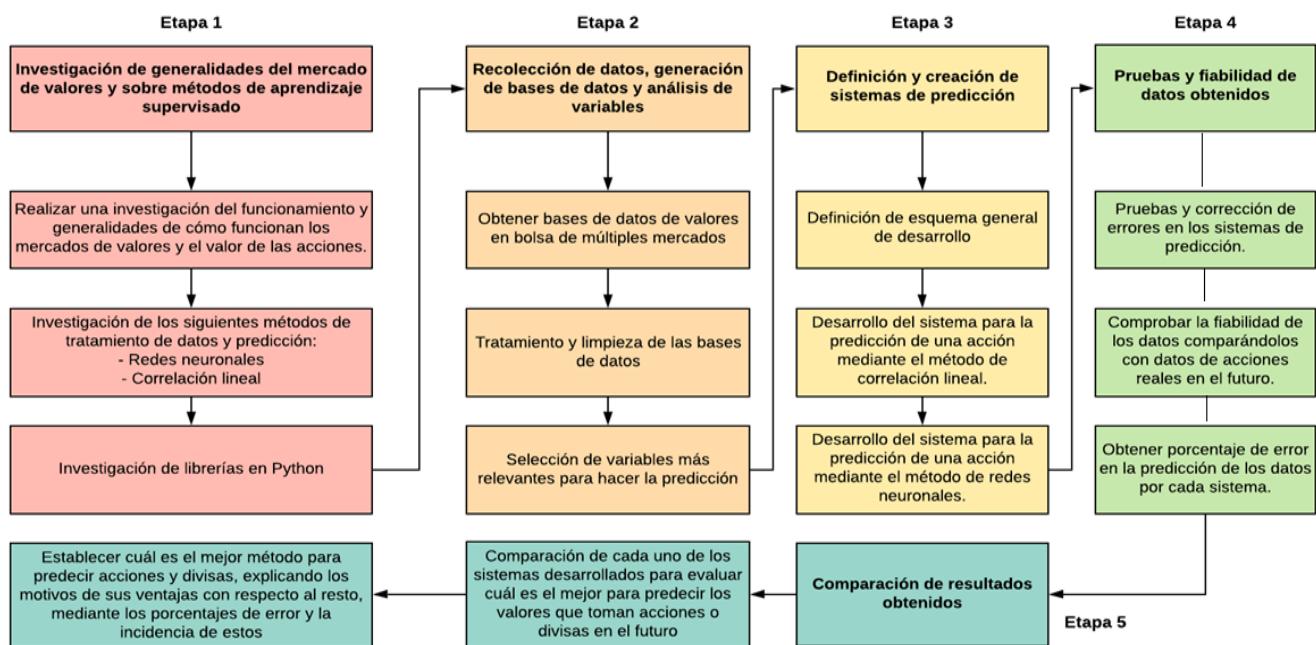


Figura 15. Etapas metodológicas

En estas etapas se hará todo el trabajo necesario para la construcción de los cuatro modelos de predicción (Dos para redes neuronales y 2 para correlación lineal) mediante los cuales se generarán distintas pruebas para posteriormente destacar cuál fue el mejor modelo de predicción para predecir el valor de una acción.

### 8.1 Tipo de investigación

El presente trabajo hará uso de distintos enfoques de investigación con el propósito de encontrar el método más eficiente para encontrar valores futuros en una acción mediante métodos de correlación lineal y redes neuronales. Por lo tanto, la investigación se abordará por distintos frentes con el objetivo de obtener una mejor calidad en los datos con los que se trabaja, así como de un conjunto de métodos eficientes. A continuación, se abordan estos tipos de enfoques en la investigación.

Por su objetivo:

## **Investigación aplicada**

Para este enfoque se busca centrar a la investigación en encontrar los métodos y variables que permitan dar respuesta a cada uno de los objetivos planteados para la investigación. Por tanto, cada uno de los puntos a desarrollar tratarán un problema en específico y por consiguiente un objetivo a resolver, enfocándose solo en ese punto sin la necesidad de ver todo el conjunto del problema.

Por su nivel de profundidad:

### **Investigación explicativa**

Esta investigación tratará de dar respuesta a las causas y consecuencias de un determinado fenómeno, que en este caso sería el de responder a cada una de las preguntas de investigación que se abordaron al inicio de este documento. Además de explicar el porqué de las cosas se tratará de llegar a ese estado en cuestión.

Esta investigación tendrá un enfoque que permita determinar un principio de causa-efecto que permitan determinar los modelos complejos que den origen a los resultados de la investigación.

Por el tipo de datos empleados la investigación será:

### **Cuantitativa**

Este tipo de investigación hará uso de un análisis de la realidad (en este caso las variables que permitan predecir el valor de una acción, así como de posibles valores de este en el futuro) esto mediante el empleo del uso de procedimientos basados en la medición.

Este enfoque permitirá obtener un mayor nivel de inferencia en la investigación, así como la de ofrecer la posibilidad de realizar experimentos con una mayor facilidad y de obtener explicaciones a partir de hipótesis en los datos.

## **8.2 Nivel de la investigación**

Esta investigación se encuentra en el nivel predictivo pues se encarga de la estimación de eventos futuros producto del valor que tomen las diversas variables dentro del mercado. Además de lo anterior se van a aplicar técnicas estadísticas específicas que permitan estimar estos valores con la mayor precisión posible.

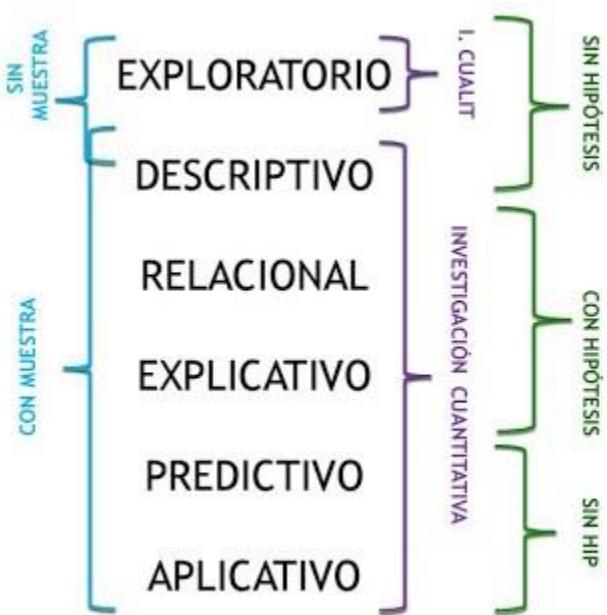


Figura 16. Niveles de Investigación

Tal como se puede apreciar la presente investigación se encuentra en este nivel, pues para obtener los datos objetivo es necesario contar con una muestra de datos bien nutrida, no se cuenta con una hipótesis central de trabajo y obedece al tipo de investigación cuantitativa.

### 8.3 Procedimientos para la investigación

Para la realización de la investigación se tomó como referencia el enfoque que da Castro en [33] donde básicamente describe al proceso de investigación como una “*serie de fases en las que el investigador se concentra en una tarea específica de la misma*”. A continuación, se detallan los procedimientos para alcanzar los objetivos de la investigación basado en el enfoque de Castro:

#### Recolección de información

Este procedimiento es vital, pues la recolección de grandes cúmulos de datos es necesaria para alimentar los sistemas y así obtener mejores resultados. Entre más y mejores datos se recolecten mejores resultados se obtendrán.

#### Procesamiento y análisis de los datos

En esta etapa se seleccionarán los datos que se analizarán y se procederá a pasarlo al sistema para que este los analice y a partir de ello obtenga una predicción.

#### Fiabilidad y validez

Una vez que los datos son seleccionados y analizados es muy importante conocer si estos son buenos para el propósito de la investigación y si estos merecen ser creídos. Esto para lograr establecer un control de calidad en la investigación.

## **8.4 Etapas para alcanzar los objetivos de la investigación**

### **Etapa 1: Investigación de generalidades del mercado de valores y sobre métodos de aprendizaje supervisado**

1.1 Realizar una investigación del funcionamiento y generalidades de cómo funcionan los mercados de valores y el valor de las acciones.

1.2 Investigación de los siguientes métodos de tratamiento de datos y predicción:

- Redes neuronales

- Correlación lineal

1.3 Investigación de librerías en Python para tratamiento de datos.

1.4 Investigación de librerías en Python para el uso de redes neuronales en la predicción de datos.

1.5 Investigación de librerías en Python para el uso de la correlación lineal en la predicción de datos.

### **Etapa 2: Recolección de datos, generación de bases de datos y análisis de variables**

2.1 Investigar fuentes de datos sobre un determinado mercado de acciones.

2.2 Obtener bases de datos de valores en bolsa de múltiples mercados.

2.3 Tratamiento y limpieza de las bases de datos

2.4 Análisis de datos.

2.5 Selección de variables más relevantes para hacer la predicción.

### **Etapa 3: Definición y creación de sistemas de predicción**

3.1 Definición de esquema general de desarrollo.

3.2 Uso y tratamiento de las bases de datos mediante librería de Python.

3.3 Definición del método principal para la predicción de datos usando correlación lineal.

3.4 Definición del método principal para la predicción de datos usando redes neuronales.

3.5 Desarrollo del sistema para la predicción de una acción mediante el método de correlación lineal.

3.6 Desarrollo del sistema para la predicción de una acción mediante el método de redes neuronales.

### **Etapa 4: Pruebas y fiabilidad de datos obtenidos**

4.1 Pruebas y corrección de errores en los sistemas de predicción.

4.2 Comprobar la validez de los datos obtenidos.

4.3 Comprobar la fiabilidad de los datos comparándolos con datos de acciones reales en el futuro.

4.4 Obtener porcentaje de error en la predicción de los datos por cada sistema.

#### **Etapa 5. Comparación de resultados obtenidos**

5.1 Comparación de cada uno de los sistemas desarrollados para evaluar cuál es el mejor para predecir los valores que toman acciones o divisas en el futuro, esto mediante tablas comparativas con respecto a datos reales y simulaciones.

5.2 Establecer cuál es el mejor método para predecir acciones y divisas, explicando los motivos de sus ventajas con respecto al resto, mediante los porcentajes de error y la incidencia de estos.

## 9. Desarrollo

### 9.1 Recolección de datos, generación de bases de datos y análisis de variables

#### 9.1.1 Selección de mercados o empresas a predecir

Para la selección de datos se tomó como referencia a Espinoza en [43] en el que menciona es posible obtener un conjunto de datos importante a partir del National Association of Securities Dealers Automated Quotations (NASDAQ) que básicamente contiene los precios históricos de alrededor de 3,0637 compañías enlistadas en esta bolsa de valores en un periodo de tiempo que abarca desde el 8 de febrero de 1971. En esta lista de datos se obtendrán una larga y considerable fuente de datos, pues está en constante actualización pues va agregando empresas nuevas año con año. La siguiente tabla muestra cómo se clasifica por industria y su valor en el mercado.

Tabla 2. Clasificación de las compañías en el NASDAQ [43]

Industria	Nº compañías	Valor en USD	Porcentaje
n/a	282	\$ 152,224,931,576.01	2%
<b>Basic Industries</b>	78	\$ 55,563,817,509.27	1%
<b>Capital Goods</b>	183	\$ 215,499,205,446.41	2%
<b>Consumer Durables</b>	83	\$ 75,447,540,058.21	1%
<b>Consumer Non-Dur.</b>	104	\$ 256,796,404,981.66	3%
<b>Consumer Services</b>	369	\$ 1,881,683,727,473.55	22%
<b>Energy</b>	80	\$ 61,624,004,753.95	1%
<b>Finance</b>	603	\$ 517,776,818,993.46	6%
<b>Healthcare</b>	593	\$ 1,329,473,781,823.79	15%
<b>Miscellaneous</b>	99	\$ 245,080,675,126.22	3%
<b>Public Utilities</b>	68	\$ 80,291,551,810.66	1%
<b>Technology</b>	465	\$ 3,644,142,785,568.86	42%
<b>Transportation</b>	56	\$ 110,369,203,845.52	1%
<b>Total</b>	3063	\$8,625,974,448,967.58	100%

Se analizará diez de las empresas con mejores cotizaciones en bolsa dentro de NASDAQ, esto porque se busca entender el comportamiento ascendente del valor de una acción en un periodo de tiempo fijo que puede variar de empresa a empresa. Las empresas por analizar se enlistan a continuación:

1. Apple
2. Amazon
3. Berkshire Hathaway
4. Facebook
5. Alphabet
6. Johnson & Johnson
7. Microsoft
8. Procter & Gamble
9. Visa Inc.
10. Walmart

Si bien estas empresas tienen cotizaciones en bolsa dispares su comportamiento en general es ascendente y por lo tanto bueno para estimar valores a la alza.

### 9.1.2 Descripción de datos a obtener

Los atributos en forma de datos que tomaremos para construir la base de datos son (fecha, apertura, cierre, alto, bajo, volumen y ajuste de cierre), esto porque son los datos disponibles en las mayorías de bases de datos que se encuentran en la red con este tipo de registros, por lo que serán las variables que el sistema analizará para obtener las predicciones. A continuación, se enlista en forma de tabla lo que cada dato significa:

Tabla 3. Descripción de los atributos [43]

Atributo	Descripción
Fecha	Nos indica el momento en el que fueron tomado los datos, este valor puede ser el año, mes, semana, día, hora y por minuto.
Apertura	Esto indica el precio apertura del día por la acción
Alto	Esto indica el precio máximo pagado durante el periodo por la acción
Bajo	Esto indica el precio mínimo pagado durante el periodo por la acción
Cierre	Esto indica el último precio pagado durante el periodo por la acción
Volumen	Esto indica el número de acciones comercializadas de la acción durante el periodo
Ajuste de cierre	Esto indica el precio ajustado incluyendo cualquier distribución o acción corporativa que se produjera en cualquier momento antes de abrir al periodo siguiente.

### 9.1.3 Preparación de los datos

El objetivo que se perfila para esta fase en el tratamiento de los datos es el de unir, transformar y manipular los datos de manera que se pueda obtener datos con una alta calidad para así obtener predicciones más certeras.

#### Datos Elegidos

La información obtenida de cada compañía fue sacada de [35] el viernes 3 de febrero del 2021, por lo que los datos estarán actualizados a este día. Los datos obtenidos por cada compañía son los siguientes:

1. Date
2. Open
3. High
4. Low
5. Close
6. Adj Close

## 7. Volume

Para consultar un ejemplo de los datos obtenidos consultar el anexo K.

### Limpieza de datos

Para la limpieza de datos se eliminaron diversos registros que estaban incompletos, pues o les faltaban varios datos o incluso no contenían ningún dato, este procedimiento de realizó para las diez bases de datos con las que se cuenta.

### Formato de datos

Para finalmente condensar los datos en un formato se decidió utilizar el formato de archivo de tipo valores separados por coma (CSV), pues es el que ofrece mayores facilidades para el trabajo con las librerías de Python.

## 9.2 Definición y creación de sistemas de predicción

### 9.2.1 Definición de esquema general de desarrollo

Para el esquema general de desarrollo para el sistema de regresión lineal se toman los siguientes criterios:

1. Uso de datos en Python con librería PANDAS, para que el sistema pueda leer los datos de un archivo CSV para leer los datos de las bases de datos.
2. Convertir los datos en un vector de una sola dimensión.
3. Dividir el conjunto de datos en el conjunto de entrenamiento y el conjunto de prueba
4. Utilizar la librería sklearn, con la siguiente sintaxis:

```
class sklearn.linear_model.LinearRegression(*, fit_intercept=True, normalize=False, copy_X=True,  
n_jobs=None, positive=False)
```

Lo que permitirá ajustar los datos a un modelo lineal con coeficientes  $w = (w_1, \dots, w_p)$  para minimizar la suma residual de cuadrados entre los objetivos del conjunto de datos y los objetivos dichos por la aproximación lineal.

5. Ajustar el modelo
6. Encontrar el error y comprobar la precisión del modelo.
7. Graficar la predicción comparando el valor real contra el valor de predicción.
8. Comparar la precisión de cada modelo de bases de datos.

Para el esquema general de desarrollo para el sistema de redes neuronales se toman los siguientes criterios:

1. Uso de datos en Python con librería PANDAS, para que el sistema pueda leer los datos de un archivo CSV para leer los datos de las bases de datos.
2. Importar el conjunto de entrenamiento
3. Obtener solo el precio de las acciones abiertas para la entrada de RNN.
4. Utilizar la normalización como función de escala.

5. Obtener las entradas y salidas
6. Restringir la entrada y la salida según el funcionamiento de LSTM.
7. Construir la RNN
8. Inicializar el RNN
9. Agregar la capa de entrada y la capa LSTM
10. Añadir la capa de salida
11. Compilar el RNN
12. Compilar todas las capas juntas.
13. Adaptar el RNN al conjunto de entrenamiento.
14. Hacer las predicciones y visualizar los resultados
15. Encontrar el error y comprobar la precisión del modelo.
16. Graficar la predicción comparando el valor real contra el valor de predicción.
17. Comparar la precisión de cada modelo de bases de datos.

### **9.2.2 Uso y tratamiento de las bases de datos mediante librería de Python**

Para el tratamiento de los datos se hará uso de la librería PANDAS que básicamente se encargará de los siguientes puntos:

- Ayudara a dar definición a nuevas estructuras de datos basadas en los arrays (arreglos) de la librería NumPy pero con funcionalidades agregadas.
- Permite leer y escribir archivos en formato CSV, que es el formato donde se encuentran las bases de datos.
- Ayuda a acceder a los datos mediante nombres de filas y columnas para guardarlos en variables para ser utilizados de forma posterior.
- Por ser una librería de uso eficiente su uso es bastante óptimo para realizar cualquier tratamiento de datos que estén en un archivo CSV.

A continuación, se presenta la manera en que se tomaran los datos del archivo CSV para guardarlos en una variable:

```
df = pd.read_csv('GOOG.csv')

df.dropna(inplace=True)

df = df.reset_index()

prices = df['Close'].tolist()

dates = df.index.tolist()
```

Este código se encarga de leer el archivo de las acciones de Google limpiándola de datos nulos con df.dropna para después guardar todos los datos en index y después guárdalos en prices y en dates.

### 9.2.3 Definición de los métodos principales para la predicción de datos usando correlación lineal

#### Consideraciones Generales

Para la creación de los métodos de correlación lineal se tomarán distintos enfoques que se ajusten a este modelo de predicción, buscando así en cada uno reducir el margen de error y aumentar la precisión en la predicción del modelo.

Para el manejo de los datos se usará el mismo enfoque tanto para limpiar como para capturar los datos en csv o mediante la obtención de datos en la web de la misma fuente mediante la librería Pandas de Python.

#### Método 1:

1: Las importaciones de librerías en su mayoría son para tratamiento de datos, graficar, y para el modelo de regresión lineal. A continuación, todas las importaciones:

```
import pandas as pd  
  
import numpy as np  
  
import matplotlib.pyplot as plt  
  
from sklearn.linear_model import LinearRegression  
  
from sklearn.model_selection import train_test_split  
  
import datetime
```

Razones de importaciones:

Pandas: Tratamiento y manejo de los datos de los archivos csv para pasarlo a un vector con el que puedan trabajar el resto de las librerías.

Numpy: Permite trabajar con vectores y matrices con un gran número de dimensiones, esto además de contener una gran colección de funciones matemáticas que permitirán por ejemplo redimensionar un vector, para que así sea leído de una forma más sencilla por el resto de las librerías en el programa.

Matplotlib.pyplot: Esta librería permitirá realizar una gráfica para tener una descripción más gráfica de los resultados.

Sklearn.linear\_model: Esta librería permitirá construir y adaptar el modelo de regresión lineal.

Sklearn.model\_selection: Permite dividir matrices o vectores en subconjuntos de datos que pueden ser aleatorios, en este caso se usarán para obtener los datos de entrenamiento.

Datetime: Para uso y manejo de fechas

2: Primero se convierten los datos de (dates, prices) a vectores de una sola dimensión, esto para que sea más sencillo.

```
dates = np.reshape(dates, (len(dates), 1))
```

```
prices = np.reshape(prices, (len(prices), 1))
```

El método np.reshape crea una nueva forma en las matrices de dates y prices y las convierte a un vector de una sola dimensión sin cambiar su contenido.

3: Una vez preparados los datos se comienza a trabajar con el modelo de regresión lineal guardándolo en la variable regressor que básicamente servirá para ajustar el marco de trabajo a un modelo lineal con coeficientes  $w = (w_1, \dots, w_p)$  que permitirá minimizar la suma residual de cuadrados entre los puntos que se observaran en el en el conjunto de datos de los vectores con lo que se trabajara y de los objetivos en la aproximación lineal. [36]

```
regressor = LinearRegression()
```

4. Después se van a dividir los datos en los siguientes conjuntos:

- a) xtrain: Datos de entrenamiento para x.
- b) xtest: Datos de prueba para x.
- c) ytrain: Datos de entrenamiento para y.
- d) ytest: Datos de prueba para y.

Esta división se hará mediante train\_test\_split que permitirá hacer esta división mediante los siguientes criterios:

test\_size: Flotante, que toma valores entre 0 y 1 y que representa la proporción del conjunto de datos para incluir en la división de prueba. [36]

random\_state: Controla la mezcla aplicada a los datos antes de aplicar la división. Pase un int para una salida reproducible a través de múltiples llamadas a funciones. [36]

Random\_state permite que todas las divisiones que se vayan generando sean reproducibles. Scikit-learn usa permutaciones aleatorias para generar todas las divisiones en el modelo. Random\_state proporciona una semilla que generara números aleatorios, lo que permitirá que los números aleatorios se generen en el mismo orden.

Tanto test\_size como random\_state cambiaron su valor en las pruebas para ajustar el modelo.

```
xtrain, xtest, ytrain, ytest = train_test_split(dates, prices, test_size=0.25, random_state=42)
```

5. Ajustar el modelo

Regressor ajusta el modelo lineal, con el uso de los siguientes parámetros:

Xtrain: Serán los datos usados para el entrenamiento.

Ytrain: Serán los valores objetivos.

```
regressor.fit(xtrain, ytrain)
```

6. A continuación, se va a predecir usando el modelo lineal con los datos de prueba de X, guardando los resultados en la variable preds:

```
preds = regressor.predict(xtest)
```

7. Después se encuentra el error y se comprueba la puntuación del modelo obtenida.

Para obtener la puntuación del modelo se usa regressor.score que regresa  $R^2$  que determina la calidad del modelo. Entre más cercano sea el valor a 1 el modelo es más preciso.

```
rms = np.sqrt(np.mean(np.power((np.array(ytest)-np.array(preds)), 2)))
```

```
print(rms)
```

```
print('Score of the Linear Regression Model', regressor.score(xtest, ytest))
```

8. Graficar los resultados obtenidos

```
plt.scatter(xtrain, ytrain, color='red', label='Precio Actual')
```

```
plt.plot(xtrain, regressor.predict(xtrain), color='blue', label='Precio predecido')
```

```
plt.title('Precio Actual VS Precio predecido')
```

```
plt.legend()
```

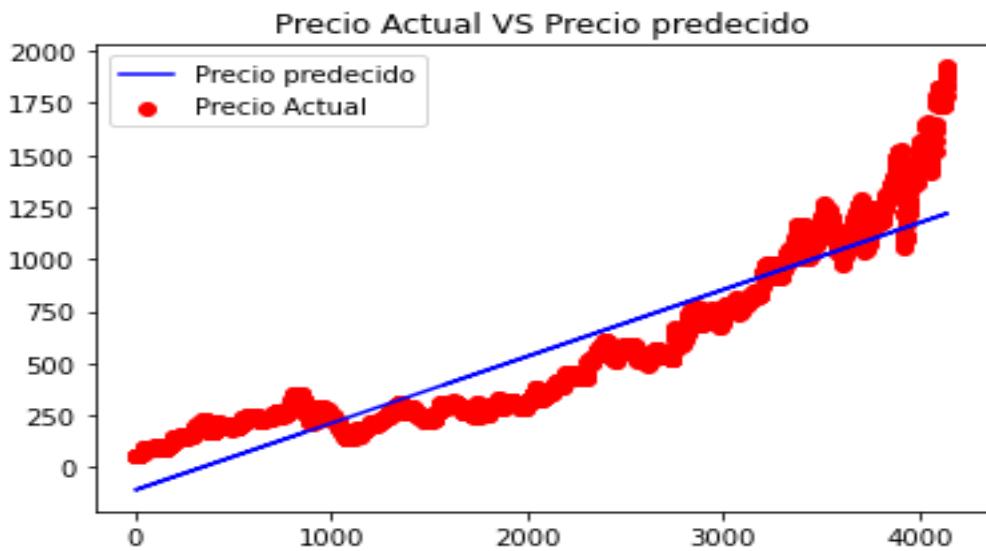
```
plt.show()
```

Una vez terminado el modelo, se obtienen los siguientes resultados para la empresa Alphabet:

163.48420765870654

Score of the Linear Regression Model 0.8466574526173501

Con la siguiente gráfica:



Gráfica 1. Predicción de acciones de Alphabet

## Método 2:

1: Las importaciones de librerías no varían demasiado con respecto a los del modelo anterior, por lo que se expondrán solo las que tienen un uso distinto. A continuación, todas las importaciones:

```

import pandas as pd

import datetime

import pandas_datareader.data as web

from pandas import Series, DataFrame

from sklearn.linear_model import LinearRegression

from sklearn import linear_model

from sklearn import model_selection

from sklearn.model_selection import train_test_split

import matplotlib.pyplot as plt

from matplotlib import style

from sklearn import preprocessing

```

Razones de importaciones:

-Pandas\_datareader.data as web: Extrae datos de varias fuentes de Internet, en este caso de Yahoo en un DataFrame de pandas. [37]

-Sklearn import preprocessing: Cambian los vectores de características sin procesar a una representación que sea más adecuada para el modelo.

- Matplotlib import style: Para diseño de la gráfica que genera el modelo.

2. A diferencia del modelo anterior los datos se obtienen de forma distinta sin dejar de ser los mismos, esto para no tener la necesidad de tener el archivo o mostrar que hay distintas formas de obtener las bases de datos ya limpias y listas para trabajar.

En esta etapa se definen las fechas de inicio y fin de los registros de acciones que se deseen analizar.

Nota: Estos datos deben corresponder a los que se tomaron para la construcción de las bases en formato csv.

```
start = datetime.datetime(2004, 8, 19)
```

```
end = datetime.datetime(2021, 2, 3)
```

3. A continuación se obtienen los datos con la librería Pandas del mismo sitio donde se obtuvieron los archivos csv. Aquí las variables start y end se toman para definir el rango de tiempo del cual se tomaron los valores de acciones de la empresa. Además, se indican las siglas de la empresa de la cual se estén tomando los datos.

Posteriormente se obtienen las primeras n filas con df.head().

```
Df = web.DataReader("GOOG", 'yahoo', start, end)
```

```
df.head()
```

4. Se obtienen los siguientes valores a partir de los datos obtenidos:

PCT\_Change: Obtiene el cambio porcentual entre la apertura y el cierre.

HL\_PCT: Obtiene el movimiento porcentual de low y high en un día.

Estos datos se agregan para tener una mayor cantidad de información y así alimentar a las bases de otras variables que puedan otorgar mayor precisión al sistema.

```
Df['HL_PCT'] = (Df['High'] - Df['Low']) / Df['Close'] * 100.0
```

```
Df['PCT_change'] = (Df['Close'] - Df['Open']) / Df['Open'] * 100.0
```

```
Df.head()
```

5: Preparar Datos:

Para preparar los datos se sigue el siguiente orden:

- Se elige la columna a predecir que será la de Adj Close.

```
forecast_col = 'Adj Close'
```

- b) Llenar los valores NAN con el último valor válido, mediante el métodofillna de Pandas.

```
df.fillna(method='ffill', inplace=True)
```

- c) Se seleccionan el número de días que se van a pronosticar

```
forecast_out = 30
```

- d) Se crea la etiqueta 'forecast\_out'

```
df['label'] = df[forecast_col].shift(-forecast_out)
```

```
df.tail()
```

```
df = df.drop(labels='Adj Close', axis=1)
```

## 6. Guardar variables del Modelo:

- a) Se definen las características de la matriz X excluyendo la columna de etiquetas que se acaba de crear.

```
X = np.array(df.drop(['label'], 1))
```

- b) Preprocesamiento de los datos en X para escalar características, es decir estandarizar los datos.

```
X = preprocessing.scale(X)
```

- c) X contendrá las últimas filas 'n = Forecast\_out' para las que no se tiene datos de etiqueta

```
X_lately = X[-forecast_out:]
```

```
X = X[:-forecast_out]
```

```
df.dropna(inplace=True)
```

- d) De manera similar se define a Y con el vector de etiqueta para los datos, que es la variable donde se hará la predicción:

```
y = np.array(df['label'])
```

## 7. Creación de conjuntos de entrenamiento y prueba

El uso de la validación cruzada básicamente mezcla los datos y, de acuerdo con el criterio test\_size se dividen los datos en datos de prueba y de entrenamiento.

- a. test\_size = 0.2 ==> indica que se tomará el 20% de los datos en el test data

```
X_train, X_test, y_train, y_test = model_selection.train_test_split(X, y, test_size=0.2)
```

- b. Ahora es el momento de utilizar la regresión lineal. Primero se dividen los datos de la siguiente manera:

1. 80% Datos de entrenamiento

## 2. 20% Datos de prueba

Luego, se usa la regresión lineal para entrenar y probar datos.

### Entrenamiento:

```
reg = linear_model.Ridge(alpha=.5)
```

```
reg.fit(X_train, y_train)
```

### Prueba:

```
confidencereg = reg.score(X_test, y_test)
```

```
print('Score of the Linear Regression Model', confidencereg)
```

## 8. Predicción

Se va a predecir usando el modelo lineal con los datos de prueba de X\_lately, guardando los resultados en la variable forecast\_set\_reg:

```
forecast_set_reg = reg.predict(X_lately)
```

-Se agrega la columna de Predicción.

```
df['Prediccion'] = np.nan
```

## 9. Luego para graficar los precios en función de las fechas se implementa el siguiente código que solo agrega fechas para los días de la predicción.

```
last_date = df.iloc[-1].name
```

```
last_unix = last_date.timestamp()
```

La siguiente línea solo indica la cantidad de segundos en un día:

```
one_day = 86400
```

```
next_unix = last_unix + one_day
```

Iteración a través del conjunto de datos para hacer que las funciones futuras no sean un número.

```
for i in forecast_set_reg:
```

```
next_date = datetime.datetime.fromtimestamp(next_unix)
```

```
next_unix += 86400
```

```
df.loc[next_date] = [np.nan for _ in range(len(df.columns)-1)]+[i]
```

```
df.tail()
```

11. Graficar los resultados obtenidos con el diseño de ggplot.

```
style.use('ggplot')

df['Close'].plot()

df['Prediccion'].plot()

plt.legend(loc=4)

plt.xlabel('Fecha')

plt.ylabel('Precio')

plt.show()
```

Una vez terminado el modelo, se obtienen los siguientes resultados para la empresa Alphabet:

Score of the Linear Regression Model 0.9818686404418769

Con la siguiente gráfica:



Gráfica 2. Predicción de acciones de Alphabet modelo 2

#### 9.2.4 Definición de los método principales para la predicción de datos usando redes neuronales

##### Consideraciones Generales

Para la creación de los métodos de redes neuronales se tomarán distintos enfoques que se ajusten a este modelo de predicción, buscando así en cada uno reducir el margen de error y aumentar la precisión en la predicción del modelo.

Para el manejo de los datos se usará el mismo enfoque que se usó en el uso de correlación lineal tanto para limpiar como para capturar los datos en csv o mediante la obtención de datos en la web de la misma fuente mediante la librería Pandas de Python.

### **Librerías Usadas en el desarrollo**

#### **-Keras:**

Keras es una librería que se usara en los modelos desarrollados para el uso de redes neuronales, está escrita en Python y capaz de ejecutarse tanto sobre Tensorflow como sobre Theano. [38] Fue usada tomando los siguientes criterios:

- Es de sencillo y rápido prototipado gracias a su buena modularidad, minimalismo y extensibilidad. [38]
- Es capaz de soportar tanto el uso de modelos de redes neuronales convolucionales como recurrentes y una combinación de ambas. [38]
- Soporta distintos esquemas de conectividad arbitrarios que son capaces de incluir el entrenamiento multi-entrada y multi-salida. [38]
- Puede ejecutarse tanto en CPU como en GPU. [38]
- Es compatible con distintas versiones de Python que van de la 2.7 a la 3.5. [38]

#### **Método 1:**

Este método hace uso de redes neuronales recurrentes junto con Long Short-Term Memory (LSTM) para predecir los precios de las acciones de las distintas empresas seleccionadas.

1: Las importaciones de librerías en su mayoría son para tratamiento de datos, graficar, y para la construcción del modelo de redes neuronales. A continuación, todas las importaciones:

```
import numpy as np  
  
import matplotlib.pyplot as plt  
  
import pandas as pd  
  
from keras.models import Sequential  
  
from keras.layers import Dense  
  
from keras.layers import LSTM
```

```
import math  
  
from sklearn.metrics import mean_squared_error
```

Razones de importaciones:

Pandas: Tratamiento y manejo de los datos de los archivos csv para pasarlo a un vector con el que puedan trabajar el resto de las librerías.

Numpy: Permite trabajar con vectores y matrices con un gran número de dimensiones, esto además de contener una gran colección de funciones matemáticas que permitirán por ejemplo redimensionar un vector, para que así sea leído de una forma más sencilla por el resto de las librerías en el programa.

Matplotlib.pyplot: Esta librería permitirá realizar una gráfica para tener una descripción más gráfica de los resultados.

Sequential: Es un modelo apropiado para hacer uso de una pila simple de capas donde cada capa tiene exactamente un tensor de entrada y un tensor de salida. [39]

Dense: Permite construir la capa densa que básicamente es la capa regular de la red neuronal profundamente conectada. Es la capa de uso más común y frecuente. Esta realiza la siguiente operación en la entrada y devuelve una salida. [40]

```
output = activation(dot(input, kernel) + bias)
```

LSTM: Las redes neuronales recurrentes del tipo memoria a corto y largo plazo (LSTM) serán usadas en este modelo pues proporcionan una alta precisión en la predicción del modelado de secuencias en varios dominios de aplicación, en este caso para la predicción de las acciones. [41]

Math: Esta función proporciona al modelo de ciertas funciones matemáticas bien definidas.

Mean\_squared\_error: Va a ayudar a medir el promedio de los cuadrados del error. Es decir, la diferencia entre los valores estimados y el valor verdadero de las acciones.

1. Importación el conjunto de entrenamiento, este solo se predice el "precio de las acciones abiertas" para la empresa extrayendo una columna.

```
training_set = pd.read_csv('GOOG_Train.csv')
```

Obteniendo solo el precio de las acciones abiertas para la entrada de RNN.

2. Convertir la forma vectorial de una sola columna en una forma de matriz, se usará 1: 2 como índice de columna. La segunda columna será ignorada y se obtiene la columna de precio de acciones abiertas en forma de matriz. La salida será una matriz de 2 dimensiones.

```
training_set = training_set.iloc[:, 1:2].values
```

3. Escala de funciones que utilizará la normalización como función de escala. El rango predeterminado para MinMaxScaler es de 0 a 1, que es lo que se quiere alcanzar. De igual forma se ajustará al juego de entrenamiento y lo escalará y reemplazará el juego original.

```
from sklearn.preprocessing import MinMaxScaler
```

```
sc = MinMaxScaler()
```

```
training_set = sc.fit_transform(training_set)
```

4. Obtención de las entradas y salidas restringiendo la entrada y la salida según el funcionamiento de LSTM.

```
X_train = training_set[0:4121]
```

```
y_train = training_set[1:4122]
```

5. Remodelación: adición de intervalo de tiempo como dimensión de entrada.

```
X_train = np.reshape(X_train, (4121, 1, 1))
```

6. Construcción del modelo RNN

7. Inicializar el modelo RNN creando un objeto de la clase Sequential para crear el RNN.

```
regressor = Sequential()
```

8. Agregar una capa de entrada y la capa LSTM usando la función de activación sigmoidea y (con ningún intervalo de tiempo y con 1 atributo como entrada)

```
regressor.add(LSTM(units = 4, activation = 'sigmoid', input_shape = (None, 1)))
```

9. Añadir la capa de salida con una neurona en la capa de salida para una salida unidimensional

```
regressor.add(Dense(units = 1))
```

10. Compilar el RNN compilando todas las capas juntas. La pérdida ayuda en la manipulación de pesos en NN.

```
regressor.compile(optimizer = 'adam', loss = 'mean_squared_error')
```

11. Adaptación del RNN al equipo de entrenamiento. El número de épocas se aumentó a lo recomendado para una mejor convergencia. Batch\_size es un parámetro de descenso de gradiente que controla la cantidad de muestras de entrenamiento para trabajar antes de que se actualicen los parámetros internos del modelo.

```
regressor.fit(X_train, y_train, batch_size = 32, epochs = 200)
```

12. Hacer las predicciones y visualizar los resultados y obtener el precio real de las acciones del 2021.

```
test_set = pd.read_csv('GOOG_Test.csv')  
  
real_stock_price = test_set.iloc[:, 1:2].values
```

13. Obtener el precio de las acciones previsto para 2021

```
inputs = real_stock_price  
  
inputs = sc.transform(inputs)  
  
inputs = np.reshape(inputs, (20, 1, 1))  
  
predicted_stock_price = regressor.predict(inputs)  
  
predicted_stock_price = sc.inverse_transform(predicted_stock_price)
```

14. Visualización de los resultados

```
plt.plot(real_stock_price, color = 'red', label = 'Precio real de las acciones')  
  
plt.plot(predicted_stock_price, color = 'blue', label = 'Precio de acción predecido')  
  
plt.title('Predicción del precio de las acciones')  
  
plt.xlabel('Tiempo')  
  
plt.ylabel('Precio de mercado')  
  
plt.legend()  
  
plt.show()
```

15. Hacer predicciones para todo el conjunto de datos, obtención del precio real de las acciones de 2004 a 2020

```
real_stock_price_train = pd.read_csv('GOOG_Train.csv')  
  
real_stock_price_train = real_stock_price_train.iloc[:, 1:2].values
```

16. Obtención del precio de las acciones previsto de 2004 a 2020

```
predicted_stock_price_train = regressor.predict(X_train)  
  
predicted_stock_price_train = sc.inverse_transform(predicted_stock_price_train)
```

17. Visualización de los resultados

```
plt.plot(real_stock_price_train, color = 'red', label = 'Precio real de las acciones')
```

```

plt.plot(predicted_stock_price_train, color = 'blue', label = 'Precio de acción previsto')

plt.title('Predicción del precio de las acciones')

plt.xlabel('Tiempo')

plt.ylabel('Precio de mercado')

plt.legend()

plt.show()

```

## 18. Evaluación de la RNN

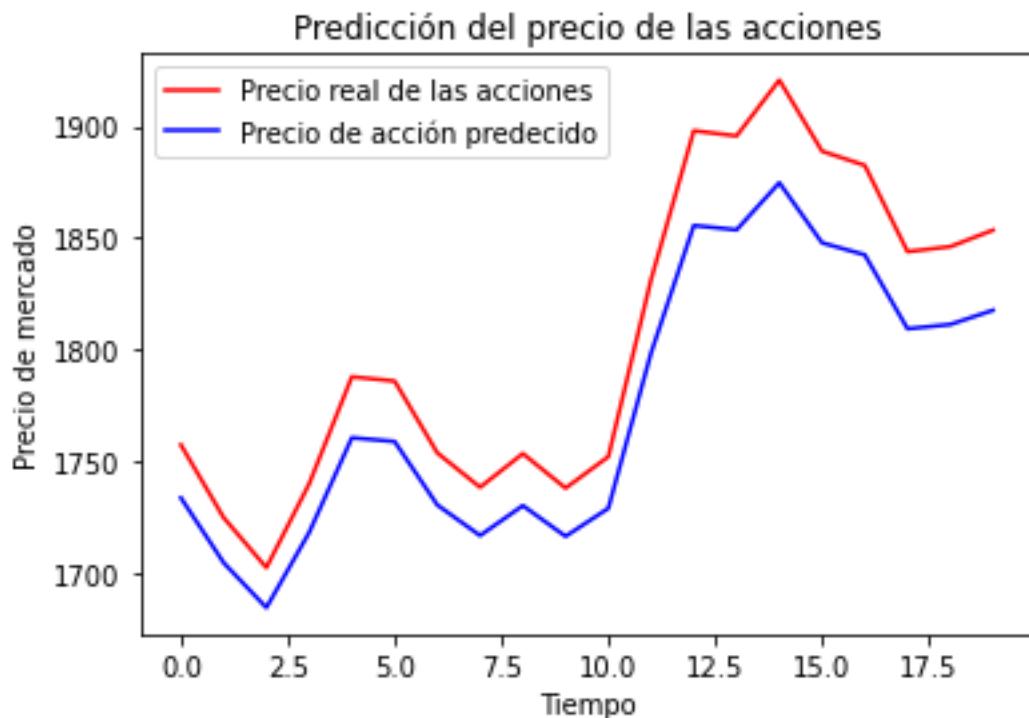
```

rmse = math.sqrt(mean_squared_error(real_stock_price, predicted_stock_price))

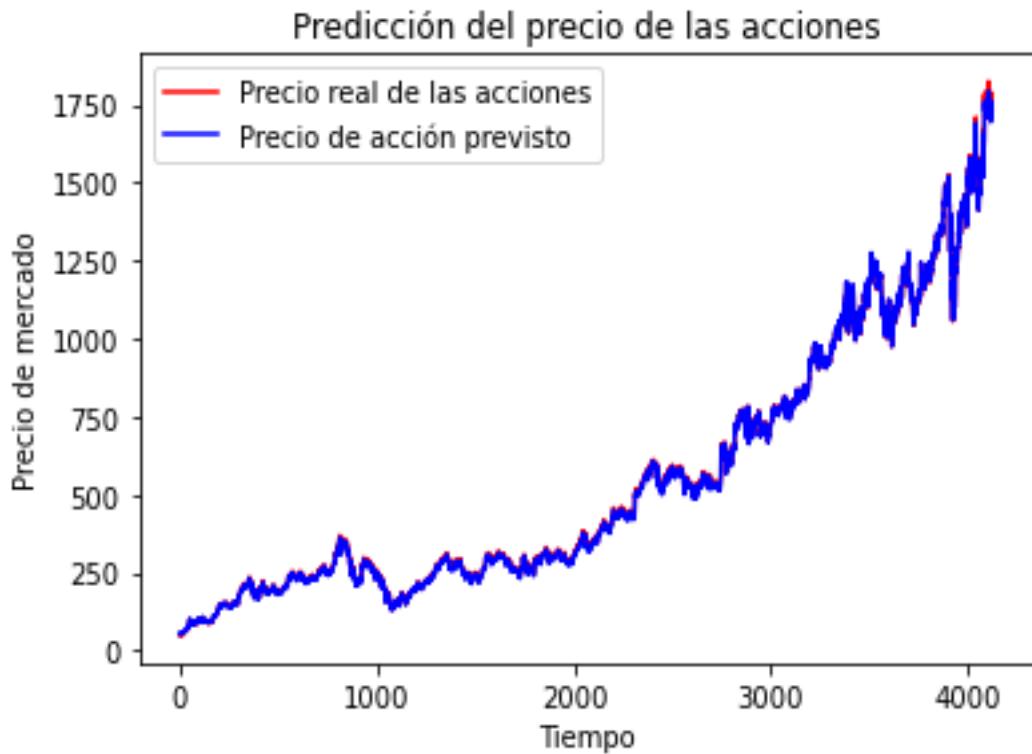
print('error:', rmse)

```

Una vez terminado el modelo, se obtienen los siguientes resultados para la empresa Alphabet:



Gráfica 3. Gráfica de resultados de modelo de redes neuronales 1



*Gráfica 4. Resultados de modelo de redes neuronales 2.*

## Método 2

El siguiente modelo está construido para hacer una predicción de OHLC que es un promedio de precios de acciones de apertura, altos, bajos. Esto usando la red neuronal recurrente LSTM.

Este modelo es empleado por que dan una enorme capacidad de reconocer y de predecir largas secuencias de datos en un periodo de tiempo extenso. Este tipo de redes basan su teoría en una especie de bucles que permiten que la salida de la red neuronal o de una parte de esta pueda servir como una entrada de la propia red neuronal en el siguiente momento. [42]

El modelo de desarrollo será el siguiente:

1: Las importaciones de librerías en su mayoría son similares a las del modelo anterior y son usadas para el tratamiento de datos, graficar, y para la construcción del modelo de redes neuronales en base a LSTM. A continuación, todas las importaciones:

```
import pandas as pd
```

```
import matplotlib.pyplot as plt
```

```
import numpy as np
```

```
import math
```

```

from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import mean_squared_error
from keras.models import Sequential
from keras.layers import Dense, Activation
from keras.layers import LSTM
import preprocessing

```

2. Se genera una semilla para convertir números aleatorios en predecibles, que después trabajara la librería numpy.

```
np.random.seed(7)
```

3. Se importan el conjunto de datos extrayendo una columna.

```

dataset = pd.read_csv('GOOG.csv', usecols=[1,2,3,4])
dataset = dataset.reindex(index = dataset.index[::-1])

```

4. Creación de un índice para obtener flexibilidad mediante la creación de matrices basadas en rangos numéricos.

```
obs = np.arange(1, len(dataset) + 1, 1)
```

5. Se capturan los distintos indicadores de predicción

```

OHLC_avg = dataset.mean(axis = 1)
HLC_avg = dataset[['High', 'Low', 'Close']].mean(axis = 1)
close_val = dataset[['Close']]

```

6. Se grafican los indicadores de ('OHLC promedio', 'HLC promedio', 'Precio de cierre')

```

plt.plot(obs, OHLC_avg, 'r', label = 'OHLC promedio')
plt.plot(obs, HLC_avg, 'b', label = 'HLC promedio')
plt.plot(obs, close_val, 'g', label = 'Precio de cierre')
plt.legend(loc = 'upper right')
plt.show()

```

7. Preparación de la base de datos

Se transforma la matriz a una sola dimensión para que pueda ser trabajada por la librería KERAS.

```
OHLC_avg = np.reshape(OHLC_avg.values, (len(OHLC_avg),1))
```

Se escalan las funciones que utilizará la normalización como función de escala. El rango predeterminado para MinMaxScaler es de 0 a 1, que es lo que se quiere alcanzar. De igual forma se ajustará al juego de entrenamiento y lo escalará y reemplazará el juego original.

```
scaler = MinMaxScaler(feature_range=(0, 1))
```

```
OHLC_avg = scaler.fit_transform(OHLC_avg)
```

8. Separar los datos de entrenamiento de los de prueba

```
train_OHLC = int(len(OHLC_avg) * 0.75)
```

```
test_OHLC = len(OHLC_avg) - train_OHLC
```

```
train_OHLC, test_OHLC = OHLC_avg[0:train_OHLC,:], OHLC_avg[train_OHLC:len(OHLC_avg),:]
```

9. Conjunto de datos (Para un tiempo T, y valores para un tiempo T +1)

Después de convertir el conjunto de datos en promedio OHLC, se convierte en datos de una columna. Esto se ha convertido en datos de series de tiempo de dos columnas, la primera columna que consta del precio de las acciones del tiempo t y la segunda columna del tiempo t + 1. Todos los valores se han normalizado entre 0 y 1.

```
trainX, trainY = preprocessing.new_dataset(train_OHLC, 1)
```

```
testX, testY = preprocessing.new_dataset(test_OHLC, 1)
```

10. Se modifican los datos de entrenamiento y de prueba.

Aquí se obtienen las entradas de entrenamiento y de prueba restringiendo la entrada según el funcionamiento de LSTM.

```
trainX = np.reshape(trainX, (trainX.shape[0], 1, trainX.shape[1]))
```

```
testX = np.reshape(testX, (testX.shape[0], 1, testX.shape[1]))
```

```
step_size = 1
```

11. Generación del modelo LSTM

```
model = Sequential()  
  
model.add(LSTM(32, input_shape=(1, step_size), return_sequences = True))  
  
model.add(LSTM(16))  
  
model.add(Dense(1))  
  
model.add(Activation('linear'))
```

12. Compilación del modelo y su entrenamiento

Se aplican dos capas LSTM secuenciales y se utiliza una capa densa para construir el modelo RNN utilizando la Keras como la librería de aprendizaje. Dado que se trata de una tarea de regresión, la activación 'lineal' se ha utilizado en la capa final.

El 75% de los datos se utilizan para entrenamiento. El optimizador Adagrad (algoritmo de gradiente adaptativo) se utiliza para una convergencia más rápida.

```
model.compile(loss='mean_squared_error', optimizer='adagrad')  
model.fit(trainX, trainY, epochs=200, batch_size=1, verbose=2)
```

### 13. Realización de la predicción

```
trainPredict = model.predict(trainX)  
  
testPredict = model.predict(testX)
```

### 14. Desnormalización de los datos para graficar

```
trainPredict = scaler.inverse_transform(trainPredict)  
  
trainY = scaler.inverse_transform([trainY])  
  
testPredict = scaler.inverse_transform(testPredict)  
  
testY = scaler.inverse_transform([testY])
```

### 15. Entrenamiento de RMSE

La métrica de precisión de la prueba es el error cuadrático medio (RMSE).

```
trainScore = math.sqrt(mean_squared_error(trainY[0], trainPredict[:,0]))  
  
print('Entrenamiento RMSE: %.2f % (trainScore))
```

### 16. Prueba de RMSE

```
testScore = math.sqrt(mean_squared_error(testY[0], testPredict[:,0]))  
  
print('Prueba RMSE: %.2f % (testScore))
```

### 17. Creación del conjunto de un conjunto de datos similares para graficar las predicciones de entrenamiento.

```
trainPredictPlot = np.empty_like(OHLC_avg)  
  
trainPredictPlot[:, :] = np.nan  
  
trainPredictPlot[step_size:len(trainPredict)+step_size, :] = trainPredict
```

### 18. Creación del conjunto de un conjunto de datos similares para graficar las predicciones de prueba.

```
testPredictPlot = np.empty_like(OHLC_avg)
```

```

testPredictPlot[:, :] = np.nan

testPredictPlot[len(trainPredict)+(step_size*2)+1:len(OHLC_avg)-1, :] = testPredict

```

19. Desnormalización del conjunto de datos principal.

```
OHLC_avg = scaler.inverse_transform(OHLC_avg)
```

20. Graficar los principales valores de OHCL, predicciones de entrenamiento y de prueba.

```

plt.plot(OHLC_avg, 'g', label = 'dataset original')

plt.plot(trainPredictPlot, 'r', label = 'conjunto de entrenamiento')

plt.plot(testPredictPlot, 'b', label = 'precio de acción de la predicción / conjunto de prueba')

plt.legend(loc = 'upper right')

plt.xlabel('Tiempo en días')

plt.ylabel('Valor OHLC de las acciones')

plt.show()

```

21. Predicción de valores futuros

```

last_val = testPredict[-1]

last_val_scaled = last_val/last_val

next_val = model.predict(np.reshape(last_val_scaled, (1,1,1)))

print("Valor del último día:", np.asscalar(last_val))

print("Valor al día siguiente:", np.asscalar(last_val*next_val))

```

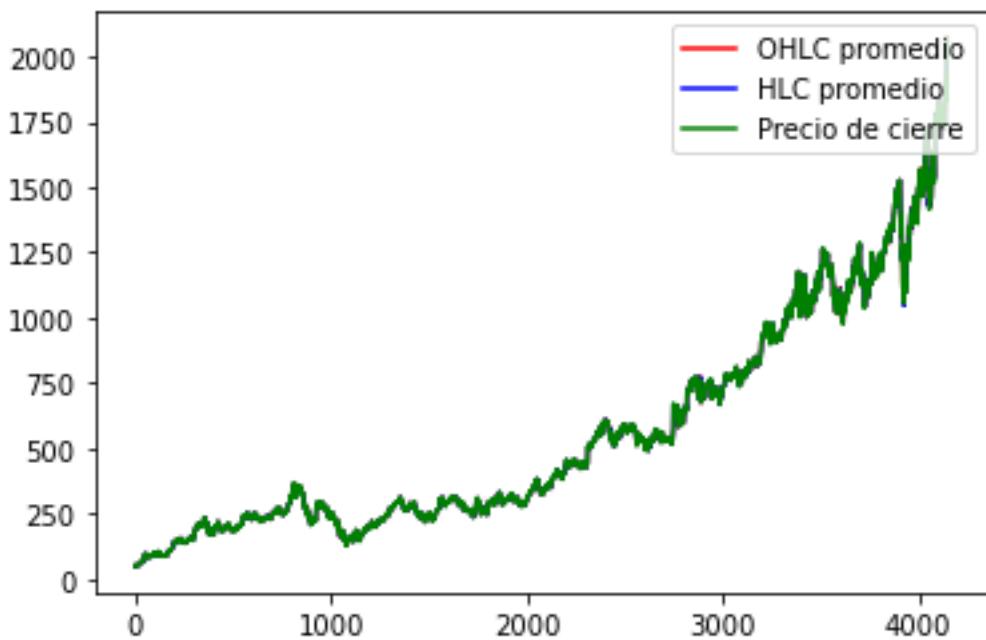
Una vez terminado el modelo, se obtienen los siguientes resultados para la empresa Alphabet:

*Entrenamiento RMSE: 9.31*

*Prueba RMSE: 77.38*

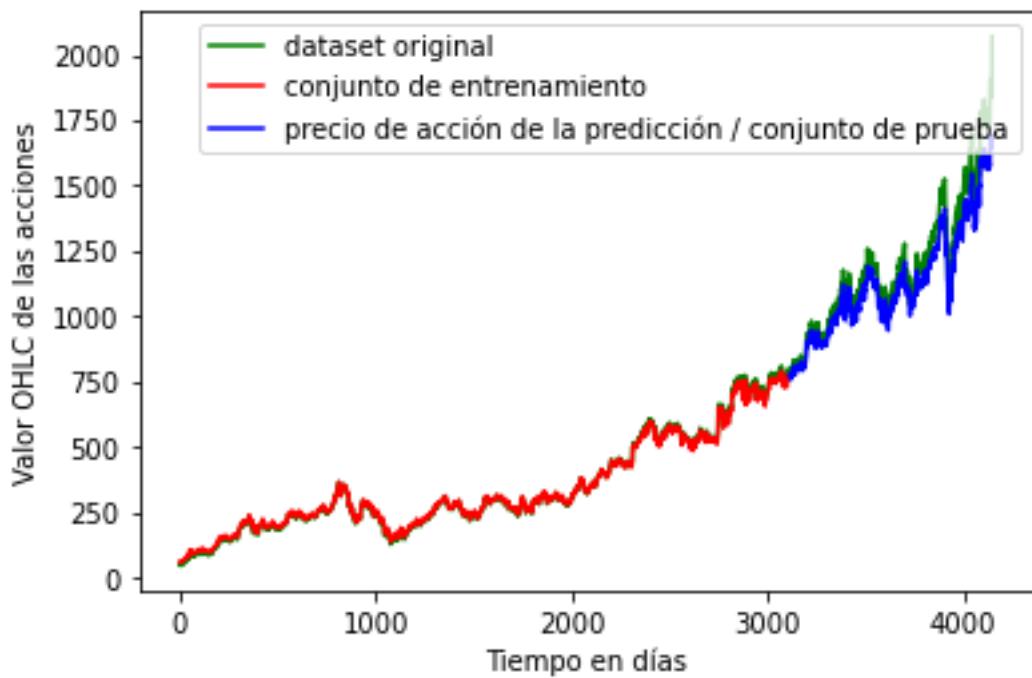
*Valor del último día: 1676.724609375*

*Valor al día siguiente: 1455.519287109375*



Gráfica 5. Datos Originales Alphabet

Después del entrenamiento y de la generación del modelo de predicción mediante LSTM con Keras, la curva es ajustada junto con el precio de las acciones originales:



Gráfica 6. Redes Neuronales mediante Keras y LSTM

## 10. Experimentación y pruebas

En esta sección se harán las distintas pruebas y ejecuciones de los cuatro modelos desarrollados para generar la mayor cantidad de escenarios posibles y a partir de ello deducir cual fue el sistema que mejores resultados obtuvo.

Las pruebas se van a realizar alternando entre los distintos parámetros correspondientes a cada modelo, asegurando los siguientes criterios de calidad:

1. Generación de diversidad
2. Tiempo de ejecución
3. Minimización de error
4. Maximización de precisión del modelo

Cada una de las pruebas desarrolladas se harán en el software Spyder haciendo uso de la versión 3.8 de Python, en un equipo con características medias de poder computacional (Intel core i5 8250U), con lo que el poder de cómputo no sea un factor determinante a la hora de construir el modelo de predicción.

Los experimentos no variarán con respecto a los modelos construidos, las únicas variables que podrán cambiar de mercado a mercado son los siguientes:

1. Base de datos históricos que va a variar de empresa a empresa.
2. Los parámetros de entrada de cada modelo (Que serán los mismos para todas las empresas, solo variando de modelo a modelo, esto para adaptarse a la naturaleza de origen del modelo).

### **10.1 Pruebas y fiabilidad de datos obtenidos**

Se realizarán un total de 120 pruebas y ejecuciones de las cuales corresponden los siguientes criterios:

- Tres pruebas de correlación lineal del modelo uno.
- Tres pruebas de correlación lineal del modelo dos.
- Tres pruebas de sistema de redes neuronales del modelo uno.
- Tres pruebas de sistema de redes neuronales del modelo dos.

Esto para cada una de las 10 empresas a analizar.

1. Apple
2. Amazon
3. Berkshire Hathaway
4. Facebook
5. Alphabet
6. Johnson & Johnson

7. Microsoft
8. Procter & Gamble
9. Visa Inc.
10. Walmart

Dando un total de 12 ejecuciones por empresa a las cuales se les asignaran diversos criterios de entrada dependiendo de los modelos analizados, esto para otorgar diversidad a los resultados.

Las ejecuciones serán anexadas en el apartado de anexos, donde se colocará tanto las gráficas y salidas que cada modelo tuvo al ser ejecutado.

A continuación, y en forma de tabla se anexan los resultados de cada modelo por empresa:

#### **Modelo de correlación lineal 1:**

Consultar el anexo L para observar las tablas con los resultados generados.

#### **Análisis de resultados:**

Este modelo da como resultado el coeficiente de determinación que básicamente es la proporción de la varianza total de la variable explicada (en este caso el de la predicción) por la regresión. El coeficiente de determinación da como resultado la calidad del ajuste de un modelo a la variable que se pretender explicar.

El resultado del coeficiente que este modelo da como resultado es un número real que oscila entre 0 y 1 y depende del grado de dependencia entre las variables si se acerca más o menos a una buena predicción. Cuanto más se acerca este valor está a 1 mayor será el ajuste y la calidad del modelo a la variable 'close' que en este caso es el que se tomó como referencia para hacer la predicción. A continuación, se enlistan los modelos de mejor a peor basado en el mejor resultado por modelo.

El coeficiente generado explica en un puntaje de (coeficiente\*100) al modelo generado. Por ejemplo, un valor de 0.93 para el coeficiente daría una asociación del 93% de la variable 'close' con respecto a la variable generada.

*Tabla 4. Resultados Correlación lineal 1*

Posición	Empresa	Mejor Resultado	Número de datos de entrada	test_size	random_state
1	Facebook	0.9362192549227766	2,192	0.25	42
2	Visa Inc.	0.8834415430908416	3,243	0.25	42
3	Walmart	0.8502902356009842	12,216	0.45	60
4	Alphabet	0.8466574526173501	4,144	0.25	42
5	Berkshire Hathaway	0.8276653121217437	10,310	0.35	50
6	Procter & Gamble	0.817961967262511	14,875	0.35	50
7	Johnson & Johnson	0.7843288053584068	14,875	0.35	50

8	Amazon	0.553508491147729	5,970	0.25	42
9	Microsoft	0.5427033518608348	8,796	0.45	60
10	Apple	0.47377626032801745	10,122	0.35	50

La empresa que mostro mejores resultados de lejos fue Facebook pues además de ser el modelo con menos datos de entrada fue la que obtuvo mejores resultados con un puntaje de 0.93 lo que indica una calidad alta asociada al modelo de predicción. Con un resultado similar Visa arrojo resultados bastante buenos pues también tiene un número bajo de datos de entrada con el que logro obtener un puntaje de 0.88. Para ambos modelos se usó como parámetros de `test_size=0.25` y `random_state=42`.

Los peores resultados se dieron con modelos asociados a un alto número de entradas, ejemplo de ello es Apple con 10,122 registros y un puntaje de 0.47 del modelo, esto se asocia a la poca dependencia que hay entre la variable dependiente 'close' y el resto de las variables. Esto es producto de registros con información con mucho ruido derivadas de datos atípicos o con mucha variabilidad entre los datos.

En cuanto a los criterios de entrada se demostró que el valor de `test_size=45` y `random_state=42` fue el mejor para generar mejores relaciones de dependencia y por tanto del puntaje del modelo. Esto tiene que ver mucho con la naturaleza de los datos de origen y el grado de dependencia entre ellos.

Este modelo de forma general se puede considerar malo para hacer predicciones, pues un porcentaje del 47% para uno de los casos deja claro que tiene una fuerte dependencia del origen de los datos y de su grado de dependencia.

### Modelo de correlación lineal 2:

Consultar el anexo M para observar los resultados correspondientes a este modelo:

### Análisis de Resultados:

Este modelo sigue una aproximación bastante similar al modelo pasado, pues utiliza la correlación lineal para obtener una predicción. A diferencia del pasado este hace uso de ajustes en el modelo que claramente da como resultados más calidad con resultados que no van a menos del 90% de precisión para todas las ejecuciones para las 10 empresas. A continuación, los mejores resultados por empresa de mejor a peor:

Tabla 5. Resultados de Correlación lineal 2

Posición	Empresa	Mejor Resultado	Número de datos de entrada	<code>test_size</code>
1	Berkshire Hathaway	0.9935796496287518	10,310	0.4
2	Amazon	0.9898938203996032	5,970	0.2
3	Visa Inc.	0.9866743272875365	3,243	0.6
4	Apple	0.9838453740255115	10,122	0.4
5	Aphabet	0.9824098581663854	4,144	0.6
6	Microsoft	0.979424802031503	8,796	0.6
7	Johnson & Johnson	0.9678352811191353	14,875	0.6
8	Facebook	0.9611791254656632	2,192	0.4

9	Procter & Gamble	0.9587024510109063	14,875	0.6
10	Walmart	0.8502902356009842	12,216	0.6

Este modelo tiene resultados muy buenos para todos los casos teniendo un 99% de grado de precisión en el mejor de los casos. En este caso el criterio de datos de entrada y el parámetro `test_size` no definió el grado de exactitud del modelo, pues el grado de variación entre cada uno de los modelos no varía mucho si estos criterios se modifican. Lo que si definió el grado de exactitud fue el grado de dependencia entre la variable dependiente 'Adj Close' y el resto de las variables, que en el peor de los casos dio un resultado de 85% por lo que se puede considerar un modelo bastante bueno para hacer predicciones de datos históricos, en este caso de predicciones.

### Modelo de Redes Neuronales 1:

Para la construcción de este modelo se tomaron solo 20 registros para las pruebas que se tomaran como referencia para hacer la predicción. Mientras que el resto de los registros son usados como datos de entrenamiento.

El único parámetro que se manipula para generar distintos escenarios para las pruebas es la variable 'Epochs' que básicamente se encarga de definir el número de veces que el modelo de la red neuronal funcionará en base a los datos de entrenamiento. Para este modelo se tomaron los siguientes valores para esta variable:

- a) 200
- b) 500
- c) 750

Esta variabilidad entre la variable se tomó así para generar escenarios lo más diversos posibles para explorar distintas soluciones para cada empresa.

Consultar anexo N para observar los resultados resumidos para este modelo.

### Análisis de Resultados:

Este modelo proporciona buenos resultados de forma general, pues a parte de tener un tiempo de ejecución relativamente corto para la mayoría de los modelos se adaptó bastante bien dado márgenes de error que oscilan en un porcentaje de menos del 10% de error. A continuación, los mejores resultados por empresa de mejor a peor:

Tabla 6. Resultados Redes Neuronales 1

Posición	Empresa	Mejor Resultado	Número de datos de entrada	Epochs
1	Procter & Gamble	0.1260535385257277	14,875	500
2	Walmart	0.24826549553799138	12,216	750
3	Visa Inc.	0.3607054663621247	3,243	750
4	Microsoft	0.5679991031777459	8,796	200
5	Facebook	0.695361994798917	2,192	750
6	Johnson & Johnson	0.774811124242707	14,875	200
7	Apple	1.3253522436824308	10,122	750

8	Amazon	4.729788087593539	5,970	500
9	Alphabet	10.598201891930568	4,144	500
10	Berkshire Hathaway	695.8421687274326	10,310	500

Como se aprecia la calidad del modelo generado no está asociado necesariamente al número de datos de entrada pues se tienen números buenos tanto a entradas grandes de datos como de pequeñas entradas. En tanto las 'Epochs' destaca que entre mayor sea este número el resultado será mucho más preciso.

El mejor resultado tuvo un error de solo el 0.12 lo que se puede decir que para esta empresa se pueden hacer predicciones bastante precisas con este modelo. Por otro lado, el peor resultado lo dio la empresa Berkshire Hathaway dio un error de 3701.2358287416564 que refleja que no es el modelo más adecuado para los datos que genera esta empresa.

Estos resultados reflejan un amplio margen de mejora para ciertos datos de entrada, este modelo deja en claro que el uso de redes LSTM que es el usado para este modelo es el de ayudar a incrementar la calidad de las predicciones solo para ciertos datos y por tanto solo para algunas empresas. La propia naturaleza de los datos de la empresa puede ser un factor restrictivo, ya que las entradas que se están utilizando pueden ser insuficientes para predecir el valor de las acciones en el tiempo.

Solo en algunos casos si se incrementa el número de 'Epochs' la calidad del modelo puede incrementarse significativamente y por tanto la calidad de la predicción.

### **Modelo de Redes Neuronales 2:**

De manera similar al modelo de redes neuronales anteriores el único parámetro que se manipula para generar distintos escenarios es la variable 'Epochs'. Para este modelo se tomaron los siguientes valores para esta variable:

- a) 200
- b) 100
- c) 50

Estas variables son tomadas para generar la diversidad necesaria para tener escenarios muy diversos en cuanto a volúmenes de entrada a la red neuronal.

Consultar anexo N para observar los resultados resumidos para este modelo.

### **Análisis de Resultados:**

Este modelo es el que más resultados de salida obtiene, pues son 4 las variables que da. A continuación que explicada cada una de ellas:

#### 1. Entrenamiento RMSE

Mide el error cuadrático medio que es generado a partir de la cantidad de error que se genera entre dos conjuntos de datos derivados del conjunto de entrenamiento que se encarga de comparar un valor que es predicho y un valor conocido.

## 2. Prueba RMSE

De forma similar mide el error cuadrático medio solo que para los datos que se toman como prueba.

## 3. Valor del último día

Valor de la variable ‘close’ del último día del conjunto de datos.

## 4. Valor al día siguiente

Valor de la variable ‘close’ del día siguiente del último día del conjunto de datos.

A continuación, los mejores resultados por empresa de mejor a peor:

Tabla 7. Resultados Redes Neuronales 2

Posición	Empresa	Número de datos de entrada	Epochs	Entrenamiento RMSE	Prueba RMSE
1	Johnson & Johnson	14,875	200	0.29	5.91
2	Microsoft	8,796	200	0.42	11.66
3	Procter & Gamble	14,875	200	0.54	2.98
4	Walmart	12,216	200	0.63	3.76
5	Visa Inc.	3,243	200	0.63	12.53
6	Apple	10,122	200	1.44	11.62
7	Facebook	2,192	200	1.93	8.20
8	Alphabet	4,144	200	20.99	143.29
9	Amazon	5,970	200	91.71	1412.56
10	Berkshire Hathaway	10,310	200	745.18	18217.11

Este modelo muestra resultados más claros y muestran el grado de dependencia que hay entre todas las variables. ‘Epochs’ muestra como la calidad de los resultados es mayor siempre cuando toma el valor de 200, lo que quiere decir que entre mayor sea el número que tome este valor la calidad por tanto será de mayor calidad.

El número de datos de entrada muestra de forma general que no es una variable determinante para que pueda afectar el valor de la predicción de los datos. Ejemplo de ello es Visa que cuenta con 3,243 datos y un RMSE para los datos de entrenamiento de 0.63, mientras que Procter & Gamble cuenta con 14,875 y un RMSE de 0.54. Por tanto, no es un factor determinante para medir la calidad que surge a partir de los datos.

De nuevo cuenta la empresa Berkshire Hathaway es la que toma los peores valores derivados de sus datos de entrada, por lo que podemos concluir que este modelo no se ajusta de la mejor manera para un modelo de redes neuronales. Mientras que el mejor resultado que se obtuvo fue de 0.29 como valor de RMSE para los datos de entrenamiento y de 5.91 para los que vienen de los que pertenecen a los de prueba, esto refleja una calidad de predicción lo suficientemente buena para hacer predicciones precisas de posibles valores futuras que tome la variable ‘close’.

## 10.2 Comparación de resultados obtenidos

Al comparar los cuatro modelos realizados podemos observar resultados bastante buenos pues en la mayoría de las empresas analizadas la precisión de la predicción se encontraba alrededor del 80%-99% de precisión. Basados en estos resultados podemos decir que de forma global los resultados apuntan a que las redes neuronales y los modelos generados a partir de la regresión múltiple tienden a rendir muy bien en la mayoría de los escenarios.

A continuación, la media de los resultados para cada modelo generado con el promedio de todos los resultados obtenidos, la moda que será obtenida a partir del rango de porcentajes que más se repita en los resultados obtenidos, con una escala que irá del 0-100% y la desviación estándar para observar qué tan dispersos están los datos con respecto a la media:

Tabla 8. Comparación de modelos

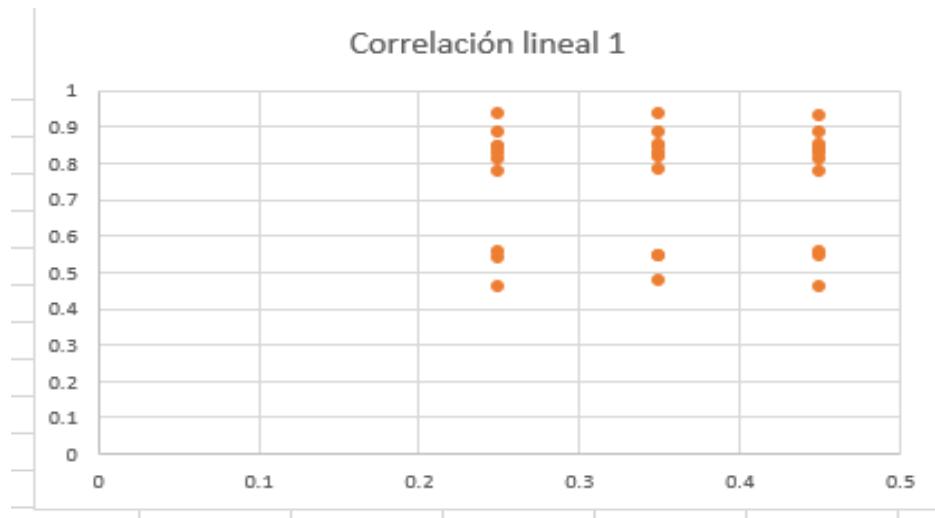
Modelo	Precisión media	Moda de precisión	Desviación Estándar
Correlación Lineal 1	74.85147029%	80%-90%	15.97128936
Correlación Lineal 2	96.43504602%	90%-100%	4.087162886
Redes Neuronales 1	-85.16212908%	90%-100%	70128.52462
Redes Neuronales 2	-25.16366667	90%-100%	32269.88516

De esta comparación se puede concluir lo siguiente:

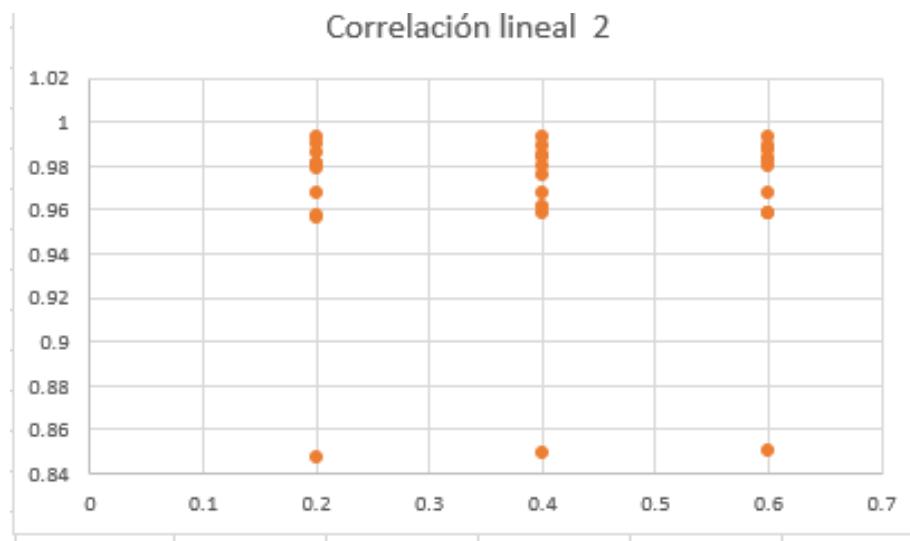
1. Los modelos de correlación tienen una mejor precisión media en sus datos debido a que sus resultados tenían menos datos atípicos.
2. Los modelos de redes neuronales obtuvieron una mala precisión media por tener datos atípicos que movían mucho el valor medio de los datos.
3. Los cuatro modelos tienen obtenido muy buenos resultados para la mayoría de los casos, tal y como se refleja en la media donde la mayoría de los resultados se encuentran en el rango de [90-100]% de precisión.
4. La desviación estándar muestra como los datos atípicos en los modelos de redes neuronales están muy dispersos con respecto a la media de los datos lo que hace que genera un valor elevado para este parámetro.
5. El modelo de correlación lineal 2 muestra una clara ventaja sobre el resto, pues es el que menos datos atípicos obtuvo a partir de la predicción, además de que el rango en el que se encuentra es alto.
6. El modelo de correlación lineal 1 fue el que peores resultados obtuvo para la mayoría de las empresas analizadas, con datos que están en el rango de [80-90]% de precisión.

A continuación, se muestra gráficos de dispersión para observar cuánto afecta una variable a otra, comparando los dos modelos de correlación lineal respectivamente y de igual forma para los de redes neuronales:

**Gráficas de dispersión de modelos de correlación lineal:**



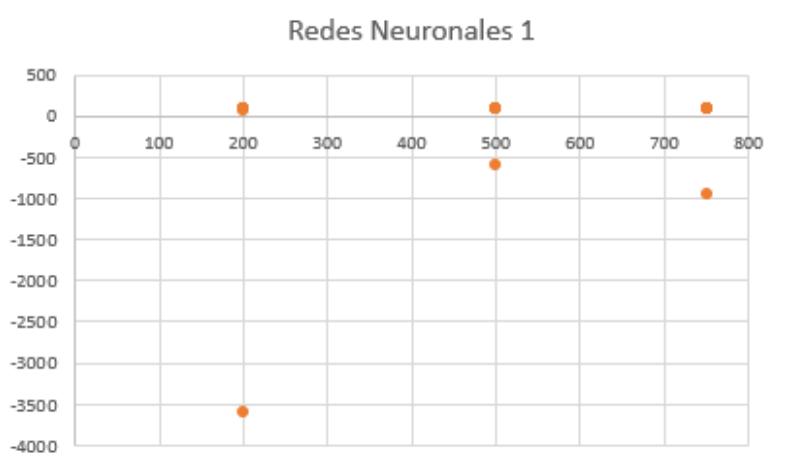
Gráfica 7. Gráfica de dispersión de modelos de correlación lineal I



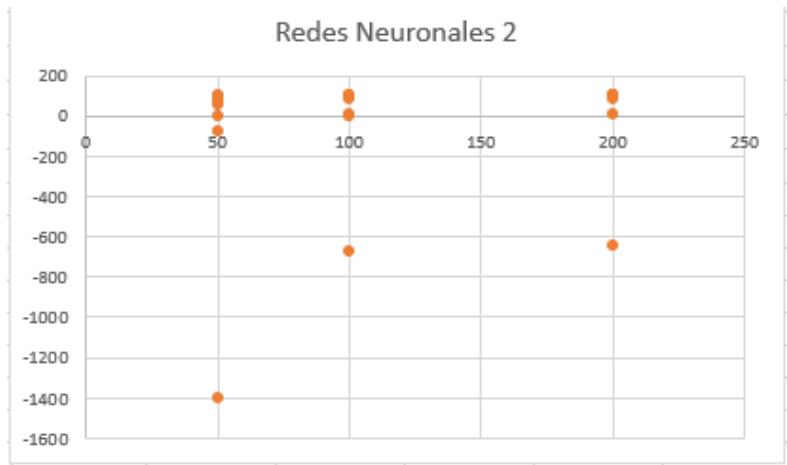
Gráfica 8. Gráfica de dispersión de modelos de correlación lineal II

A partir de ambas gráficas podemos observar que los resultados obtenidos a partir de la variable test\_size no varía mucho cuando este es cambiado, este tiende a incrementar varias decimas según se incremente este valor. Este último comportamiento se puede visualizar de mejor manera el segundo modelo de correlación que hace uso de valores de test\_size algo mayores que los que tiene el primero.

#### Gráficas de dispersión de modelos de redes neuronales:



*Gráfica 9. Gráfica de dispersión de modelos de redes neuronales I*



*Gráfica 10. Gráfica de dispersión de modelos de redes neuronales II*

De igual forma que en los modelos de correlación lineal los resultados no variaron dado el parámetro de entrada del modelo que en este caso corresponde a Epochs. Este valor solo tiende un poco más a la alza conforme se va incrementando.

Esto cambia solo en los registros con valores atípicos donde el grado de dependencia cambia el comportamiento de forma radical donde puede llegar a tener mejores valores de precisión con valores reducidos de Epochs.

Las comparaciones de todos los resultados obtenidos muestran que la tarea de predicción cuantitativa y bajo ciertas condiciones en los datos de entrada, el procedimiento clásico de regresión lineal obtuvo mejores resultados que las redes neurales debido a su estabilidad en la mayoría de las empresas analizadas.

Esto se refleja en la siguiente gráfica:



Gráfica 11. Comparación de los 4 modelos de predicción

Donde:

- Serie 1 corresponde al modelo de correlación línea 1
- Serie 2 corresponde al modelo de correlación línea 2
- Serie 3 corresponde al modelo de redes neuronales 1
- Serie 4 corresponde al modelo de redes neuronales 2

Se puede concluir dada la gráfica que:

- Los modelos de correlación lineal son buenos para generar escenarios con predicciones buenas para la mayoría de las empresas analizadas.
- Los modelos de redes neuronales fueron los que llegaron a tener una predicción mucho más precisa, pero solamente para un número limitado de las empresas analizadas, ya que este modelo no respondió bien a cierto grupo de datos de entrada.
- Ambos modelos se pueden considerar buenos para hacer predicciones precisas.
- En la mayoría de los casos incrementar el valor tanto de la variable test\_size como Epochs produce mejores resultados.

## 11. Análisis e interpretación de resultados

Para el análisis de resultados se tomará el mejor resultado y peor resultado por modelo, analizado los siguientes puntos:

1. Los resultados describen el comportamiento real de los datos en el mercado.
2. Los datos a partir de los cuales se genera el modelo benefician o perjudican el valor de la predicción.
3. El modelo empleado beneficia o afecta la precisión del modelo de predicción generado.
4. Es conveniente invertir en el mercado dado los resultados generados a partir del modelo de predicción.

Se analizarán los mejores y peores resultados pues son los que mejor describen el comportamiento real de los cuatro modelos, así se tiene un abanico real de los escenarios y con ello un análisis más completo. A continuación, el análisis de cada uno de los modelos:

### **Modelo de Correlación lineal 1:**

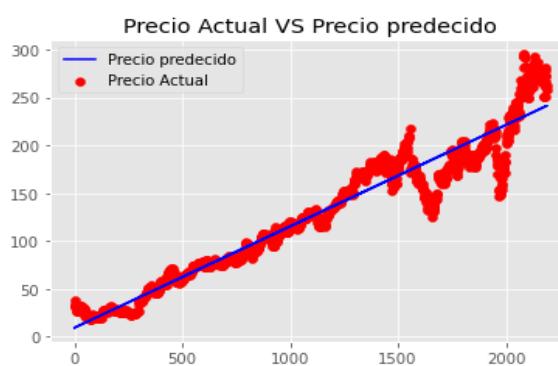
#### **a. Mejor escenario**

Empresa: Facebook

Precisión: 0.9362192549227766

Número de datos de entrada: 2,192

Gráfica:



Gráfica 12. Mejor escenario de Correlación lineal 1

#### **Los resultados describen el comportamiento real de los datos en el mercado**

En este modelo los datos describen simplemente el comportamiento hacia el cual tienden los datos, por lo tanto, no son óptimos si se quiere hacer una predicción con márgenes de precisión altos. Aunque si se toma tan solo para hacer un análisis de tendencias que pueda tomar la acción en el tiempo sin que se ajuste con gran precisión al valor real se garantizan buenos resultados.

Por ejemplo, el valor real del último día analizado para esta empresa es:

Fecha: 03/02/2021 Valor de la acción real en close: 266.649994

Fecha: 03/02/2021 Valor de la acción de la predicción en close: 159.69893247

Como se muestra en estos resultados el valor que se obtiene a partir de la predicción para este modelo no se debe tomar tal cual fuera el valor que realmente tomara la acción en ese día, pero si permite hacerse una buena idea de que si la pendiente de los datos tiene pendiente positiva o negativa y con ello saber si es conveniente o no hacer una inversión para ese mercado.

#### **Los datos a partir de los cuales se genera el modelo benefician o perjudican el valor de la predicción**

Para modelos de correlación lineal es muy sencillo observar si la naturaleza de los datos de origen afecta o no la calidad del modelo predictivo, esto porque el coeficiente que este modelo da como resultado.

Ya que si este valor se acerca a 1 mayor será el ajuste y la calidad del modelo a la variable 'close', que quiere decir que la variable dependiente close tiene una alta correlación lineal de los datos independientes

que en este caso son el resto (Open, High, Low, Adj Close, Volume) lo que dice que hay un alto grado de dependencia entre la variable dependiente y las independientes.

Para este modelo se obtuvo un valor 0.9362192549227766 lo que muestra que la misma naturaleza de los datos genera un coeficiente de correlación lineal alto y con ello una predicción que se puede llegar a considerar buena.

#### **El modelo empleado beneficia o afecta la precisión del modelo de predicción generado**

Para este caso el modelo empleado beneficio a la predicción de la acción de Facebook, pues resultó ser el que mejores resultados tuvo en comparación del resto de mercados para este modelo en concreto, además de que fue en el que mejor desempeño hubo para este mercado. De esto se puede concluir lo siguiente:

1. Los datos de origen tienen una buena dependencia entre ellos lo que genera que el modelo se ajuste bastante bien a este mercado.
2. Facebook fue el mercado que tiene un mayor coeficiente de correlación lineal simple tiene entre sus datos.
3. Este modelo solo explica tendencias en el mercado, no valores exactos o con poco margen de distancia con respecto a los reales.

#### **Es conveniente invertir en el mercado dado los resultados generados a partir del modelo de predicción**

Dados los datos generados de la predicción se puede observar una tendencia a la alta, con una pendiente positiva que va creciendo conforme pasa el tiempo. Dado este escenario este mercado muestra que hacer una inversión para este mercado no es precisamente mala idea, ya que a partir del modelo se puede concluir que hacer una predicción para obtener ganancias en el corto plazo no es rentable pues la pendiente tiende a ser horizontal. Por lo que una inversión a largo plazo sería lo más conveniente.

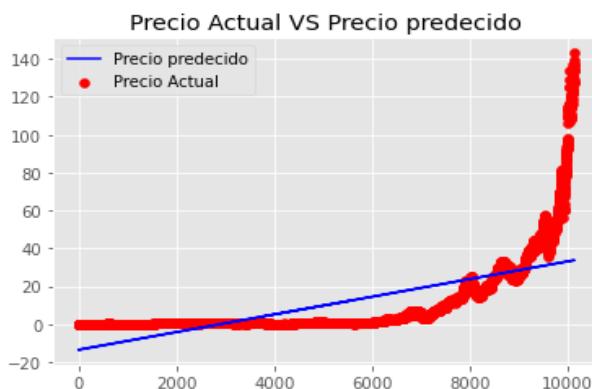
#### **b. Peor escenario**

Empresa: Apple

Precisión: 0.47377626032801745

Número de datos de entrada: 10,122

Gráfica:



Gráfica 13. Peor escenario de Correlación lineal 1

### **Los resultados describen el comportamiento real de los datos en el mercado**

Para el caso de Apple los datos tuvieron resultados bastante pobres para describir de forma correcta el comportamiento del mercado dadas las predicciones. Como se muestra en la gráfica la recta tiende a ajustarse poco al cambio brusco que existe de un momento a otro en el valor de la acción. Esto se explica en los siguientes puntos:

1. Este modelo no se ajusta correctamente a empresas con cambios grandes en el valor de la variable a predecir, tal y como es el caso de Apple.
2. Los cambios bruscos en la variable dependiente tienden a provocar un bajo coeficiente de correlación lineal.

Por lo anterior se puede concluir que este modelo no logró describir de la mejor manera el comportamiento del mercado provocado por cambios de gran amplitud de un momento a otro en la variable dependiente.

### **Los datos a partir de los cuales se genera el modelo benefician o perjudican el valor de la predicción**

Los datos de entrada para este modelo lejos de beneficiar a la predicción la perjudican pues Apple al iniciar las décadas del 2010 tuvo un crecimiento en bolsa que se puede considerar atípico para un modelo de esta naturaleza, lo que en general hace la variable close de la predicción tienda a ajustarse poco a los datos reales.

### **El modelo empleado beneficia o afecta la precisión del modelo de predicción generado**

Al ser un modelo que no tolera cambios bruscos en el valor de la variable dependiente y al ser Apple una empresa que describe este comportamiento se puede concluir que el modelo de correlación lineal 1 afecta de gran manera la calidad de la predicción generada.

### **Es conveniente invertir en el mercado dado los resultados generados a partir del modelo de predicción**

Al ser un modelo que genera valores muy pobres para esta empresa, no se puede generar una conclusión lo suficientemente buena para deducir si es conveniente o no hacer una inversión en cualquier plazo de tiempo para el modelo.

Estos escenarios se generan precisamente cuando una empresa tiende a tener una gran variabilidad en los precios de una acción en un corto periodo de tiempo.

### **Modelo de Correlación lineal 2:**

#### **a. Mejor escenario**

Empresa: Berkshire Hathaway

Precisión: 0.9935796496287518

Número de datos de entrada: 10,310

Gráfica:



Gráfica 14. Mejor escenario de Correlación lineal 2

### **Los resultados describen el comportamiento real de los datos en el mercado**

Este modelo tuvo una precisión bastante alta a la hora de construir el modelo, por lo que podemos deducir que el margen de precisión en la predicción es bastante alto tal y como se puede apreciar en los siguientes datos generados a partir del modelo de correlación lineal 2 para el mercado de Berkshire Hathaway:

Tabla 9. Comparación de resultados Correlación lineal 2 (Mejor)

Fecha	Valor real de la acción	Valor de la predicción
2021-01-12	350460.00	350822.210991
2021-01-13	353194.00	347646.704499
2021-01-14	352801.00	348988.290063
2021-01-15	350320.00	353243.078943
2021-01-16	352474.00	353686.471346

Sin bien el resultado no es preciso al máximo nivel la predicción del valor es bastante cercana a la del valor real de la acción. Por lo que la predicción generada a partir del modelo de correlación lineal 2 es bastante bueno para hacer una buena descripción del curso que tomaran el precio de las acciones para este mercado en particular.

### **Los datos a partir de los cuales se genera el modelo benefician o perjudican el valor de la predicción**

Para este modelo los datos de entrenamiento y de prueba son de gran ayuda para construir un buen modelo de predicción, pues como se puede apreciar la correlación que llegan a generar tiene un valor muy cercano a uno. Lo anterior solo refleja que para este modelo las variables independientes tienen una buena correlación lineal con respecto a la variable dependiente 'Close', lo que genera resultados muy parecidos más no iguales a los de la realidad.

### **El modelo empleado beneficia o afecta la precisión del modelo de predicción generado**

Este modelo en particular obtuvo resultados muy eficientes en la mayoría de los escenarios, y este no es la excepción, pues género en coeficiente de correlación lineal muy alto y con ello un valor en la predicción

muy cercano a uno que en este caso fue de 0.9935796496287518, que refleja que en particular para Berkshire Hathaway este modelo es el óptimo para predecir el valor de una acción.

### **Es conveniente invertir en el mercado dado los resultados generados a partir del modelo de predicción**

Dados los datos generados a partir del modelo esta empresa tiene una clara tendencia hacia la alza en sus acciones con una pendiente de crecimiento hacia arriba, por lo que se puede concluir que es conveniente hacer inversiones de corto alcance para obtener una buena inversión de capital.

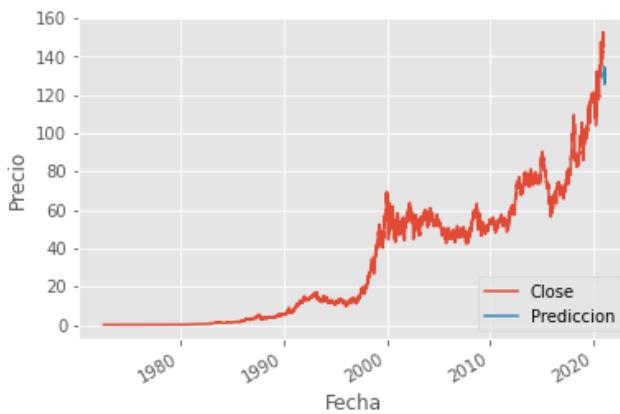
#### **b. Peor escenario**

Empresa: Walmart

Precisión: 0.8502902356009842

Número de datos de entrada: 0.6

Gráfica:



Gráfica 15. Peor escenario de Correlación lineal 2

### **Los resultados describen el comportamiento real de los datos en el mercado**

A pesar de que fue la peor precisión para este modelo, lejos de ser mala obtuvo unos buenos resultados, pues como se logra apreciar en la siguiente tabla, los valores tuvieron una buena aproximación con el valor real:

Tabla 10. Comparación de resultados Correlación lineal 2 (Peor)

Fecha	Valor real de la acción	Valor de la predicción
2021-01-12	148.35	130.458866
2021-01-13	146.84	126.276200
2021-01-14	146.36	125.823167
2021-01-15	144.04	126.415914
2021-01-16	142.79	128.379036

Si bien el modelo no está tan cercano a los datos reales, si fue capaz de describir comportamientos y tendencias en todos los casos, por lo que se puede considerar lo suficientemente bueno para lograr una buena descripción del comportamiento que tomen los valores de la acción en 'Close'.

## **Los datos a partir de los cuales se genera el modelo benefician o perjudican el valor de la predicción**

En este caso fueron los datos de entrada los que precisamente redujeron el valor del coeficiente de correlación lineal, pues como se puede apreciar en la gráfica el mercado de Walmart tiene tendencias a la alza, baja y de sostenimiento de la acción en un plazo de tiempo considerable y como vimos en el modelo pasado este tipo de comportamientos erráticos en el valor de las acciones hace que el coeficiente de correlación se vea afectado y con ello disminuya la precisión en el modelo generado.

## **El modelo empleado beneficia o afecta la precisión del modelo de predicción generado**

En este caso afecta más de lo que beneficia, pero tampoco se puede considerar como un mal modelo para hacer predicciones en este mercado en particular, pero debido a esos comportamientos poco sostenidos en los históricos del valor de 'Close' se recomienda usar otro tipo de modelo y en general para mercados con este tipo de comportamientos no lineales.

## **Es conveniente invertir en el mercado dado los resultados generados a partir del modelo de predicción**

Según los resultados obtenidos no es recomendable hacer inversiones a largo plazo para este mercado en particular, pues como se logra apreciar al tener comportamientos no lineales en el tiempo muestra como de un momento puede ir a la alza o a la baja de forma muy acelerada. Solo se recomienda hacer inversiones pequeñas para cortos periodos de tiempo en el que no se arriesgue tanto capital ni tiempo de inversión.

### **Modelo de Redes neuronales 1:**

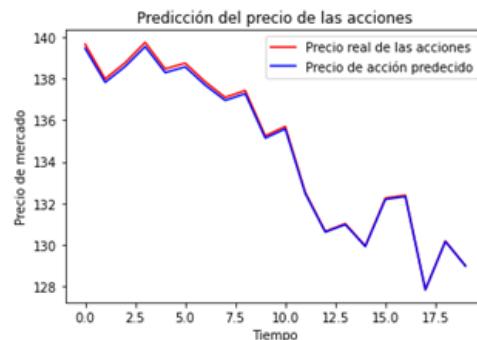
#### **a. Mejor escenario**

Empresa: Procter & Gamble

Precisión: 99.87394646%

Número de datos de entrada: 14,875

Gráfica:



Gráfica 16. Mejor escenario de redes neuronales 1

## **Los resultados describen el comportamiento real de los datos en el mercado**

Tabla 11. Comparación de resultados Redes Neuronales I (Mejor)

Fecha	Valor real de la acción	Valor de la predicción
26-01-2021	133.089996	137.25214
27-01-2021	128.380005	136.09885
28-01-2021	130.360001	137.98691
29-01-2021	128.210007	139.019
01-02-2021	128.970001	138.42133

Este modelo resultó ser uno de los más precisos, pues logra ajustarse con valores bastante cercanos a los que se tienen en el mercado real, como se logra apreciar existe una diferencia en promedio de 10 dólares entre el valor que toma la predicción y el real. Por lo que se puede tomar como una excelente referencia para conocer el comportamiento del mercado y con ello tener una aproximación a posibles valores futuros en la acción.

#### **Los datos a partir de los cuales se genera el modelo benefician o perjudican el valor de la predicción**

Este modelo tuvo una precisión en su mejor escenario de 99.87394646%, por lo que los más de 14000 registros que tuvo de origen. Esto se da porque el comportamiento que tienen los datos históricos tiene un comportamiento regular con el tiempo con un comportamiento con una complejidad casi lineal, que provoca que sea mucho más fácil de procesar por la red neuronal.

#### **El modelo empleado beneficia o afecta la precisión del modelo de predicción generado**

Este modelo en particular fue el mejor para predecir los datos de esta empresa, pues tuvo valores bastante cercanos al 100% en la precisión del modelo, arrojando valores bastante próximos a los que se obtienen en el mercado real. Esto derivado de un correcto uso de los parámetros del modelo.

Este modelo tuvo resultados bastante buenos en contraste con el tiempo de ejecución y costos computacionales, por lo que si hacemos una comparación es uno de los que mejores resultados arroja dados estos dos parámetros.

#### **Es conveniente invertir en el mercado dado los resultados generados a partir del modelo de predicción**

Este modelo tiene comportamientos que tienden estar a la alza y a la baja en períodos de tiempo relativamente estables, pero siempre con una tendencia a la alza en el tiempo, por lo anterior se recomienda hacer inversiones a largo plazo previendo que si se invierte en períodos de tiempo cortos es preferente vender la acción en un futuro cercano.

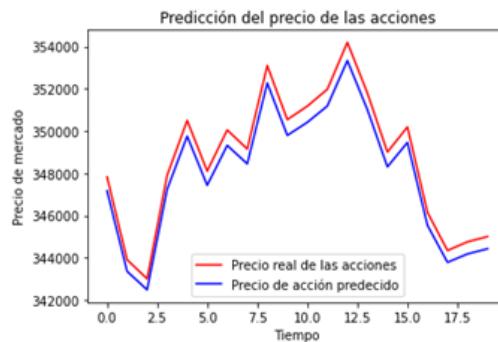
#### **b. Peor escenario**

Empresa: Berkshire Hathaway

Precisión: -595.8421687%

Número de datos de entrada: 10,310

Gráfica:



Gráfica 17. Peor escenario de redes neuronales 1

**Los resultados describen el comportamiento real de los datos en el mercado**

Tabla 12. Comparación de resultados Redes Neuronales I (Peor)

Fecha	Valor real de la acción	Valor de la predicción
26-01-2021	335850.62	349200
27-01-2021	336431.47	343450
28-01-2021	341262.84	348220
29-01-2021	344291.84	344100
01-02-2021	343585.3	346175

A pesar de que el modelo tuvo una precisión bastante mala, como se puede apreciar en la anterior tabla los resultados generados son capaces de ajustarse bastante bien al comportamiento del mercado real, con datos que se ajustan bien a las alzas y bajas del mercado, pero mal al valor específico que toma el valor de la acción al cierre de esta.

**Los datos a partir de los cuales se genera el modelo benefician o perjudican el valor de la predicción**

Los datos de esta empresa tienden a perjudicar el valor global de la precisión del modelo por tener un comportamiento algo inestable en el tiempo, como se puede apreciar en la siguiente gráfica obtenida de [43]:



*Gráfica 18. Berkshire Hathaway Acciones*

Donde se puede apreciar un comportamiento con un crecimiento muy acelerado en el tiempo para pasar de un momento a otro a una baja bastante pronunciada. Estos comportamientos erráticos afectan a toda clase de modelos de predicción y en particular para este modelo genera esos escasos valores de precisión, por lo que podemos concluir que los datos de entrenamiento si están afectando la calidad del modelo generado y con ello el valor de la predicción.

#### **El modelo empleado beneficia o afecta la precisión del modelo de predicción generado**

El modelo demostró ajustarse bastante bien al comportamiento del mercado, pero lejano al valor real de la acción, por lo que podemos decir que este modelo es excelente para ajustarse a predicciones de comportamiento en el tiempo, pero bastante malo para cuando se tienen datos de origen erráticos.

Si se está analizando una empresa con comportamientos más lineales y sin tantos cambios bruscos el resultado en la predicción será bastante bueno, pero si es lo contrario como en el presente caso, estos valores serán poco precisos.

#### **Es conveniente invertir en el mercado dado los resultados generados a partir del modelo de predicción**

Basados en los resultados obtenidos y derivado de un análisis del mercado podemos decir que, si es conveniente invertir en este mercado y a distintos plazos de tiempo, pues en cualquier escenario que se haga de predicción en el tiempo los valores para esta empresa siempre van a la alza, con bajas probabilidades de ir a la baja.

#### **Modelo de Redes neuronales 2:**

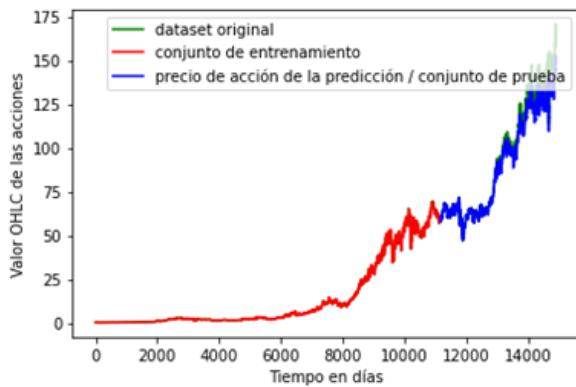
##### **a. Mejor escenario**

Empresa: Johnson & Johnson

Precisión: 99.72%

Número de datos de entrada: 14,875

Gráfica:



Gráfica 19. Mejor escenario de redes neuronales 2

### Los resultados describen el comportamiento real de los datos en el mercado

A partir del modelo generado se obtienen los siguientes datos:

Tabla 13.Comparación de resultados Redes Neuronales II (Mejor)

Fecha	Valor real de la acción	Valor de la predicción
03-02-2021	160.5	146.1570281982422
04-02-2021	160.99	128.91632080078125

Estos datos están reflejando como puede existir una gran variabilidad entre los datos reales y los pertenecientes a la predicción que se ve en un margen de tan solo dos días, esto no sucede así para periodos de tiempo más largos donde este sistema es capaz de predecir de una buena forma el comportamiento que puede tomar la acción.

### Los datos a partir de los cuales se genera el modelo benefician o perjudican el valor de la predicción

Para este caso los datos de los cuales parte el modelo benefician al sistema, pues generan un error que fue bastante bajo en contraste con otras empresas que tienen errores que superan el 100% de error. Como en modelos anteriores este comportamiento surge como respuesta a un valor de cierre que es sostenido en el tiempo a la alza con un crecimiento sostenido y semilineal que beneficia a la red neuronal LSTM para hacer la predicción en el tiempo.

### El modelo empleado beneficia o afecta la precisión del modelo de predicción generado

El modelo aporto una buena descripción del comportamiento que puede tomar una acción en un largo periodo de tiempo, no así con los resultados de la predicción con una diferencia de hasta 30 dólares.

Además de lo anterior el modelo refleja la alta sensibilidad a el cambio de parámetros y sobre todo a la gran diversidad de escenarios que puede generar a partir de la modificación de los datos de entrada.

### Es conveniente invertir en el mercado dado los resultados generados a partir del modelo de predicción

Dadas las predicciones para esta empresa y basados en el comportamiento exclusivamente que toman las predicciones para esta empresa es conveniente invertir en valores para esta empresa, pues tiene un crecimiento sostenido a la alza con crecimientos prolongados en cortos periodos de tiempo. En este escenario se recomienda hacer inversiones a mediano plazo para evitar posibles cambios bruscos en periodos de tiempo o muy largos o cortos.

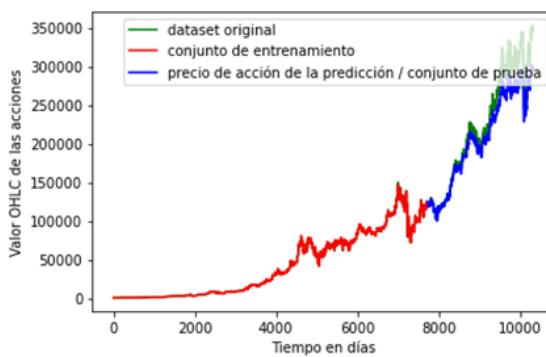
### b. Peor escenario

Empresa: Berkshire Hathaway

Precisión: -645.18%

Número de datos de entrada: 10,310

Gráfica:



Gráfica 20. Peor escenario de redes neuronales 2

### Los resultados describen el comportamiento real de los datos en el mercado

A partir del modelo generado se obtienen los siguientes datos:

Tabla 14. Comparación de resultados Redes Neuronales II (Peor)

Fecha	Valor real de la acción	Valor de la predicción
03-02-2021	350701	303950.3125
04-02-2021	354,760	265474.75

El sistema como se puede apreciar tanto en la gráfica como en los datos generados a partir de la predicción no fue capaz de hacer una buena descripción del comportamiento del mercado en el tiempo, pues a partir del día 8000 los datos empiezan a tener una gran variabilidad en la información lo que genera que el margen entre el valor de la predicción y el real se vaya incrementando, lo que genera una mala descripción del comportamiento.

### Los datos a partir de los cuales se genera el modelo benefician o perjudican el valor de la predicción

Los datos usados para este modelo evidencian la falta de linealidad y relación de dependencia entre los datos, pues lejos de beneficiar al modelo llegan a provocar errores de más del 600% de error de los datos reales con respecto a los generados por la predicción.

Esto refleja como para los sistemas de redes neuronales los datos son un verdadero protagonista que pueden o beneficiar considerablemente al valor de la acción o por el contrario afectar de sobremanera los resultados obtenidos.

#### **El modelo empleado beneficia o afecta la precisión del modelo de predicción generado**

Este modelo no es conveniente para utilizar como modelo de predicción para esta empresa, pues las cifras de error generadas a partir del solo reflejan la debilidad que tienen estos sistemas de usarse de forma generalizada, pues requieren de una análisis exhaustivo previo de si los datos con los cuales se va a trabajar son o no convenientes para ser usados por la red neuronal.

#### **Es conveniente invertir en el mercado dado los resultados generados a partir del modelo de predicción**

No es conveniente formular una hipótesis lo suficientemente confiable a partir del modelo de predicción generada para poder establecer si es o no conveniente invertir en este mercado, pues los márgenes de error son muy grandes para poder tomarlos en consideración.

#### **¿Cuál sistema de predicción es mejor para predecir valores futuros de acciones?**

Una vez analizados los resultados y comparados con los reales podemos decir de forma precisa que el método que tuvo un mejor comportamiento para la mayoría de los escenarios fue el de correlación lineal 2 pues fue el que obtuvo las siguientes ventajas:

- El modelo de correlación obtuvo la mejor precisión media en sus datos debido a que sus resultados tenían menos datos atípicos.
- Este modelo tiene una media de precisión donde la mayoría de los resultados se encuentran en el rango de [90-100].
- Este modelo fue el que tuvo una mayor estabilidad para la mayoría de las empresas analizadas.
- Fue el método que logró acercarse más al valor real de la acción en el mercado, pues como se dijo anteriormente este logra correlación lineal 2 hacer una buena descripción del curso que tomarán el precio de las acciones para los mercados.

#### **¿Cuáles fueron los mercados con una predicción más cercana al valor real de la acción?**

1. Berkshire Hathaway- correlación lineal 2
2. Procter & Gamble- redes neuronales 1
3. Walmart - correlación lineal 2

#### **¿Cuándo es correcto invertir o no en un determinado mercado?**

Antes de responder a esta pregunta se tienen las siguientes consideraciones:

-Se debe considerar que una acción está sujeta al comportamiento de una empresa y las decisiones que ésta tome, por lo que tener una predicción exacta del valor que pueda tomar sería algo muy complejo pues la cantidad de variables que entran en juego es enorme.

-De igual forma es importante saber que las acciones son un instrumento financiero con una renta variable, por lo que no siempre es posible garantizar la obtención de ingresos a partir de una inversión.

- La recuperación total o parcial de la inversión inicial va a depender de la capacidad de la empresa de poder de generar las suficientes utilidades y de la liquidez de los títulos de la acción.

-La naturaleza de la empresa bajo la cual se quiere invertir.

-Regulaciones por parte del estado o de la misma empresa con respecto a términos legales.

Una vez considerado esto se puede saber si es conveniente o no invertir en el mercado por los siguientes pasos:

1. Antes de hacer una inversión o de incluso usar un modelo de predicción es muy importante que se conozca a la empresa en la que se desea invertir, esto va desde el giro de la empresa, utilidades, activos, comportamiento en el mercado, posible futuro, etc... Una vez analizado si la empresa tiene un buen comportamiento, un buen marco de negocio y una situación financiera favorable se pasa al siguiente paso, en caso contrario se omite y de preferencia no se debe hacer la inversión.

2. Se debe hacer un estudio preliminar de inversión, investigando en foros o canales de inversión, así como si es posible con otros inversionistas sobre perspectivas y conveniencias de inversión. (Si los inversionistas la evitan o hacen inversiones bajas se deben de evitar inversiones para esa empresa). De aquí se debe hacer una lista de las empresas convenientes a inversión y se pasa al siguiente paso.

3. Una vez hecho lo anterior entra los modelos de predicción (en particular recomiendo utilizar el modelo de correlación lineal 2 o uno semejante), que son los que precisamente nos permitirán depurar la lista de empresas y obtener predicciones a futuro del curso de la empresa y con ello saber finalmente si es conveniente o no invertir en ella basado en las siguientes consideraciones:

-El modelo dice que es conveniente hacer una inversión a cualquier plazo de tiempo, pues el comportamiento va siempre a la alza con previsiones de crecimiento favorables.

-El valor de la acción que se predijo demuestra que se tendrá un retorno de inversión.

-De preferencia analizar la empresa por los cuatro modelos para tener un panorama más completo de si es conveniente o no invertir en ella.

#### **Derivado de la hipótesis de la tesina ahora es posible responder a la siguiente pregunta:**

¿Las redes neuronales y los métodos de correlación lineal son métodos que ayudan a predecir el valor de acciones de un determinado mercado, mediante el uso de minería de datos y una selección correcta de variables?

Ambos métodos son capaces de realizar predicciones al valor real de una acción en el futuro, además pueden llegar a describir el comportamiento real de una acción en el tiempo.

Esto mediante un trabajo exhaustivo de tratamiento de datos, recolección de datos, análisis de datos y un correcto uso de parámetros que pueden variar de método a método y que sirven para ajustar al modelo y alcanzar una precisión más alta.

## 12. Conclusiones

Este trabajo de investigación resultó de un arduo proceso científico en el que se abordaron distintas estrategias de minería de datos, métodos estadísticos y machine learning para lograr predecir con cierta precisión el valor que puede tomar una acción en el mercado en el futuro. Para lograr esto se realizó una amplia investigación sobre los métodos empleados para realizar este tipo de tareas y el cómo se emplean para este campo de aplicación.

En respuesta a los objetivos planteados por esta tesina podemos decir que existen métodos precisos que pueden llegar a predecir el valor que toma una acción en el futuro, el que más se acercó a un resultado eficiente fue el método de correlación lineal 2 que llegó a tener unos resultados bastante prometedores con cifras que se acercaban seriamente al resultado de la acción.

De todos los métodos desarrollados se obtuvieron resultados que se pueden considerar buenos y que pueden tomarse como referencia para realizar o bien una predicción del comportamiento a futuro de un mercado o en particular de su valor específico en un futuro cercano. De cada método desarrollado podemos concluir lo siguiente:

-Correlación lineal 1:

Para este modelo los resultados generados en su mayoría son buenos para describir el comportamiento hacia el cual tienden el mercado y no fue del todo óptimos si se quiere hacer una predicción con márgenes de precisión altos pues los valores si tienen cierto margen de diferencia. Aunque como se dijo anteriormente si este se toma tan solo para hacer un análisis de tendencias que pueda tomar la acción en el tiempo sin que se ajuste con gran precisión al valor real se garantizan buenos resultados para predecir escenarios.

- Correlación lineal 2:

Este modelo tuvo una precisión bastante buena para la mayoría de las diez empresas analizadas, pues a la hora de construir el modelo el margen de precisión en la predicción fue bastante alto y con ello su acercamiento al valor real de la acción en mercado.

Sin bien los resultados generados no son exactos, el nivel la predicción del valor de la acción es bastante cercana al del valor real de la acción. Por lo que la predicción generada a partir del modelo de correlación lineal 2 es bastante bueno para hacer una buena descripción del curso que tomarán el precio de las acciones para la mayoría de los mercados. Este modelo fue el que basado en los resultados resultó ser el mejor de los cuatro analizados, pues fue el que tuvo una mayor precisión media y el que estuvo más cercano al valor de la acción real de la acción.

-Redes Neuronales 1:

Este modelo resultó ser uno de los más precisos y eficientes para la mayoría de los escenarios analizados pues logra ajustarse con valores bastante cercanos a los que se tienen en el mercado real. Pero esto no sucedió en todos los casos analizados, pues este modelo resultó tener un error que alcanzó 3701.2358287416564 para una empresa. Esto refleja que no es el más modelo adecuado para usar de forma general, pues primero se tiene que analizar el mercado y los datos de entrada para a partir de ahí en base al error conocer si la predicción será de

calidad o no. Este modelo deja en claro que el uso de redes LSTM que es el usado para este modelo es el de ayudar a incrementar la calidad de las predicciones solo para ciertos datos y por tanto solo para algunas empresas. La propia naturaleza de los datos de la empresa puede ser un factor restrictivo, ya que las entradas que se están utilizando pueden ser insuficientes para predecir el valor de las acciones en el tiempo.

#### - Redes Neuronales 2:

El modelo de redes neuronales 2 demostró ajustarse correctamente al comportamiento derivado de los precios de la acción en los distintos mercados, pero no fue el óptimo para predecir el valor real de la acción, por lo que podemos decir que este modelo es excelente para ajustarse a predicciones de comportamiento en el tiempo tal y como sucedió con el modelo de correlación lineal 1.

De igual forma podemos agregar que los 3 mercados donde se obtuvieron predicciones más precisas dada el acercamiento con el valor real del mercado son las siguientes:

1. Berkshire Hathaway- correlación lineal 2
2. Procter & Gamble- redes neuronales 1
3. Walmart - correlación lineal 2

Como podemos ver se confirma que el modelo de correlación lineal 2 fue el que tuvo mejores resultados entre los cuatro desarrollados para esta investigación.

Con estos puntos analizados podemos concluir que se cumplieron los 3 objetivos específicos de investigación más el objetivo general de la investigación proyectado para este documento en el que se logró crear cuatro sistemas de predicción para valores de acciones los 10 mercados analizados. Mas la comparación derivada de los resultados de cada uno de los modelos.

### **¿Cómo se lograron los objetivos de investigación planteados?**

#### **Objetivo General:**

Para resolver el objetivo general se atendió a una metodología de investigación y desarrollo que baso sus ejes en los siguientes puntos:

1. Se Investigaron las generalidades del mercado de valores y de los métodos de aprendizaje supervisado.
2. Se recolectaron los datos, se generaron las bases de datos y finalmente se analizaron las variables con las que trabajaron los modelos.
3. Se definieron y crearon los cuatro sistemas de predicción (2 de redes neuronales y 2 de correlación lineal)
4. Se probaron y compararon los resultados obtenidos.

A partir de lo anterior fue posible comparar la precisión de cada modelo generado y se logró establecer a partir de la comparación con datos reales si fue o no el hacer una inversión en el mercado desarrollado.

#### **Objetivos específicos**

Para cumplir los objetivos específicos se diseñaron cuatro modelos de predicción (2 de redes neuronales y 2 de correlación lineal) en Python 3.8 utilizando distintas librerías de minería de datos, redes neuronales y de correlación lineal en los que se realizaron 120 ejecuciones, 30 por cada modelo para las 10 empresas seleccionadas en base a su cotización en bolsa en las que se hicieron 3 pruebas por mercado. Derivado de lo anterior se generaron distintos resultados que permitieron establecer cual fue el mejor modelo para predecir el valor que toma una acción en el futuro.

### **Trabajo Futuro**

Derivado de este trabajo se pretende realizar una investigación complementaria sobre posibles mejoras al modelo de correlación lineal 2 que fue el que mejores resultados obtuvo para tener predicciones mucho más cercanas al valor real de la acción en el futuro y complementándolo con otras técnicas que favorezcan el modelo.

Esto contemplando no solo con los datos históricos de las acciones, sino que agregando una gran cantidad de variables que pueden afectar el valor de una acción como las siguientes:

-Percepción de la empresa en la sociedad

Mediante análisis de sentimientos en redes sociales que generen escenarios de aceptación o rechazo del mercado y con ello el aumento o declive de su valor en el mercado.

-Movimientos de la propia empresa

Mediante técnicas de inteligencia artificial medir el impacto que tiene una decisión empresarial en el desempeño del valor de la acción del mercado.

-Decisiones políticas que afecten el desempeño del valor de la acción.

De forma similar al punto pasado de analizar el factor de riesgo y de impacto que puede tener una decisión política en el valor de una acción.

Explicación del cumplimiento de los objetivos planteados y las conclusiones sobre los resultados obtenidos y trabajo futuro.

Además de lo anterior se pretende realizar un trabajo parecido al presente solo que con el mercado de criptomonedas que puede llegar a ser incluso más complejo que el de los mercados y su cotización en bolsa.

### 13. Referencias

- [1] Espino Timón, C. (2017, enero). Análisis predictivo: técnicas y modelos utilizados y aplicaciones de este - herramientas Open Source que permiten su uso (N.o 1). Universitat Oberta de Catalunya. <http://openaccess.uoc.edu/webapps/o2/bitstream/10609/59565/6/caresptimTFG0117mem%C3%B2ria.pdf>
- [2] Uvidia Fassler, M., Cisneros Barahona, A., Méndez Naranjo, P., & Villa Yáñez, H. (2018, septiembre). Minería de datos para la toma de decisiones en la unidad de nivelación y admisión universitaria ecuatoriana (N.o 2). Universidad Nacional de Chimborazo (Ecuador). <https://dialnet.unirioja.es/descarga/articulo/6836545.pdf>
- [3] Núñez, E., W. Steyerberg, E., & Núñez, J. (2011, abril). Estrategias para la elaboración de modelos estadísticos de correlación (N.o 6). Universitat de Valencia. <https://www.revespcardiol.org/es-estrategias-elaboracion-modelos-estadisticos-regresion-articulo-S0300893211003502>
- [4] Pérez Ramis, M. (2017). Aplicaciones de la Ciencia de Datos para la predicción de la Demanda (N.o 1). Universidad de Sevilla. <http://bibing.us.es/proyectos/abreproj/91321/fichero/30263137J+-+Marta+P%C3%A9rez+Ramis.pdf>
- [5] Jurado, M., & Fellman, H. (2018, septiembre). CIENCIA DE DATOS UNA ALTERNATIVA DE ANÁLISIS AL CRECIMIENTO PEDAGÓGICO DEL ESTUDIANTE EN EDUCACIÓN SUPERIOR (N.o 2). Universidad Privada Domingo Savio. [http://www.scielo.org.bo/pdf/escepies/v5n2/v5n2\\_a06.pdf](http://www.scielo.org.bo/pdf/escepies/v5n2/v5n2_a06.pdf)
- [6] Mireles Vázquez, I. (2012, julio). Bolsa de Valores “¿Cómo? ¿Por qué? Y ¿Para qué?” (N.o 21). UAM. <http://tiempoeconomico.azc.uam.mx/wp-content/uploads/2017/07/21te4.pdf>
- [7] Salazar Coronel, A., Aceves López, J., & Valdez Pineda, D. (2010). IMPORTANCIA DE UNA INVESTIGACIÓN DE MERCADO (N.o 1). Instituto Tecnológico de Sonora. [https://www.itson.mx/publicaciones/pacioli/documents/no71/49a.-importancia\\_de\\_la\\_investigacion\\_de\\_mercado\\_nx.pdf](https://www.itson.mx/publicaciones/pacioli/documents/no71/49a.-importancia_de_la_investigacion_de_mercado_nx.pdf)
- [8] MARTINEZ MENDOZA, M. (2000, junio). COMO INVERTIR EN LA BOLSA MEXICANA DE VALORES (N.o 1). UNIVERSIDAD AUTONOMA DE NUEVO LEON. <http://eprints.uanl.mx/6460/1/1080111914.PDF>
- [9] Díaz, A. (2005). El mercado bursátil en el sistema financiero. México: McGraw-Hill.
- [10] Gitman, L. J. y Zutter, Ch. J. (2012). Principios de administración financiera (12.<sup>a</sup> ed.). México: Pearson Educación.
- [11] Herrera, C. E. (2003). Mercados financieros. México: SICCO-Gasca.
- [12] Villegas, E. y Ortega, R. M. (2003). Sistema financiero de México. México: McGraw-Hill.
- [13] Badii, M., Guillen, A., Cerna, E., Valenzuela, J., & Landeros, J. (2012, noviembre). Análisis de Regresión Lineal Simple para Predicción. UANL. [http://www.spentamexico.org/v7-n3/7\(3\)67-81.pdf](http://www.spentamexico.org/v7-n3/7(3)67-81.pdf)

- [14] Vinuesa, P. (2016, octubre). Correlación: teoría y práctica (N.o 1). UNAM.  
[https://www.ccg.unam.mx/~vinuesa/R4biosciences/docs/Tema8\\_correlacion.pdf](https://www.ccg.unam.mx/~vinuesa/R4biosciences/docs/Tema8_correlacion.pdf)
- [15] Matich, D. (2001, marzo). Redes Neuronales: Conceptos Básicos y Aplicaciones. (N.o 1). Universidad Tecnológica Nacional.  
[https://www.frro.utn.edu.ar/repositorio/catedras/quimica/5\\_anio/orientadora1/monograias/matich-redesneuronales.pdf](https://www.frro.utn.edu.ar/repositorio/catedras/quimica/5_anio/orientadora1/monograias/matich-redesneuronales.pdf)
- [16] Huertas López, A. (2015, 1 septiembre). Modelos predictivos para el mercado FOREX. Recuperado 18 de septiembre de 2020, de  
[https://www.um.es/documents/118351/2874787/TFM\\_HUERTAS+LOPEZ.pdf/a132c94c-a04d-483b-ba1c-a05521e7c132](https://www.um.es/documents/118351/2874787/TFM_HUERTAS+LOPEZ.pdf/a132c94c-a04d-483b-ba1c-a05521e7c132)
- [17] Deco Sampedro, J. D. (2012, 1 junio). Estudio y aplicación de técnicas de aprendizaje automático orientadas al ámbito médico: estimación y explicación de predicciones individuales. Recuperado 18 de septiembre de 2020, de  
[https://repository.uam.es/bitstream/handle/10486/12100/59264\\_Di\\_Decosampedro\\_JavierPFM.pdf?sequence=1&isAllowed=y](https://repository.uam.es/bitstream/handle/10486/12100/59264_Di_Decosampedro_JavierPFM.pdf?sequence=1&isAllowed=y)
- [18] FUENTES MEDINA, A. F. (2017, 6 julio). REALTIME DATA MINING APLICADO A LA PREDICCIÓN DE ÍNDICES DE BOLSA INCLUYENDO SOCIAL MEDIA ANALYTICS. Recuperado 18 de septiembre de 2020, de <https://upcommons.upc.edu/bitstream/handle/2117/112197/126952.pdf?sequence=1&isAllowed=y>
- [19] GARCÍA BERMÚDEZ, J. A., & ACEVEDO RAMIREZ, A. M. (s. f.). ANÁLISIS PARA PREDICCIÓN DE VENTAS UTILIZANDO MINERÍA DE DATOS EN ALMACENES DE VENTAS DE GRANDES SUPERFICIES. Recuperado 18 de septiembre de 2020, de  
<http://repositorio.utp.edu.co/dspace/bitstream/handle/11059/1339/006312G216.pdf>
- [20] Arango Londoño, A. (2012, 1 enero). Pronóstico del Índice General de la Bolsa de Valores de Colombia (IGBC) usando modelos de inferencia difusa.core.ac.uk. <https://core.ac.uk/download/pdf/11058494.pdf>
- [21] Calvo González, C., & Blanco Morago, R. (s. f.). Predicción de valores en mercados financieros. eprints.ucm.es. Recuperado 5 de noviembre de 2020, de [https://eprints.ucm.es/56720/1/1138328787-358439\\_Raquel\\_Blanco\\_Morago\\_TFG-Predicci%C3%B3n\\_de\\_valores\\_en\\_mercados\\_financieros\\_3940146\\_991008152.pdf](https://eprints.ucm.es/56720/1/1138328787-358439_Raquel_Blanco_Morago_TFG-Predicci%C3%B3n_de_valores_en_mercados_financieros_3940146_991008152.pdf)
- [22] Ordóñez Álvarez, J. (2017). Predicción del comportamiento de los mercados bursátiles usando redes neuronales. bibing.us.es. Recuperado 5 de noviembre de 2020, de  
<http://bibing.us.es/proyectos/abreproj/5782/fichero/PROYECTO+FIN+DE+CARRERA.+Jos%C3%A9+Manuel+Ord%C3%B3n%C3%A9z.pdf>
- [23] Larranaga, P., Inza, I., & Moujahid, A. (s. f.). Redes Neuronales.sc.ehu.es. Recuperado 5 de noviembre de 2020, de <http://www.sc.ehu.es/ccwbayes/docencia/mmcc/docs/t8neuronales.pdf>

- [24] LARA ROSANO, F. (s. f.). FUNDAMENTOS DE REDES NEURONALES ARTIFICIALES. <http://conceptos.sociales.unam.mx/>. Recuperado 5 de noviembre de 2020, de [http://conceptos.sociales.unam.mx/conceptos\\_final/598trabajo.pdf](http://conceptos.sociales.unam.mx/conceptos_final/598trabajo.pdf)
- [25] Areli-Toral Barrera, J. (s. f.). Redes Neuronales. cucei.udg.mx. Recuperado 5 de noviembre de 2020, de [http://www.cucei.udg.mx/sites/default/files/pdf/toral\\_jamie\\_areli.pdf](http://www.cucei.udg.mx/sites/default/files/pdf/toral_jamie_areli.pdf)
- [26] Berzal, F. (s. f.). Redes Neuronales. elvex.ugr.es. Recuperado 6 de noviembre de 2020, de <https://elvex.ugr.es/decsai/computationalintelligence/slides/N1%20Neural%20Networks.pdf>
- [27] FALLAS, J. (2012). CORRELACIÓN LINEAL. ucipfg.com. Recuperado 5 de noviembre de 2020, de [https://www.ucipfg.com/Repositorio/MGAP/MGAP05/BLOQUE-ACADEMICO/Unidad2/complementarias/correlacion\\_lineal\\_2012.pdf](https://www.ucipfg.com/Repositorio/MGAP/MGAP05/BLOQUE-ACADEMICO/Unidad2/complementarias/correlacion_lineal_2012.pdf)
- [28] UNAL. (s. f.). Modelos de regresión. bdigital.unal.edu.co. Recuperado 5 de noviembre de 2020, de [http://bdigital.unal.edu.co/46252/14/9789583393193\\_Part03.PDF](http://bdigital.unal.edu.co/46252/14/9789583393193_Part03.PDF)
- [29] Novales, A. (2010, 20 septiembre). Análisis de Regresión. ucm.es. <https://www.ucm.es/data/cont/docs/518-2013-11-13-Analisis%20de%20Regresion.pdf>
- [30] Martínez García, M. D., & Salazar Hernández, B. C. (2019, 1 junio). Análisis de regresión y correlación lineal. ueah.edu.mx. [https://www.uaeh.edu.mx/docencia/P\\_Presentaciones/icea/asignatura/administracion/2019/analisis-regresion-administracion.pdf](https://www.uaeh.edu.mx/docencia/P_Presentaciones/icea/asignatura/administracion/2019/analisis-regresion-administracion.pdf)
- [31] Ballesteros Román, A., Sánchez-Guzmán, D., & García, R. (2013, noviembre). Minería de datos educativa: Una herramienta para la investigación de patrones de aprendizaje sobre un contexto educativo (N.o 1). IPN. [http://www.lajpe.org/dec13/22-LAJPE\\_814\\_bis\\_Alejandro\\_Ballesteros.pdf](http://www.lajpe.org/dec13/22-LAJPE_814_bis_Alejandro_Ballesteros.pdf)
- [32] Hernández, R., Fernández, C., y Baptista, M.P. (2010) Metodología de la Investigación (5<sup>a</sup> Ed.). México: McGraw Hill Educación.
- [33] Castro, E., & Castro, E. (s. f.). EL PROCESO DE INVESTIGACIÓN. UN EJEMPLO. uv.es. Recuperado 4 de febrero de 2021, de <https://www.uv.es/Angel.Gutierrez/aprengeom/archivos2/homenaje/06CastroE.PDF>
- [34] Espinosa Muñoz, J. I. (2015, julio). Una aproximación a la predicción del valor de acciones en la bolsa de Valores aplicando técnicas de Data Mining (N.o 1). Escuela Técnica Superior de Ingenieros Informáticos. [http://oa.upm.es/37259/1/TM\\_ESPINOSA\\_MU%C3%91OZ\\_JAVIER.pdf](http://oa.upm.es/37259/1/TM_ESPINOSA_MU%C3%91OZ_JAVIER.pdf)
- [35] Yahoo. (s. f.). Yahoo Finanzas. finance.yahoo.com. Recuperado 9 de febrero de 2021, de <https://es.finance.yahoo.com/>
- [36] Pedregosa, F. (s. f.). sklearn.linear\_model.LinearRegression — scikit-learn 0.24.1 documentation. scikit-learn.org. Recuperado 15 de febrero de 2020, de [https://scikit-learn.org/stable/modules/generated/sklearn.linear\\_model.LinearRegression.html](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html)

- [37] Spencer, A. (2020, 7 julio). pandas-datareader — pandas-datareader 0.9.0rc1+2.g427f658 documentation. pandas-datareader.readthedocs.io. <https://pandas-datareader.readthedocs.io/en/latest/index.html>
- [38] Antona Cortés, C. (2017, enero). HERRAMIENTAS MODERNAS EN REDES NEURONALES: LA LIBRERÍA KERAS. Universidad Autónoma de Madrid. [https://repositorio.uam.es/bitstream/handle/10486/677854/antona\\_cortes\\_carlos\\_tfg.pdf?sequence=1](https://repositorio.uam.es/bitstream/handle/10486/677854/antona_cortes_carlos_tfg.pdf?sequence=1)
- [39] Team, K. (2020, 4 diciembre). Keras documentation: The Sequential model. keras.io. [https://keras.io/guides/sequential\\_model/](https://keras.io/guides/sequential_model/)
- [40] Team, K. (s. f.). Keras documentation: Dense layer. keras.io. Recuperado 24 de febrero de 2021, de [https://keras.io/api/layers/core\\_layers/dense/](https://keras.io/api/layers/core_layers/dense/)
- [41] Alcocer, U. M. R. (2018, 29 noviembre). MODELO BASADO EN REDES NEURONALES RECURRENTES LSTM PARA LA PREDICCIÓN DE LA SIGUIENTE ACTIVIDAD EN PROCESOS DE NEGOCIO (LSTM RECURRENT NEURAL NETWORK BASED-MODEL FOR THE PREDICTION OF THE NEXT ACTIVITY IN BUSINESS PROCESSES) | Ramírez Alcocer | Pistas Educativas. http://itcelaya.edu.mx/. <http://itcelaya.edu.mx/ojs/index.php/pistas/article/view/1736>
- [42] Mañas, A. M. (s. f.). Métodos basados en Deep Learning | Notas sobre pronóstico del flujo de tráfico en la ciudad de Madrid. bookdown.org. Recuperado 2 de marzo de 2021, de <https://bookdown.org/amanas/traficomadrid/m%C3%A9todos-basados-en-deep-learning.html>
- [43] Yahoo Finanzas. (2021, 7 mayo). Berkshire Hathaway Inc. (BRK-A). finanzas.yahoo.com.

## 14. Anexos

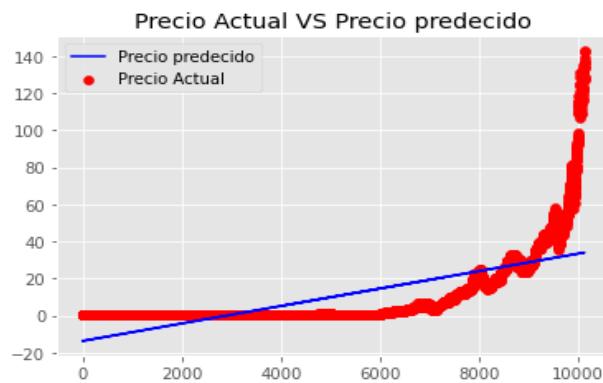
### 14.1 Anexo A. Pruebas con empresa Apple

#### -Prueba con programa de Correlación Lineal 1:

Test 1:

Atributos: test\_size=0.25, random\_state=42

**Resultados:** Score of the Linear Regression Model 0.4574994775948822

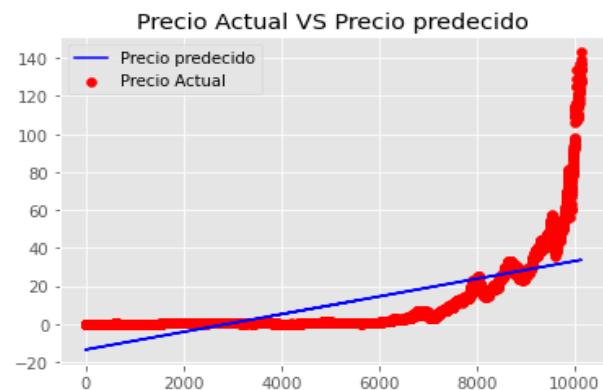


Gráfica 21. Test 1-Correlación lineal 1-Apple

Test 2:

Atributos: test\_size=0.35, random\_state=50

**Resultados:** Score of the Linear Regression Model 0.47377626032801745

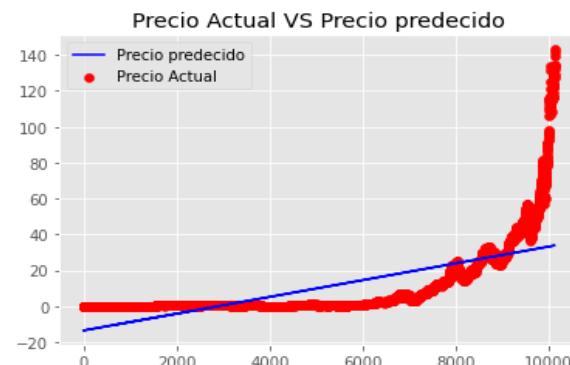


Gráfica 22. Test 2-Correlación lineal 1-Apple

Test 3:

Atributos: test\_size=0.45, random\_state=60

**Resultados:** Score of the Linear Regression Model 0.4577790125994289



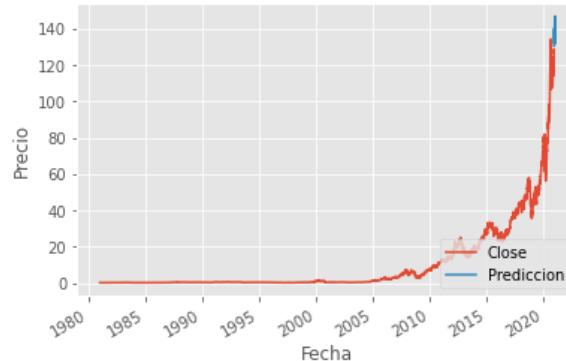
Gráfica 23. Test 3-Correlación lineal 1-Apple

-Prueba con programa de Correlación Lineal 2:

Test 1:

Atributos: test\_size=0.2

**Resultados:** Score of the Linear Regression Model 0.9808247323817118

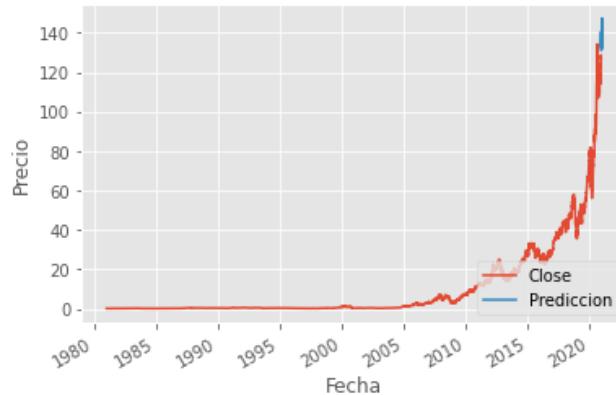


Gráfica 24. Test 1-Correlación lineal 2-Apple

Test 2:

Atributos: test\_size=0.4

**Resultados:** Score of the Linear Regression Model 0.9838453740255115

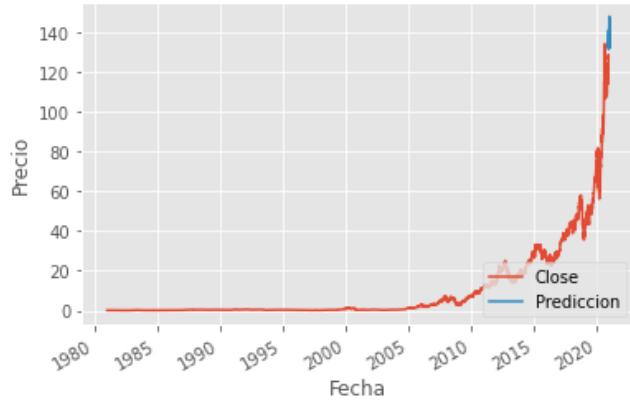


Gráfica 25. Test 2-Correlación lineal 2-Apple

Test 3:

Atributos: test\_size=0.6

**Resultados:** Score of the Linear Regression Model 0.9829777290335552



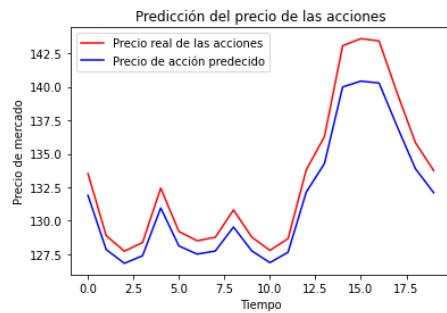
Gráfica 26. Test 3-Correlación lineal 2-Apple

#### -Prueba con programa de Redes Neuronales 1:

Test 1:

Atributos: epochs = 200

Resultados: error: 1.7961129347195295

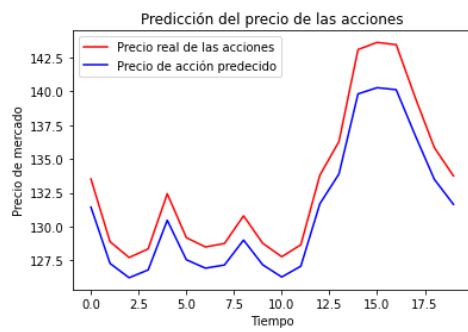


Gráfica 27. Test 1-Redes Neuronales 1-Apple

Test 2:

Atributos: epochs = 500

Resultados: error: 2.1701886389070557

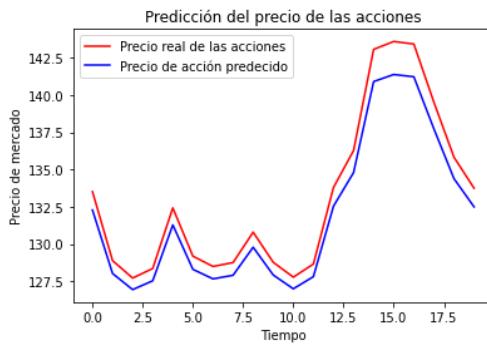


Gráfica 28. Test 2-Redes Neuronales 1-Apple

Test 3:

Atributos: epochs = 750

Resultados: error: 1.3253522436824308



Gráfica 29. Test 3-Redes Neuronales 1-Apple

**-Prueba con programa de Redes Neuronales 2:**

Test 1:

Atributos: epochs = 200

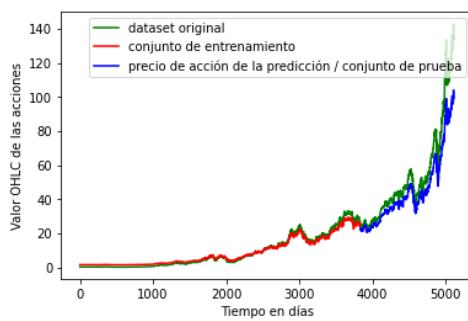
Resultados:

Entrenamiento RMSE: 1.44

Prueba RMSE: 11.62

Valor del último día: 99.15734100341797

Valor al día siguiente: 72.0592041015625



Gráfica 30. Test 1-Redes Neuronales 2-Apple

Test 2:

Atributos: epochs = 100

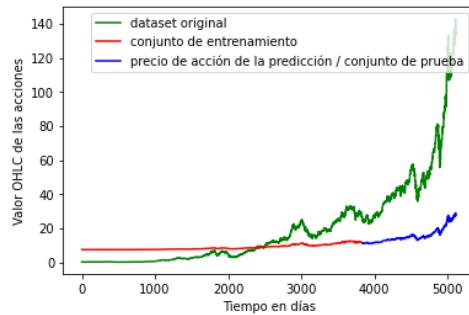
Resultados:

Entrenamiento RMSE: 7.94

Prueba RMSE: 44.25

Valor del último día: 27.780553817749023

Valor al día siguiente: 5.622982978820801



Gráfica 31. Test 2-Redes Neuronales 2-Apple

Test 3:

Atributos: epochs = 50

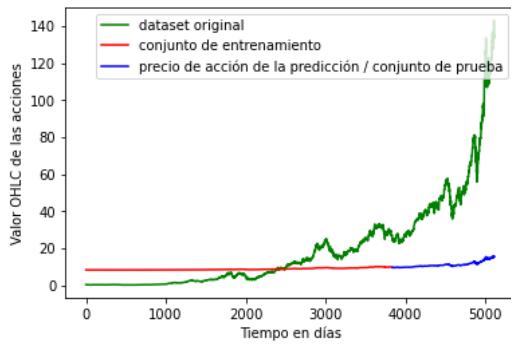
Resultados:

Entrenamiento RMSE: 8.87

Prueba RMSE: 49.49

Valor del último día: 15.459165573120117

Valor al día siguiente: 1.7005761861801147



Gráfica 32. Test 3-Redes Neuronales 2-Apple

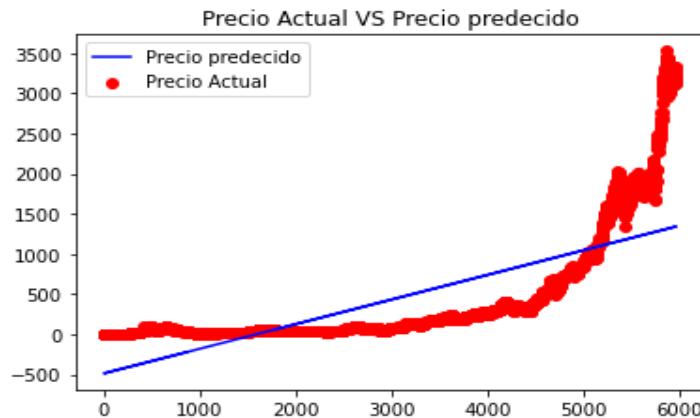
## 14.2 Anexo B. Pruebas con empresa Amazon

### -Prueba con programa de Correlación Lineal 1:

Test 1:

Atributos: test\_size=0.25, random\_state=42

**Resultados:** Score of the Linear Regression Model 0.553508491147729

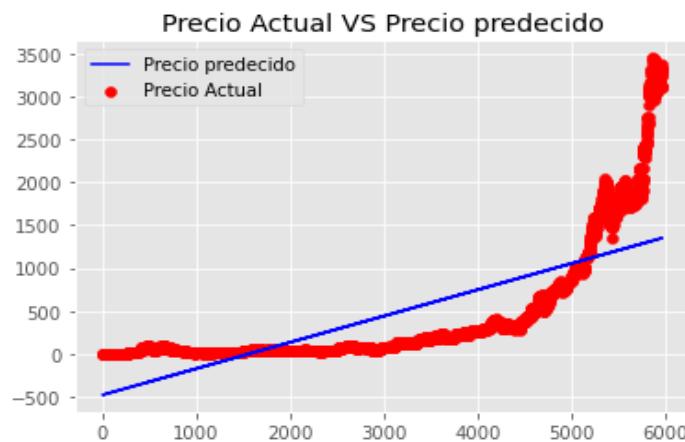


Gráfica 33. Test 1-Correlación lineal 1-Amazon

Test 2:

Atributos: test\_size=0.35, random\_state=50

**Resultados:** Score of the Linear Regression Model 0.5448903078838201

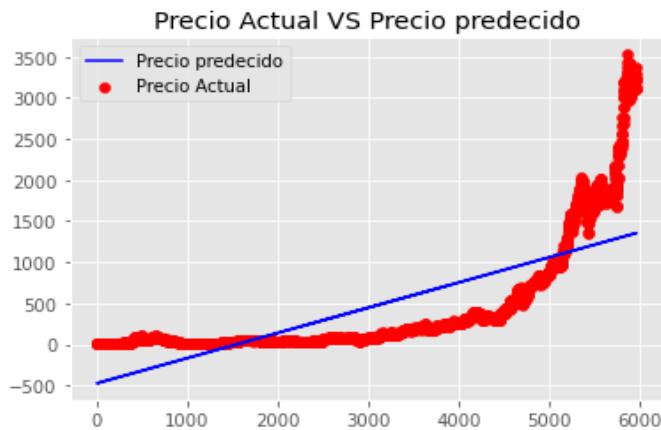


Gráfica 34. Test 2-Correlación lineal 1-Amazon

Test 3:

Atributos: test\_size=0.45, random\_state=60

**Resultados:** Score of the Linear Regression Model 0.553248747941792



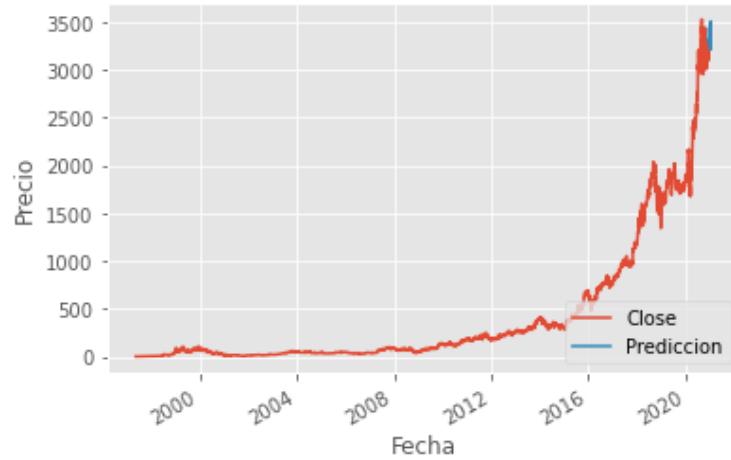
Gráfica 35. Test 3-Correlación lineal 1-Amazon

**-Prueba con programa de Correlación Lineal 2:**

Test 1:

Atributos: test\_size=0.2

**Resultados:** Score of the Linear Regression Model 0.9898938203996032

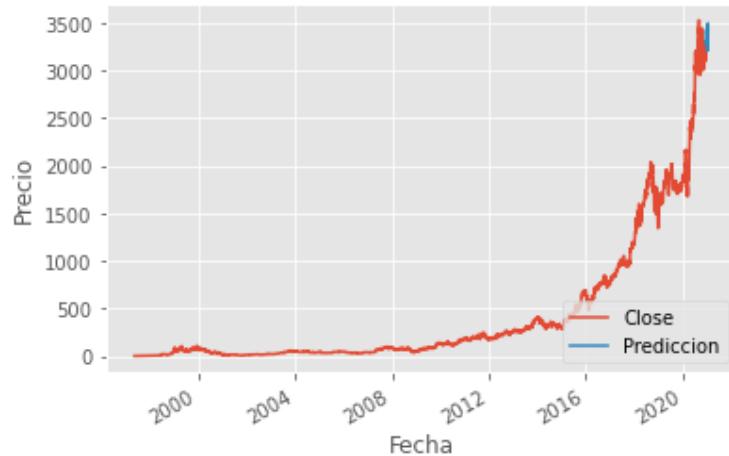


Gráfica 36. Test 1-Correlación lineal 2-Amazon

Test 2:

Atributos: test\_size=0.4

**Resultados:** Score of the Linear Regression Model 0.9886059789112015



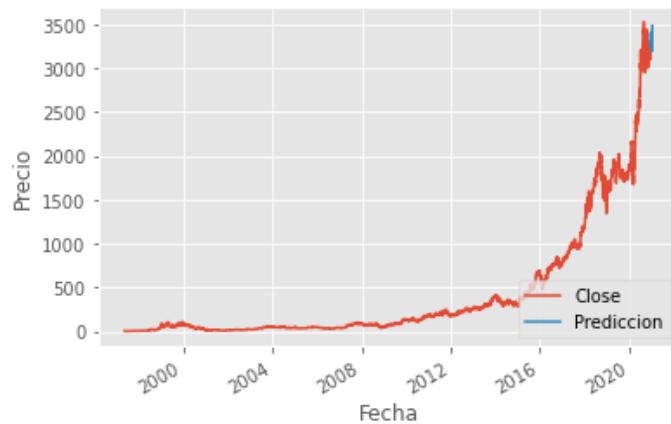
Gráfica 37. Test 2-Correlación lineal 2-Amazon

Test 3:

Atributos: test\_size=0.6

#### **Resultados:**

Score of the Linear Regression Model 0.9888760755268515



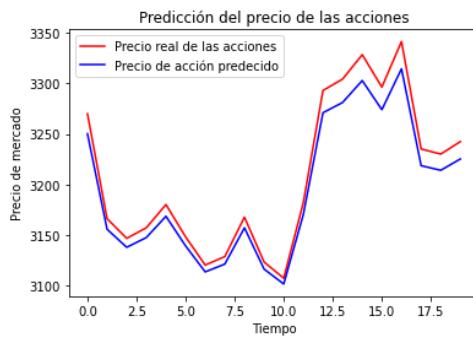
Gráfica 38. Test 3-Correlación lineal 2-Amazon

#### **-Prueba con programa de Redes Neuronales 1:**

Test 1:

Atributos: epochs = 200

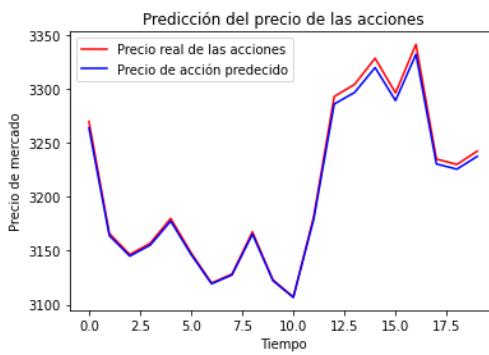
Resultados: error: 15.911142729426215



Test 2:

Atributos: epochs = 500

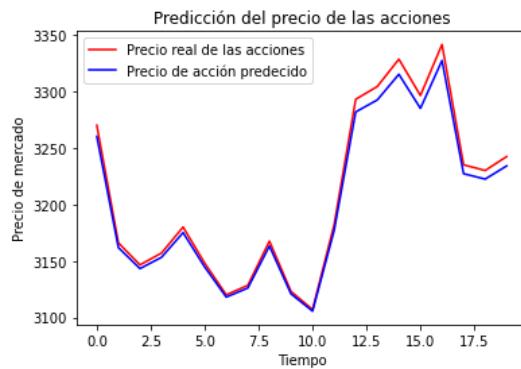
Resultados: error: 4.729788087593539



Test 3:

Atributos: epochs = 750

Resultados: error: 7.745100328710279



**-Prueba con programa de Redes Neuronales 2:**

Test 1:

Atributos: epochs = 200

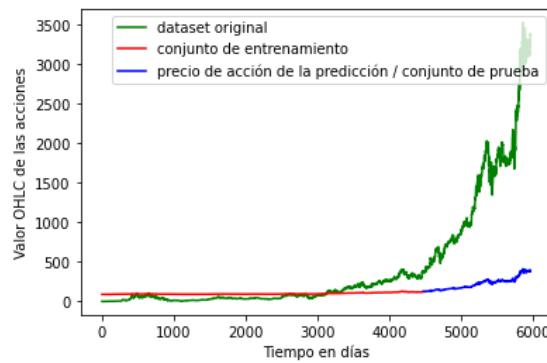
Resultados:

Entrenamiento RMSE: 91.71

Prueba RMSE: 1412.56

Valor del último día: 392.9415588378906

Valor al día siguiente: 45.910160064697266



Gráfica 42. Test 1-Redes Neuronales 2-Amazon

Test 2:

Atributos: epochs = 100

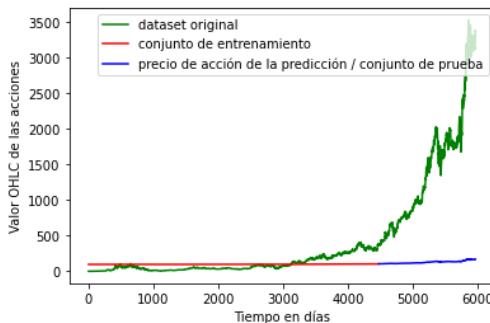
Resultados:

Entrenamiento RMSE: 98.96

Prueba RMSE: 1522.81

Valor del último día: 169.2087860107422

Valor al día siguiente: 8.289729118347168



Gráfica 43. Test 2-Redes Neuronales 2-Amazon

Test 3:

Atributos: epochs = 50

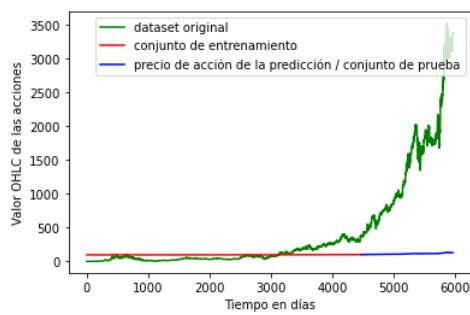
Resultados:

Entrenamiento RMSE: 100.08

Prueba RMSE: 1540.19

Valor del último día: 133.28379821777344

Valor al día siguiente: 5.071072578430176



Gráfica 44. Test 3-Redes Neuronales 2-Amazon

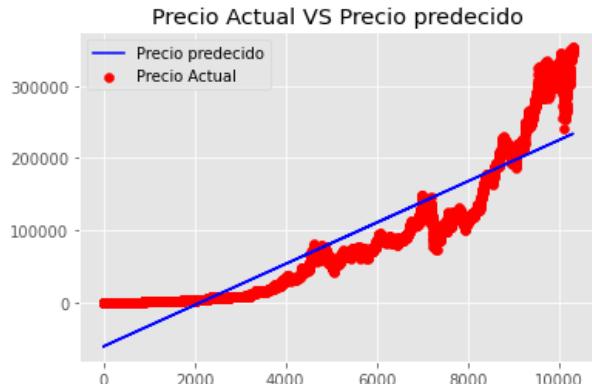
### 14.3 Anexo C. Pruebas con empresa Berkshire Hathaway

-Prueba con programa de Correlación Lineal 1:

Test 1:

Atributos: test\_size=0.25, random\_state=42

Resultados: Score of the Linear Regression Model 0.8274888429037911

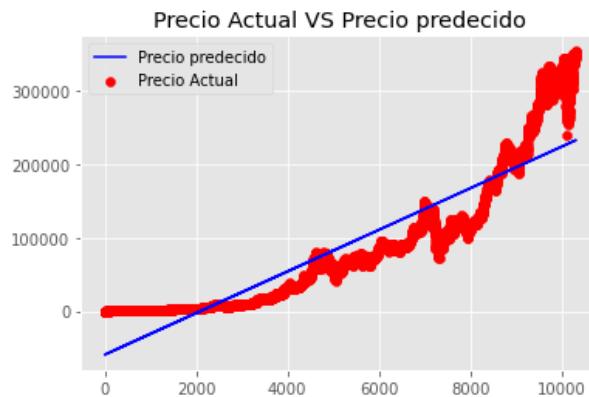


Gráfica 45.Test 1-Correlación Lineal 1-Berkshire Hathaway

Test 2:

Atributos: test\_size=0.35, random\_state=50

**Resultados:** Score of the Linear Regression Model 0.8276653121217437



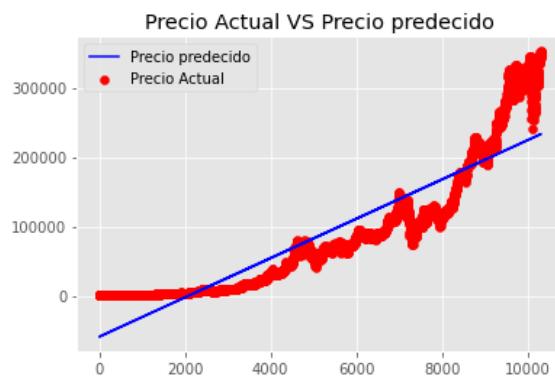
Gráfica 46.Test 2-Correlación Lineal 1-Berkshire Hathaway

Test 3:

Atributos: test\_size=0.45, random\_state=60

**Resultados:**

Score of the Linear Regression Model 0.8261837220491746



Gráfica 47.Test 3-Correlación Lineal 1-Berkshire Hathaway

-Prueba con programa de Correlación Lineal 2:

Test 1:

Atributos: test\_size=0.2

**Resultados:** Score of the Linear Regression Model 0.9930827864362123



Gráfica 48. Test 1-Correlación Lineal 2-Berkshire Hathaway

Test 2:

Atributos: test\_size=0.4

**Resultados:** Score of the Linear Regression Model 0.9935796496287518



Gráfica 49. Test 2-Correlación Lineal 2-Berkshire Hathaway

Test 3:

Atributos: test\_size=0.6

**Resultados:** Score of the Linear Regression Model 0.99337607951435



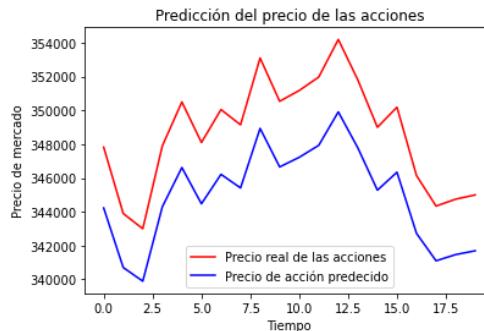
Gráfica 50. Test 3-Correlación Lineal 2-Berkshire Hathaway

## -Prueba con programa de Redes Neuronales 1:

Test 1:

Atributos: epochs = 200

Resultados: error: 3701.2358287416564

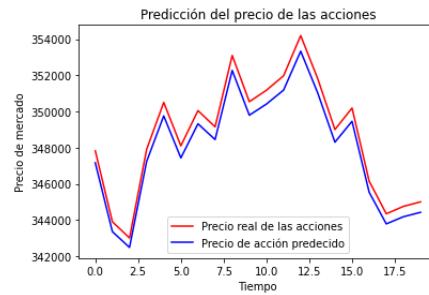


Gráfica 51. Test 1-Redes Neuronales 1-Berkshire Hathaway

Test 2:

Atributos: epochs = 500

Resultados: error: 695.8421687274326

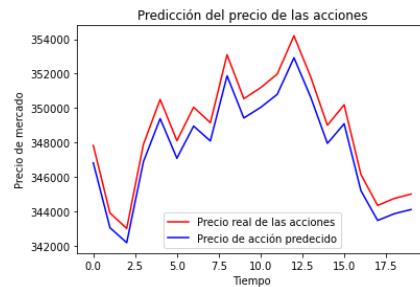


Gráfica 52. Test 2-Redes Neuronales 1-Berkshire Hathaway

Test 3:

Atributos: epochs = 750

Resultados: error: 1051.5591830138992



*Gráfica 53.Test 3-Redes Neuronales 1-Berkshire Hathaway*

**-Prueba con programa de Redes Neuronales 2:**

Test 1:

Atributos: epochs = 200

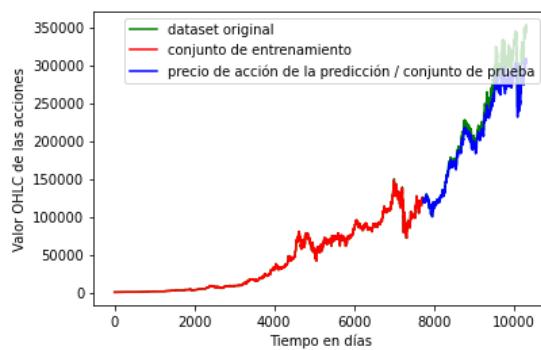
Resultados:

Entrenamiento RMSE: 745.18

Prueba RMSE: 18217.11

Valor del último día: 303950.3125

Valor al día siguiente: 265474.75



*Gráfica 54.Test 1-Redes Neuronales 2-Berkshire Hathaway*

Test 2:

Atributos: epochs = 100

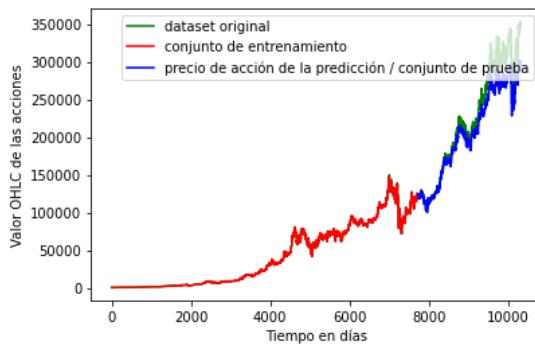
Resultados:

Entrenamiento RMSE: 771.61

Prueba RMSE: 21677.40

Valor del último día: 296962.65625

Valor al día siguiente: 253212.9375



Gráfica 55. Test 2-Redes Neuronales 2-Berkshire Hathaway

Test 3:

Atributos: epochs = 50

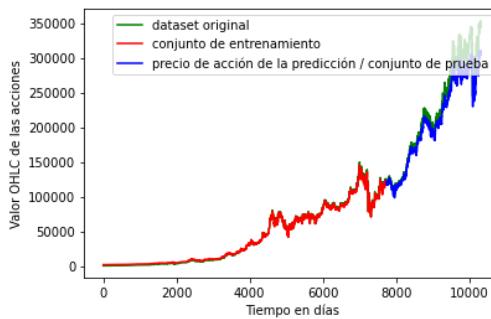
Resultados:

Entrenamiento RMSE: 1506.32

Prueba RMSE: 18489.72

Valor del último día: 305707.25

Valor al día siguiente: 268843.3125



Gráfica 56. Test 3-Redes Neuronales 2-Berkshire Hathaway

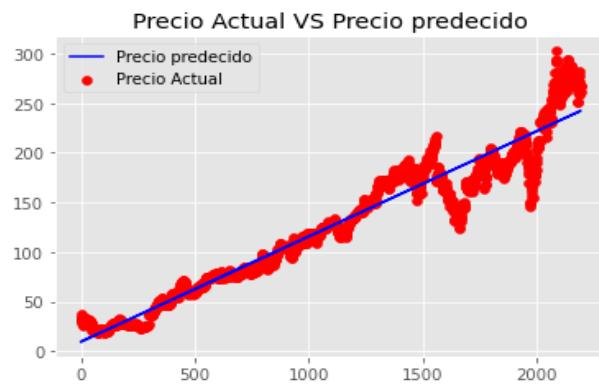
#### 14.4 Anexo D. Pruebas con empresa Facebook

-Prueba con programa de Correlación Lineal 1:

Test 1:

Atributos: test\_size=0.25, random\_state=42

**Resultados:** Score of the Linear Regression Model 0.9362192549227766

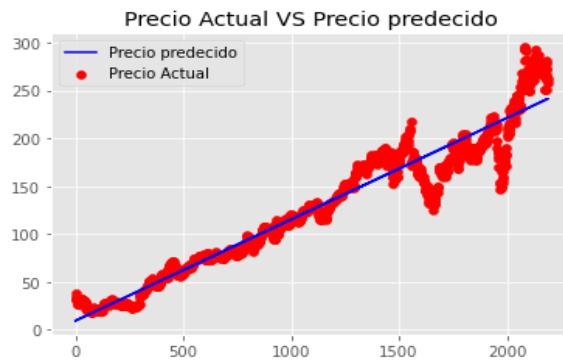


Gráfica 57. Test 1-Correlación lineal 1-Facebook

Test 2:

Atributos: test\_size=0.35, random\_state=50

**Resultados:** Score of the Linear Regression Model 0.9345774054625563

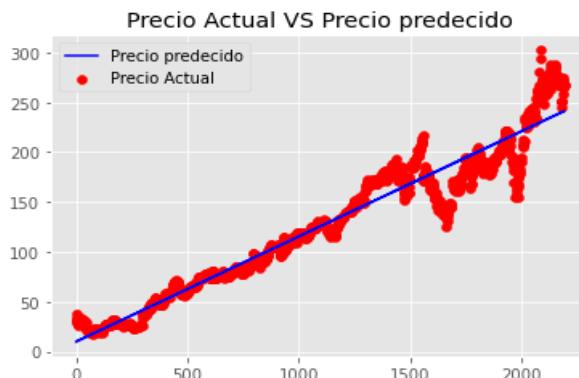


Gráfica 58. Test 2-Correlación lineal 1-Facebook

Test 3:

Atributos: test\_size=0.45, random\_state=60

**Resultados:** Score of the Linear Regression Model 0.9323610655283233



Gráfica 59.Test 3-Correlación lineal 1-Facebook

-Prueba con programa de Correlación Lineal 2:

Test 1:

Atributos: test\_size=0.2

**Resultados:** Score of the Linear Regression Model 0.956004766795847

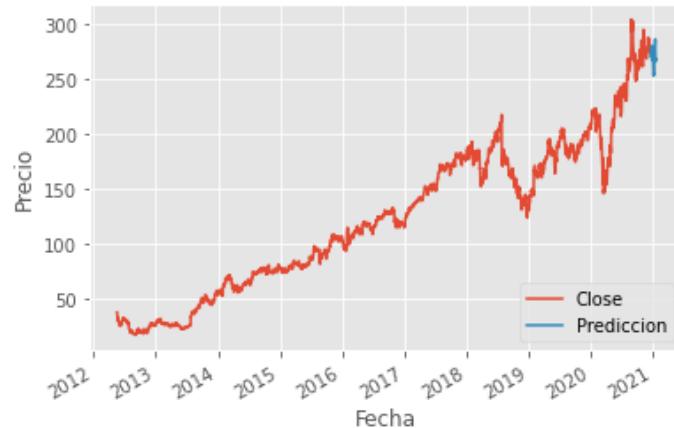


Gráfica 60.Test 1-Correlación lineal 2-Facebook

Test 2:

Atributos: test\_size=0.4

**Resultados:** Score of the Linear Regression Model 0.9611791254656632



Gráfica 61.Test 2-Correlación lineal 2-Facebook

Test 3:

Atributos: test\_size=0.6

**Resultados:** Score of the Linear Regression Model 0.9583746021108157



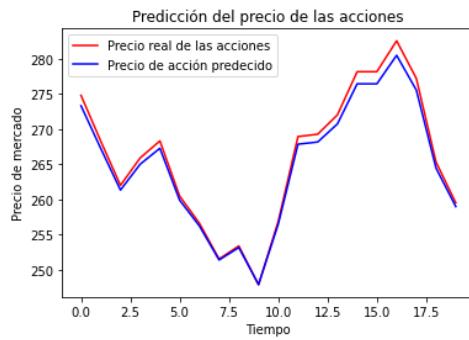
Gráfica 62. Test 3-Correlación lineal 2-Facebook

#### -Prueba con programa de Redes Neuronales 1:

Test 1:

Atributos: epochs = 200

Resultados: error: 1.1020328090080538

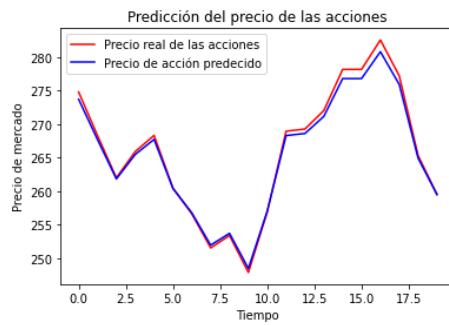


Gráfica 63. Test 1-Redes Neuronales 1 -Facebook

Test 2:

Atributos: epochs = 500

Resultados: error: 0.8062193233275442

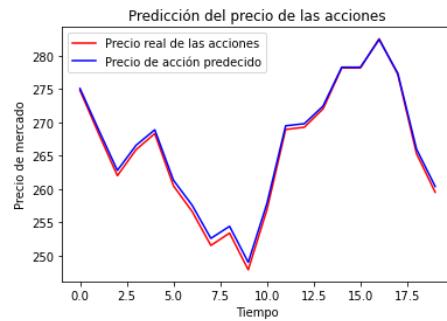


Gráfica 64. Test 2-Redes Neuronales 1 -Facebook

Test 3:

Atributos: epochs = 750

Resultados: error: 0.695361994798917



Gráfica 65. Test 3-Redes Neuronales 1 -Facebook

**-Prueba con programa de Redes Neuronales 2:**

Test 1:

Atributos: epochs = 200

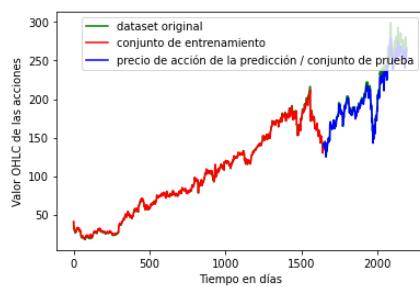
Resultados:

Entrenamiento RMSE: 1.93

Prueba RMSE: 8.20

Valor del último día: 248.95394897460938

Valor al día siguiente: 230.29002380371094



Gráfica 66. Test 1-Redes Neuronales 2 -Facebook

Test 2:

Atributos: epochs = 100

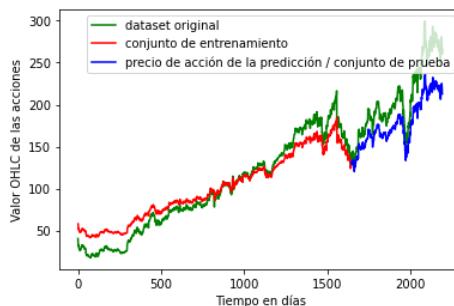
Resultados:

Entrenamiento RMSE: 14.42

Prueba RMSE: 33.03

Valor del último día: 212.9302978515625

Valor al día siguiente: 164.91575622558594



Gráfica 67. Test 2-Redes Neuronales 2 -Facebook

Test 3:

Atributos: epochs = 50

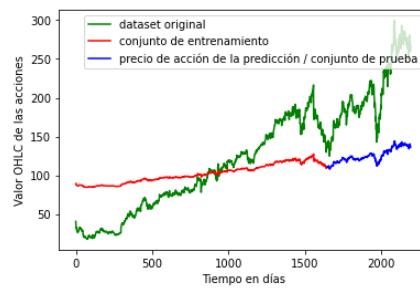
Resultados:

Entrenamiento RMSE: 41.55

Prueba RMSE: 88.60

Valor del último día: 136.53993225097656

Valor al día siguiente: 61.30923843383789



Gráfica 68. Test 3-Redes Neuronales 2 -Facebook

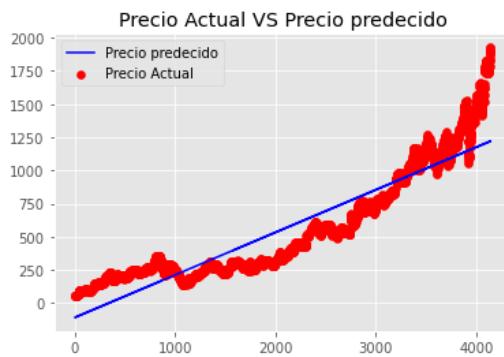
## 14.5 Anexo E. Pruebas con empresa Alphabet

### -Prueba con programa de Correlación Lineal 1:

Test 1:

Atributos: test\_size=0.25, random\_state=42

Resultados: Score of the Linear Regression Model 0.8466574526173501

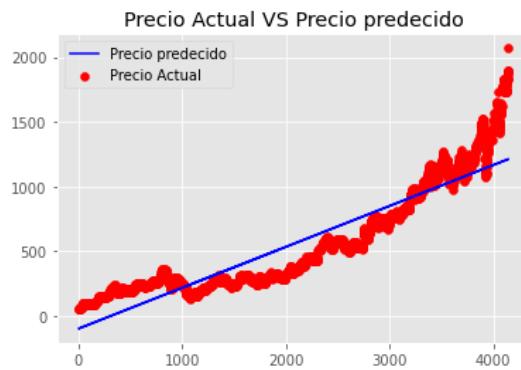


Gráfica 69.Test 1-Correlación Lineal 1 -Alphabet

Test 2:

Atributos: test\_size=0.35, random\_state=50

**Resultados:** Score of the Linear Regression Model 0.8422815261157491

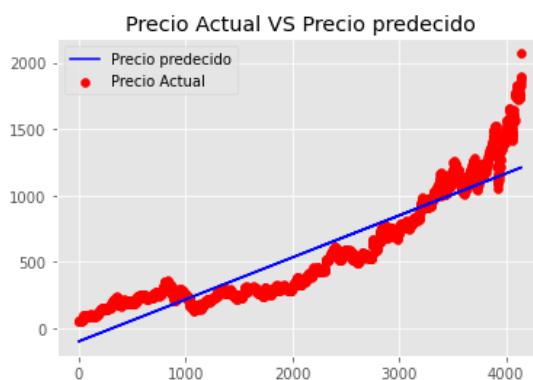


Gráfica 70.Test 2-Correlación Lineal 1 -Alphabet

Test 3:

Atributos: test\_size=0.45, random\_state=60

**Resultados:** Score of the Linear Regression Model 0.8391468413123021



Gráfica 71.Test 3-Correlación Lineal 1 -Alphabet

**-Prueba con programa de Correlación Lineal 2:**

Test 1:

Atributos: test\_size=0.2

**Resultados:** Score of the Linear Regression Model 0.9810711368647088

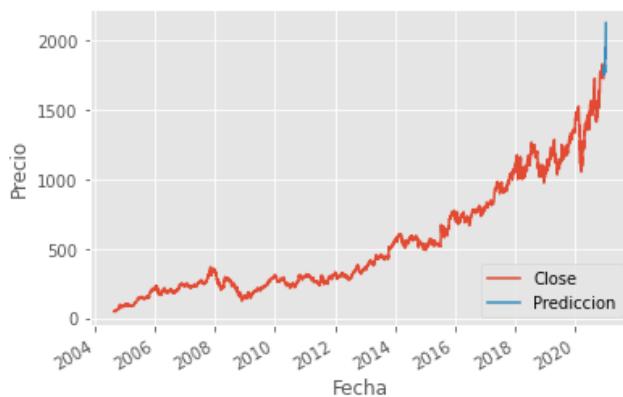


Gráfica 72. Test 1-Correlación Lineal 2 -Alphabet

Test 2:

Atributos: test\_size=0.4

**Resultados:** Score of the Linear Regression Model 0.9800258139653405

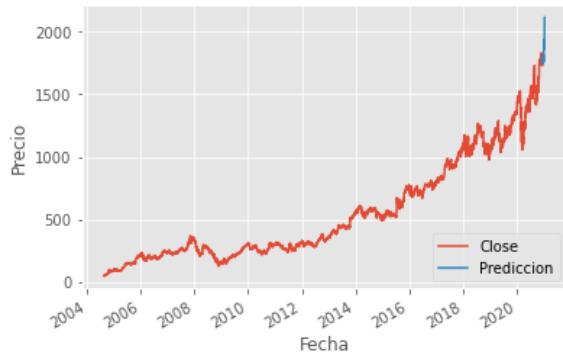


Gráfica 73. Test 2-Correlación Lineal 2 -Alphabet

Test 3:

Atributos: test\_size=0.6

**Resultados:** Score of the Linear Regression Model 0.9824098581663854



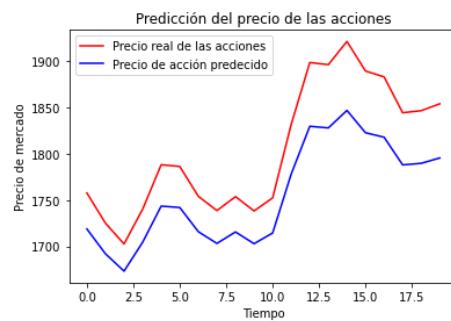
Gráfica 74. Test 3-Correlación Lineal 2 -Alphabet

#### -Prueba con programa de Redes Neuronales 1:

Test 1:

Atributos: epochs = 200

**Resultados:** error: 30.752271886562593

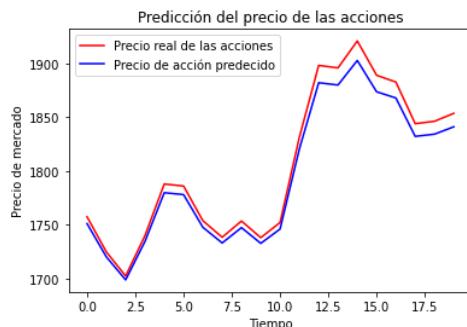


Gráfica 75. Test 1-Redes Neuronales 1 -Alphabet

Test 2:

Atributos: epochs = 500

**Resultados:** error: 10.598201891930568

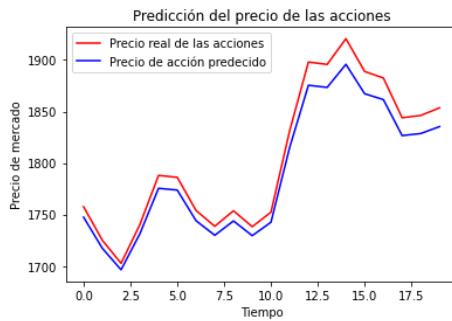


Gráfica 76. Test 2-Redes Neuronales 1 -Alphabet

Test 3:

Atributos: epochs = 750

**Resultados:** error: 15.43327313451624



Gráfica 77. Test 3-Redes Neuronales 1 -Alphabet

-Prueba con programa de Redes Neuronales 2:

Test 1:

Atributos: epochs = 200

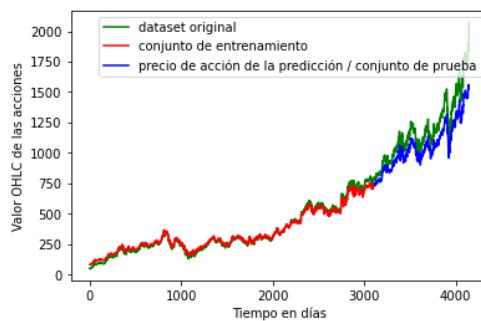
Resultados:

Entrenamiento RMSE: 20.99

Prueba RMSE: 143.29

Valor del último día: 1547.2540283203125

Valor al día siguiente: 1229.76904296875



Gráfica 78. Test 1-Redes Neuronales 2 -Alphabet

Test 2:

Atributos: epochs = 100

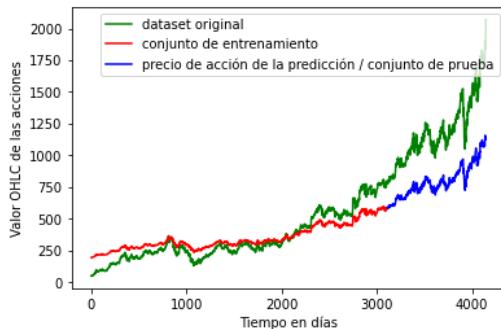
Resultados:

Entrenamiento RMSE: 87.95

Prueba RMSE: 409.96

Valor del último día: 1144.970458984375

Valor al día siguiente: 670.5430297851562



Gráfica 79. Test 2-Redes Neuronales 2 -Alphabet

Test 3:

Atributos: epochs = 50

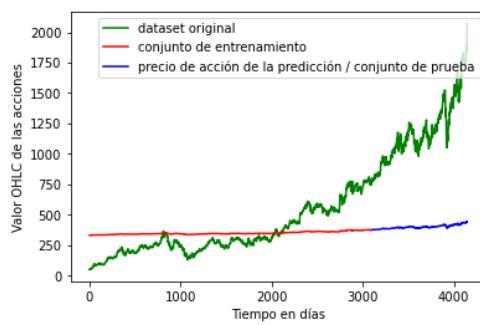
Resultados:

Entrenamiento RMSE: 175.17

Prueba RMSE: 819.51

Valor del último día: 442.14617919921875

Valor al día siguiente: 88.26483154296875



Gráfica 80. Test 3-Redes Neuronales 2 -Alphabet

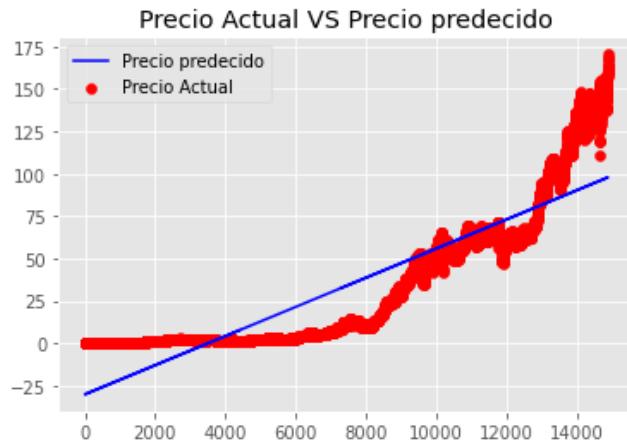
## 14.6 Anexo F. Pruebas con empresa Johnson & Johnson

-Prueba con programa de Correlación Lineal 1:

Test 1:

Atributos: test\_size=0.25, random\_state=42

**Resultados:** Score of the Linear Regression Model 0.774811124242707

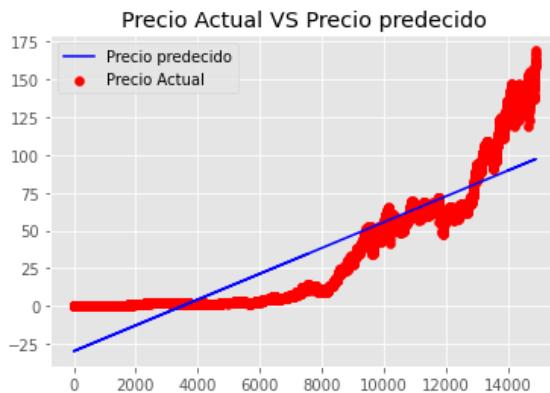


Gráfica 81. Test 1-Correlación lineal 1 -J&J

Test 2:

Atributos: test\_size=0.35, random\_state=50

**Resultados:** Score of the Linear Regression Model 0.7843288053584068

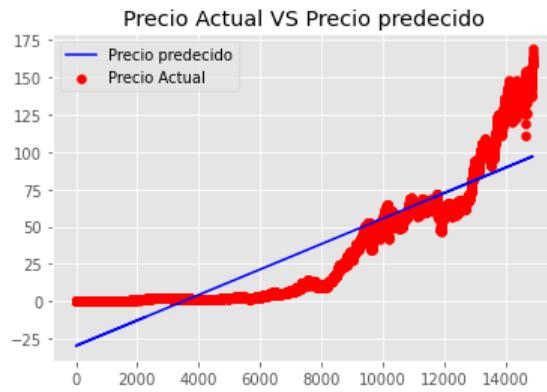


Gráfica 82. Test 2-Correlación lineal 1 -J&J

Test 3:

Atributos: test\_size=0.45, random\_state=60

**Resultados:** Score of the Linear Regression Model 0.7793808080439903



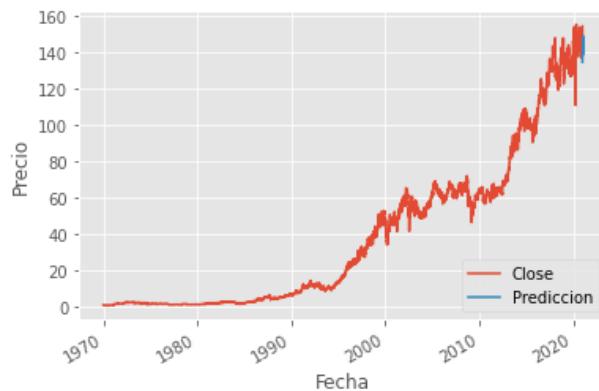
Gráfica 83. Test 3-Correlación lineal 1 -J&J

#### -Prueba con programa de Correlación Lineal 2:

Test 1:

Atributos: test\_size=0.2

**Resultados:** Score of the Linear Regression Model 0.9676117450005048

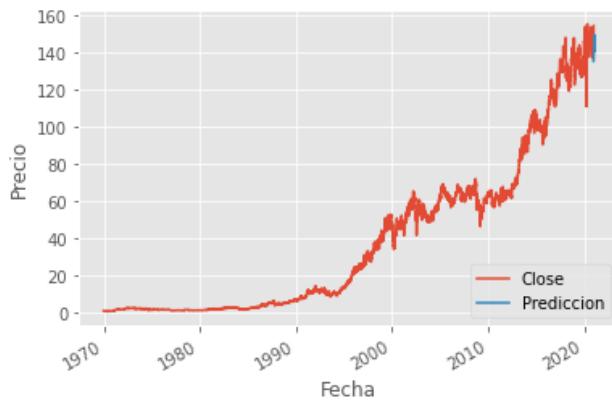


Gráfica 84. Test 1-Correlación lineal 2 -J&J

Test 2:

Atributos: test\_size=0.4

**Resultados:** Score of the Linear Regression Model 0.9676158790399497

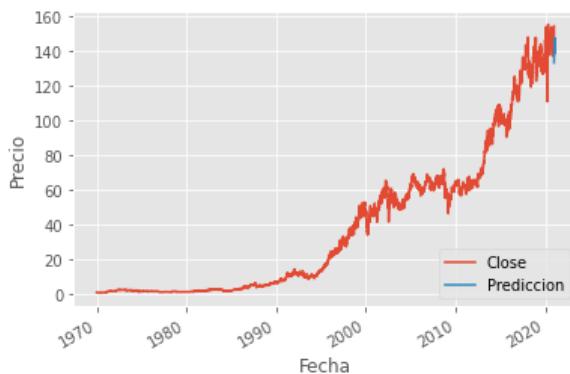


Gráfica 85. Test 2-Correlación lineal 2 -J&J

Test 3:

Atributos: test\_size=0.6

**Resultados:** Score of the Linear Regression Model 0.9678352811191353



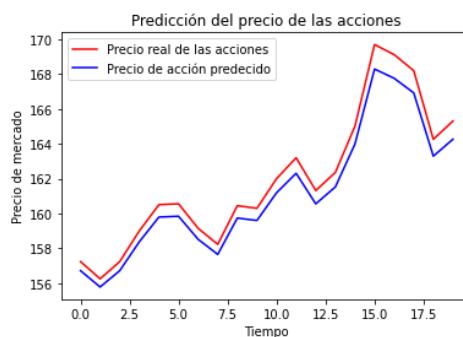
Gráfica 86. Test 3-Correlación lineal 2 -J&J

#### -Prueba con programa de Redes Neuronales 1:

Test 1:

Atributos: epochs = 200

Resultados: error: 0.8702770992256327

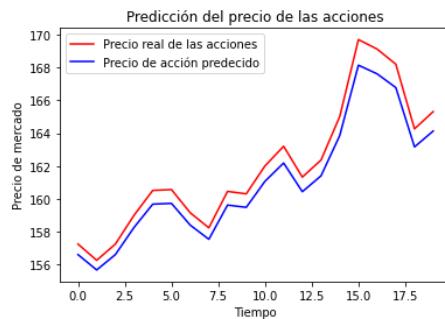


Gráfica 87. Test 1-Redes Neuronales 1 -J&J

Test 2:

Atributos: epochs = 500

Resultados: error: 0.996796383705952

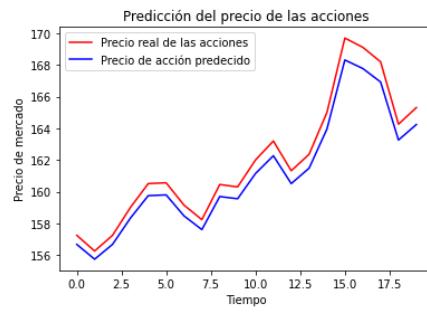


Gráfica 88. Test 2-Redes Neuronales 1 -J&J

Test 3:

Atributos: epochs = 750

Resultados: error: 0.8982144393984391



Gráfica 89. Test 3-Redes Neuronales 1 -J&J

#### -Prueba con programa de Redes Neuronales 2:

Test 1:

Atributos: epochs = 200

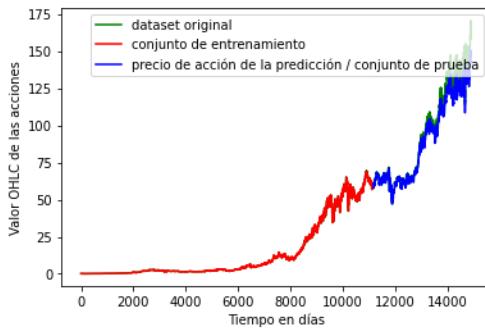
Resultados:

Entrenamiento RMSE: 0.29

Prueba RMSE: 5.91

Valor del último día: 146.1570281982422

Valor al día siguiente: 128.91632080078125



Gráfica 90. Test 1-Redes Neuronales 2 -J&J

Test 2:

Atributos: epochs = 100

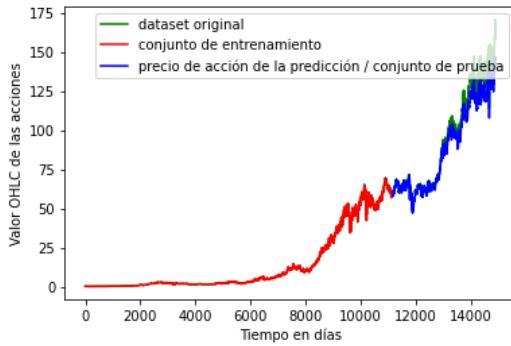
Resultados:

Entrenamiento RMSE: 0.29

Prueba RMSE: 6.82

Valor del último día: 143.1105499267578

Valor al día siguiente: 123.32766723632812



Gráfica 91. Test 2-Redes Neuronales 2 -J&J

Test 3:

Atributos: epochs = 50

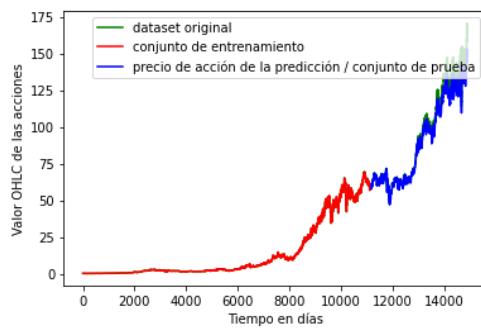
Resultados:

Entrenamiento RMSE: 0.28

Prueba RMSE: 5.01

Valor del último día: 148.38250732421875

Valor al día siguiente: 132.95484924316406



Gráfica 92. Test 3-Redes Neuronales 2 -J&J

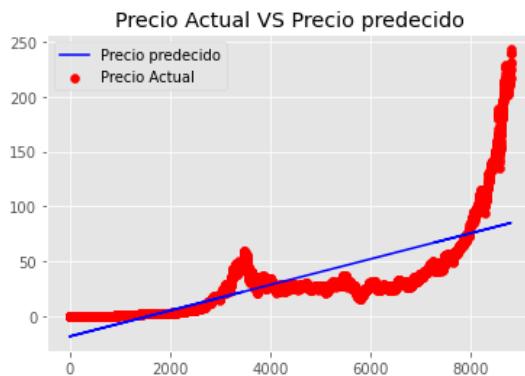
#### 14.7 Anexo G. Pruebas con empresa Microsoft

##### -Prueba con programa de Correlación Lineal 1:

Test 1:

Atributos: test\_size=0.25, random\_state=42

**Resultados:** Score of the Linear Regression Model 0.5402418699664517



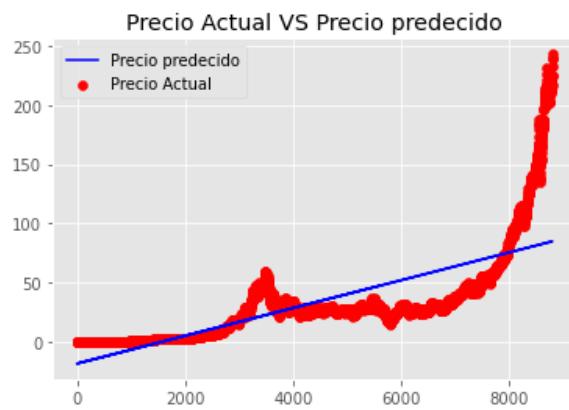
Gráfica 93. Test 1-Correlación Lineal 1 -Microsoft

Test 2:

Atributos: test\_size=0.35, random\_state=50

**Resultados:**

Score of the Linear Regression Model 0.542474832689299

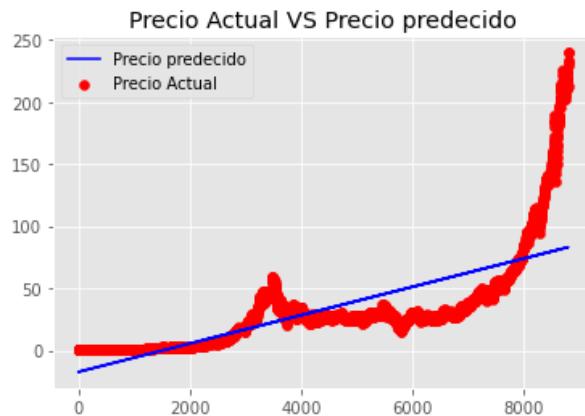


Gráfica 94. Test 2-Correlación Lineal 1 -Microsoft

Test 3:

Atributos: test\_size=0.45, random\_state=60

**Resultados:** Score of the Linear Regression Model 0.5427033518608348



Gráfica 95. Test 3-Correlación Lineal 1 -Microsoft

-Prueba con programa de Correlación Lineal 2:

Test 1:

Atributos: test\_size=0.2

**Resultados:** Score of the Linear Regression Model 0.9789609871413563



Gráfica 96. Test 1-Correlación Lineal 2 -Microsoft

Test 2:

Atributos: test\_size=0.4

**Resultados:** Score of the Linear Regression Model 0.9760197380709916

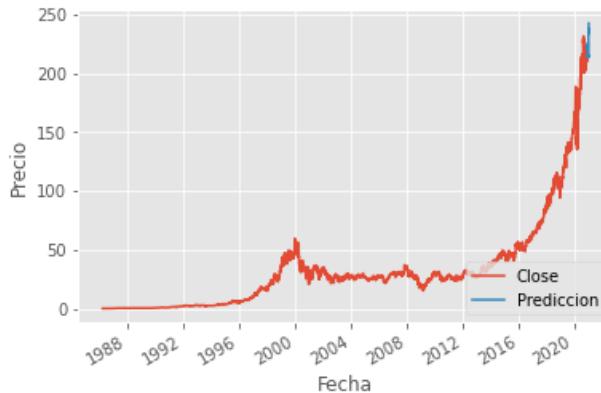


Gráfica 97. Test 2-Correlación Lineal 2 -Microsoft

Test 3:

Atributos: test\_size=0.6

**Resultados:** Score of the Linear Regression Model 0.979424802031503



Gráfica 98. Test 3-Correlación Lineal 2 -Microsoft

## -Prueba con programa de Redes Neuronales 1:

Test 1:

Atributos: epochs = 200

Resultados: error: 0.5679991031777459

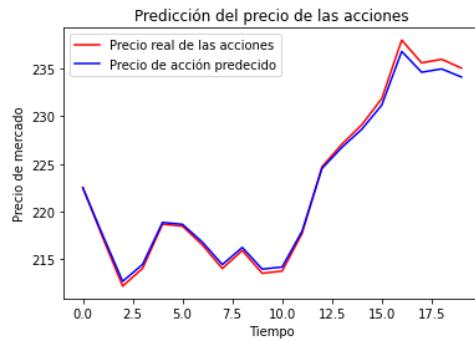


Gráfico 99. Test 1-Redes Neuronales 1 -Microsoft

Test 2:

Atributos: epochs = 500

Resultados: error: 0.9288485089356294

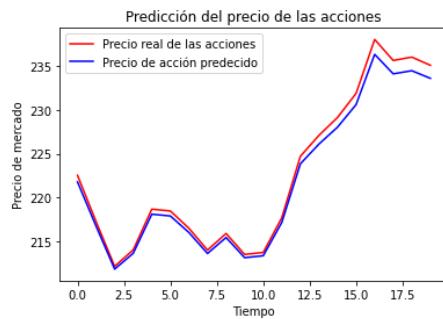
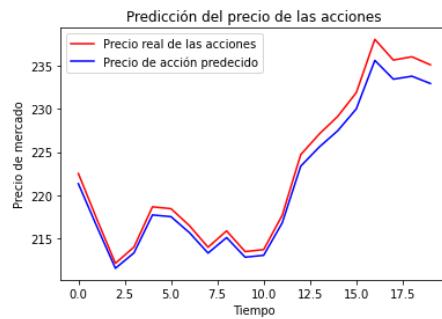


Gráfico 100. Test 2-Redes Neuronales 1 -Microsoft

Test 3:

Atributos: epochs = 750

Resultados: error: 1.3904130732022832



*Gráfica 101. Test 3-ReDES NEURONALES 1 -Microsoft*

#### **-Prueba con programa de Redes Neuronales 2:**

Test 1:

Atributos: epochs = 200

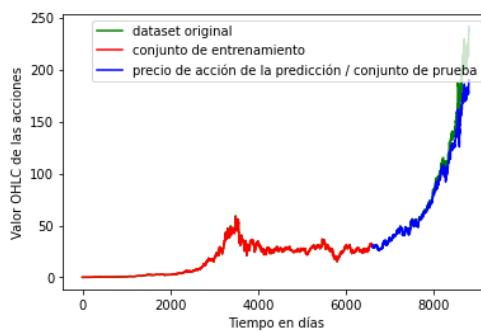
Resultados:

Entrenamiento RMSE: 0.42

Prueba RMSE: 11.66

Valor del último día: 190.08883666992188

Valor al día siguiente: 151.1627197265625



*Gráfica 102. Test 1-ReDES NEURONALES 2 -Microsoft*

Test 2:

Atributos: epochs = 100

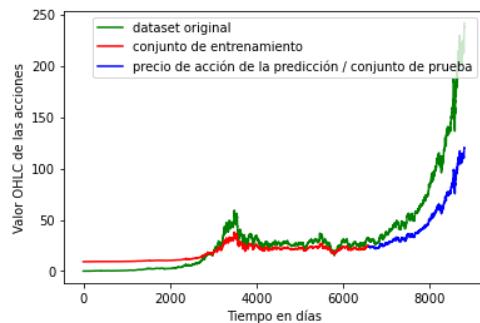
Resultados:

Entrenamiento RMSE: 7.18

Prueba RMSE: 43.49

Valor del último día: 120.4112777709961

Valor al día siguiente: 60.83268356323242



Gráfica 103. Test 2-Redes Neuronales 2 -Microsoft

Test 3:

Atributos: epochs = 50

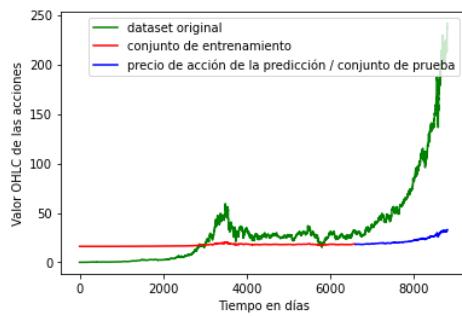
Resultados:

Entrenamiento RMSE: 12.86

Prueba RMSE: 77.81

Valor del último día: 33.11177062988281

Valor al día siguiente: 4.565129280090332



Gráfica 104. Test 3-Redes Neuronales 2 -Microsoft

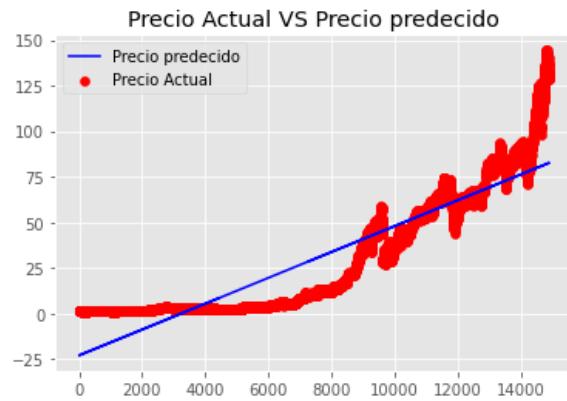
## 14.8 Anexo H. Pruebas con empresa Procter & Gamble

-Prueba con programa de Correlación Lineal 1:

Test 1:

Atributos: test\_size=0.25, random\_state=42

**Resultados:** Score of the Linear Regression Model 0.8113327832532735

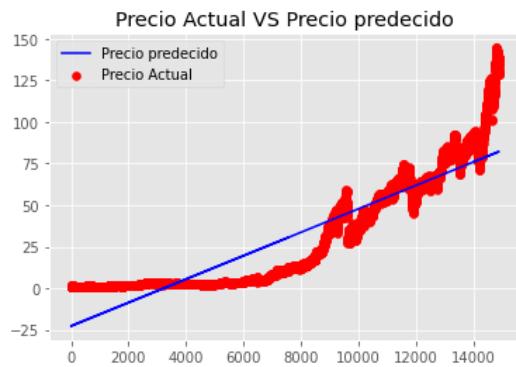


*Gráfica 105. Test 1-Correlación Lineal 1 -P&G*

Test 2:

Atributos: test\_size=0.35, random\_state=50

**Resultados:** Score of the Linear Regression Model 0.817961967262511

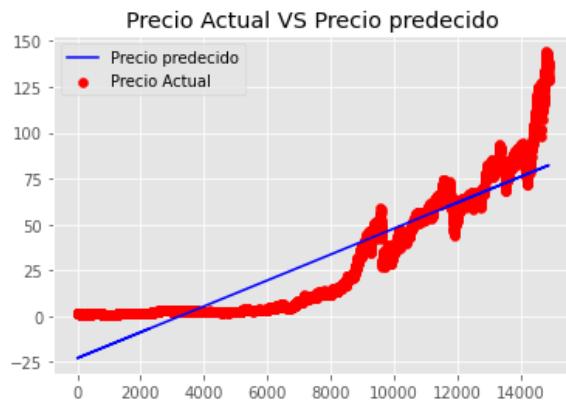


*Gráfica 106. Test 2-Correlación Lineal 1 -P&G*

Test 3:

Atributos: test\_size=0.45, random\_state=60

**Resultados:** Score of the Linear Regression Model 0.8130507175936332



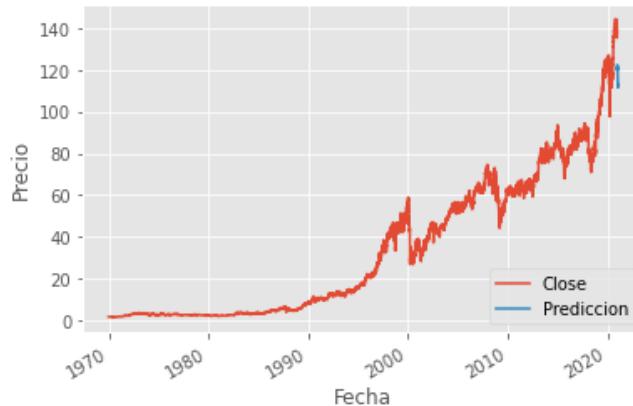
Gráfica 107. Test 3-Correlación Lineal 1 -P&G

-Prueba con programa de Correlación Lineal 2:

Test 1:

Atributos: test\_size=0.2

**Resultados:** Score of the Linear Regression Model 0.9576459914678895

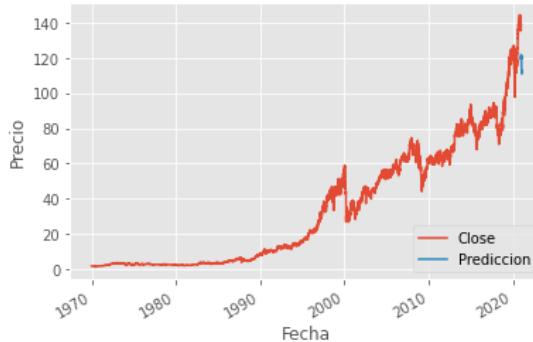


Gráfica 108. Test 1-Correlación Lineal 2 -P&G

Test 2:

Atributos: test\_size=0.4

**Resultados:** Score of the Linear Regression Model 0.9585575806028583

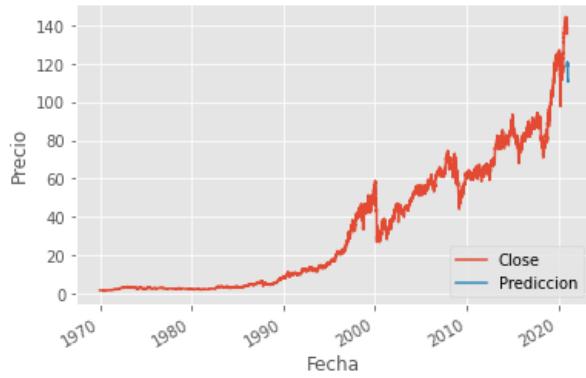


Gráfica 109. Test 2-Correlación Lineal 2 -P&G

Test 3:

Atributos: test\_size=0.6

**Resultados:** Score of the Linear Regression Model 0.9587024510109063

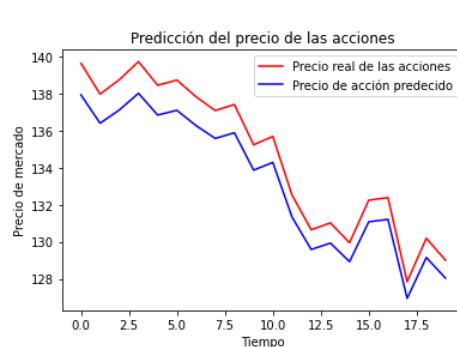


### -Prueba con programa de Redes Neuronales 1:

Test 1:

Atributos: epochs = 200

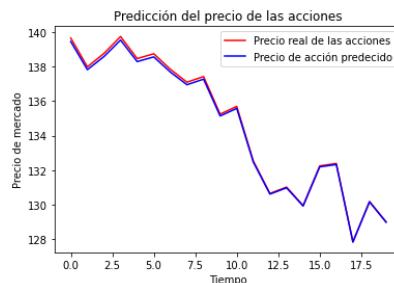
Resultados: error: 1.370565465087694



Test 2:

Atributos: epochs = 500

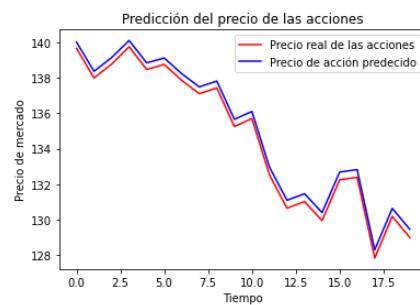
Resultados: error: 0.1260535385257277



Test 3:

Atributos: epochs = 750

Resultados: error: 0.4098822940982482



Gráfica 113. Test 3-Redes Neuronales 1 -P&G

#### -Prueba con programa de Redes Neuronales 2:

Test 1:

Atributos: epochs = 200

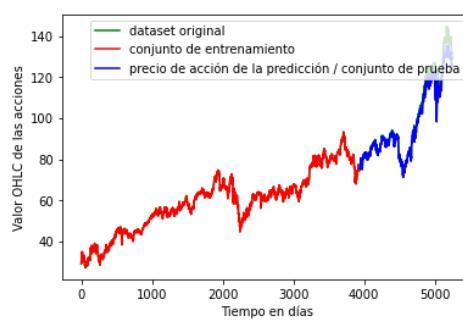
Resultados:

Entrenamiento RMSE: 0.54

Prueba RMSE: 2.98

Valor del último día: 123.88084411621094

Valor al día siguiente: 114.30765533447266



Gráfica 114. Test 1-Redes Neuronales 2 -P&G

Test 2:

Atributos: epochs = 100

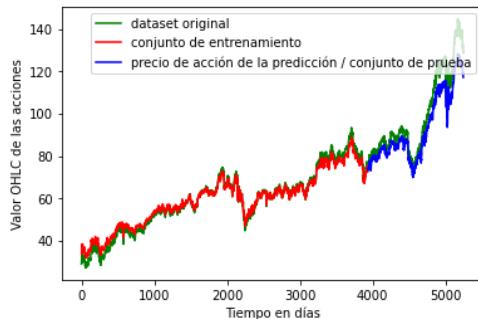
Resultados:

Entrenamiento RMSE: 2.03

Prueba RMSE: 7.18

Valor del último día: 117.08013916015625

Valor al día siguiente: 100.76918029785156



Gráfica 115. Test 2-Redes Neuronales 2 -P&G

Test 3:

Atributos: epochs = 50

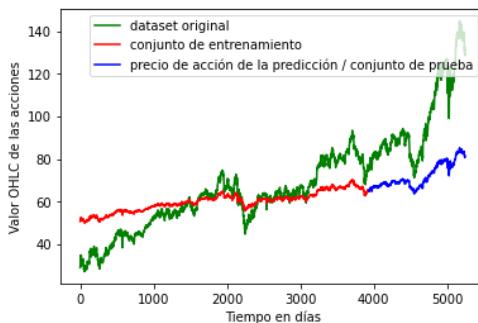
Resultados:

Entrenamiento RMSE: 9.98

Prueba RMSE: 30.34

Valor del último día: 80.74797821044922

Valor al día siguiente: 39.78200149536133



Gráfica 116. Test 3-Redes Neuronales 2 -P&G

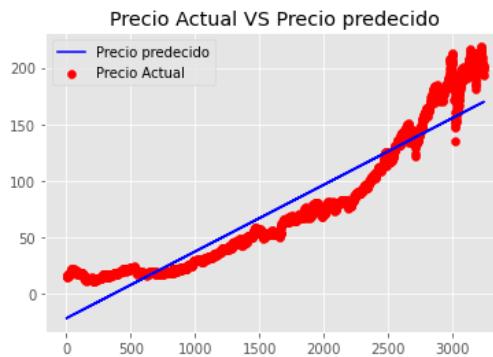
## 14.9 Anexo I. Pruebas con empresa Visa Inc.

### -Prueba con programa de Correlación Lineal 1:

Test 1:

Atributos: test\_size=0.25, random\_state=42

**Resultados:** Score of the Linear Regression Model 0.8834415430908416

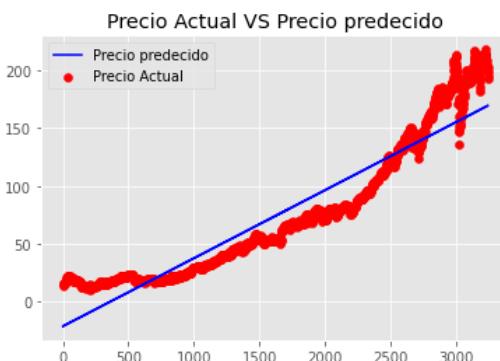


Gráfica 117. Test 1-Correlación Lineal 1 -Visa

Test 2:

Atributos: test\_size=0.35, random\_state=50

**Resultados:** Score of the Linear Regression Model 0.8830302496719609

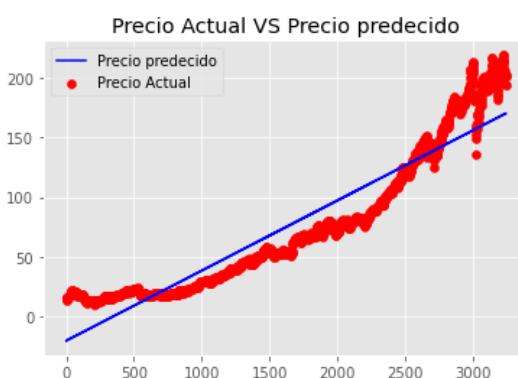


Gráfica 118. Test 2-Correlación Lineal 1 -Visa

Test 3:

Atributos: test\_size=0.45, random\_state=60

**Resultados:** Score of the Linear Regression Model 0.882787932183045



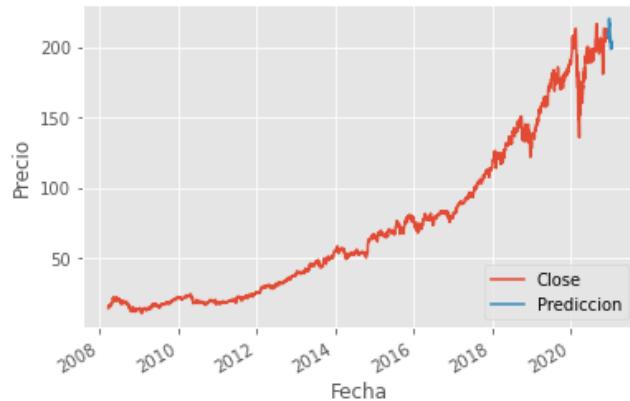
Gráfica 119. Test 3-Correlación Lineal 1 -Visa

-Prueba con programa de Correlación Lineal 2:

Test 1:

Atributos: test\_size=0.2

**Resultados:** Score of the Linear Regression Model 0.986053471223436

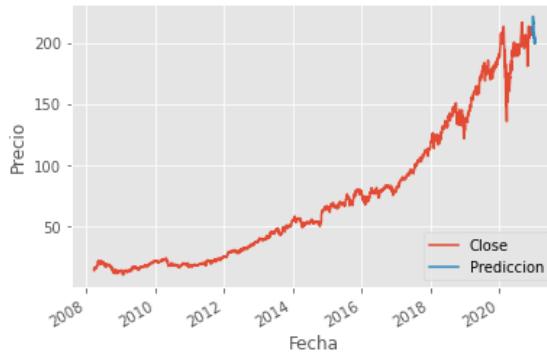


Gráfica 120. Test 1-Correlación Lineal 2 -Visa

Test 2:

Atributos: test\_size=0.4

**Resultados:** Score of the Linear Regression Model 0.9846726429020765

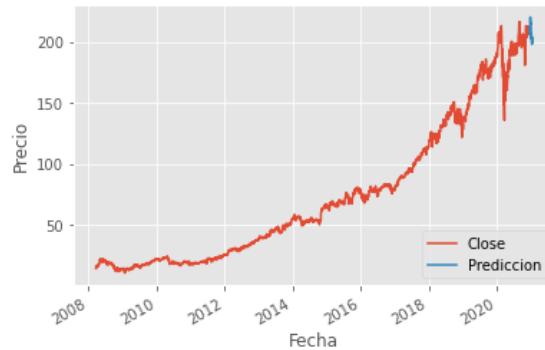


Gráfica 121. Test 2-Correlación Lineal 2 -Visa

Test 3:

Atributos: test\_size=0.6

**Resultados:** Score of the Linear Regression Model 0.9866743272875365



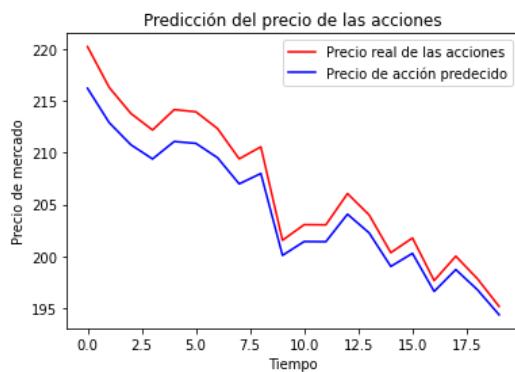
*Gráfica 122. Test 3-Correlación Lineal 2 -Visa*

#### **-Prueba con programa de Redes Neuronales 1:**

Test 1:

Atributos: epochs = 200

Resultados: error: 2.312904965668219

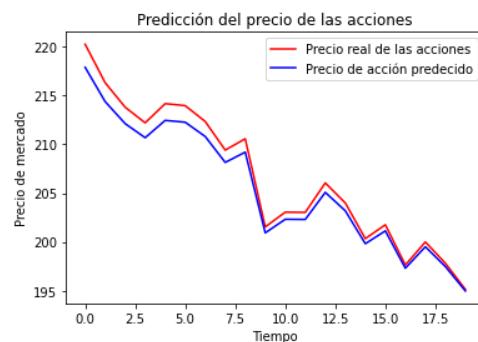


*Gráfica 123. Test 1-Redes Neuronales 1 -Visa*

Test 2:

Atributos: epochs = 500

Resultados: error: 1.2313488619765085

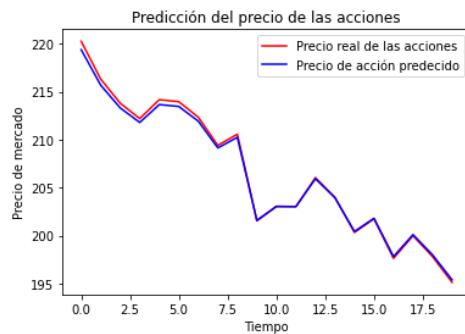


Gráfica 124. Test 2-Redes Neuronales 1 -Visa

Test 3:

Atributos: epochs = 750

Resultados: error: 0.3607054663621247



Gráfica 125. Test 3-Redes Neuronales 1 -Visa

**-Prueba con programa de Redes Neuronales 2:**

Test 1:

Atributos: epochs = 200

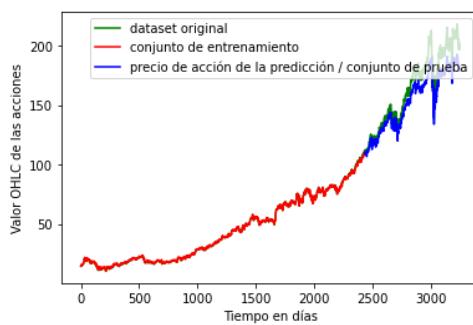
Resultados:

Entrenamiento RMSE: 0.63

Prueba RMSE: 12.53

Valor del último día: 179.0460968017578

Valor al día siguiente: 157.151611328125



Gráfica 126. Test 1-Redes Neuronales 2 -Visa

Test 2:

Atributos: epochs = 100

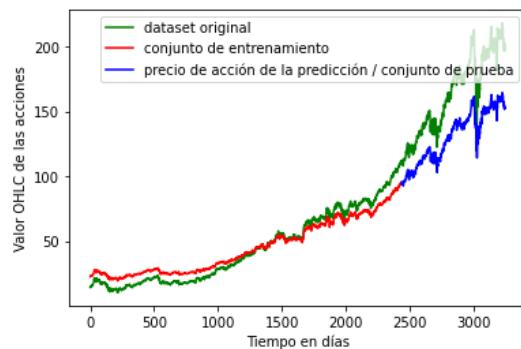
Resultados:

Entrenamiento RMSE: 6.80

Prueba RMSE: 34.67

Valor del último día: 152.31031799316406

Valor al día siguiente: 112.9681625366211



Test 3:

Atributos: epochs = 50

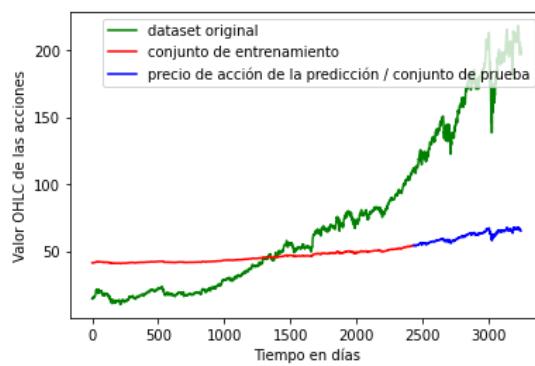
Resultados:

Entrenamiento RMSE: 23.43

Prueba RMSE: 105.40

Valor del último día: 65.7435073852539

Valor al día siguiente: 18.22549819946289



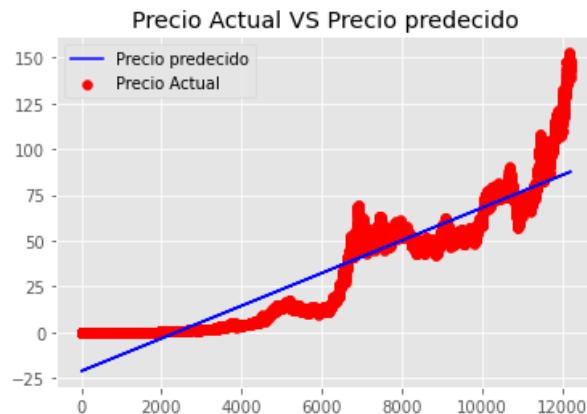
## 14.10 Anexo J. Pruebas con empresa Walmart

### -Prueba con programa de Correlación Lineal 1:

Test 1:

Atributos: test\_size=0.25, random\_state=42

**Resultados:** Score of the Linear Regression Model 0.8473734625027491

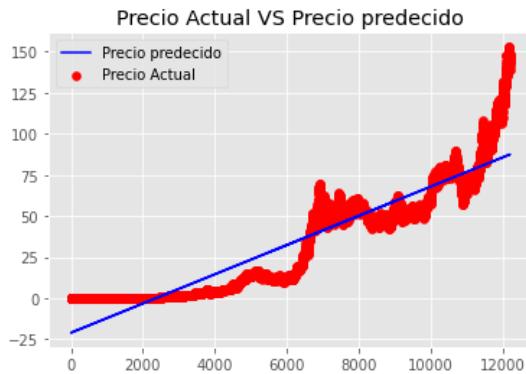


Gráfica 129. Test 1-Correlación Lineal 1 -Walmart

Test 2:

Atributos: test\_size=0.35, random\_state=50

**Resultados:** Score of the Linear Regression Model 0.8489476818179207

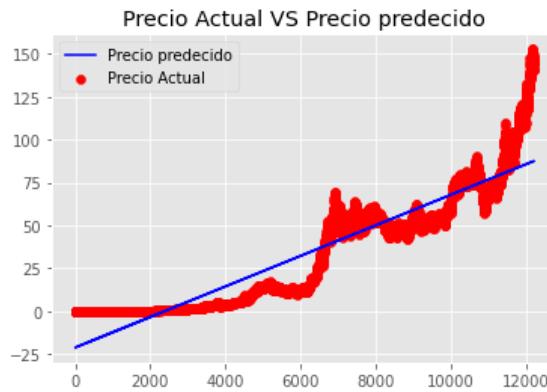


Gráfica 130. Test 2-Correlación Lineal 1 -Walmart

Test 3:

Atributos: test\_size=0.45, random\_state=60

**Resultados:** Score of the Linear Regression Model 0.8502902356009842



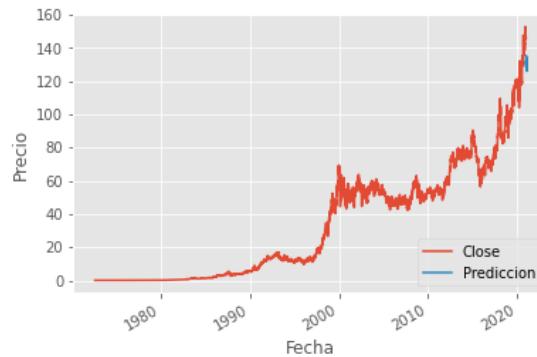
Gráfica 131. Test 3-Correlación Lineal 1 -Walmart

#### -Prueba con programa de Correlación Lineal 2:

Test 1:

Atributos: test\_size=0.2

**Resultados:** Score of the Linear Regression Model 0.9673769895622418

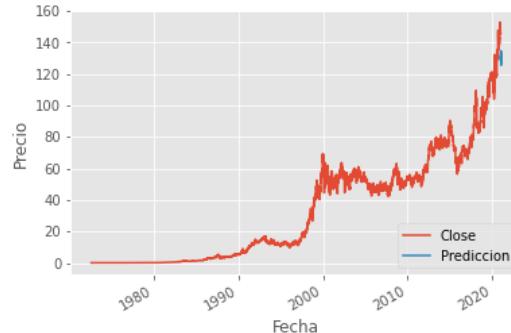


Gráfica 132. Test 1-Correlación Lineal 2 -Walmart

Test 2:

Atributos: test\_size=0.4

**Resultados:** Score of the Linear Regression Model 0.9677783509572567

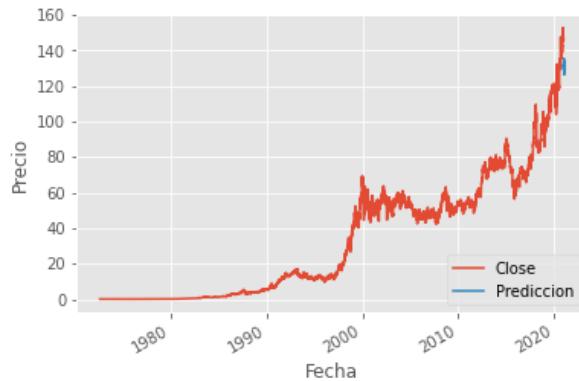


Gráfica 133. Test 2-Correlación Lineal 2 -Walmart

Test 3:

Atributos: test\_size=0.6

Resultados: Score of the Linear Regression Model 0.9668094180678853



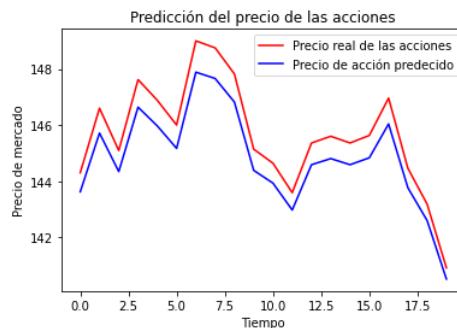
Gráfica 134. Test 3-Correlación Lineal 2 -Walmart

-Prueba con programa de Redes Neuronales 1:

Test 1:

Atributos: epochs = 200

Resultados: error: 0.8213996964735799

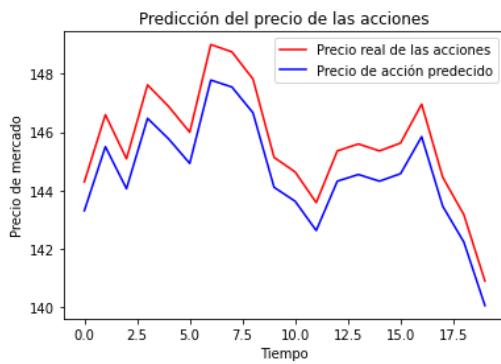


Gráfica 135. Test 1-Redes Neuronales 1 -Walmart

Test 2:

Atributos: epochs = 500

Resultados: error: 1.0547385927819648

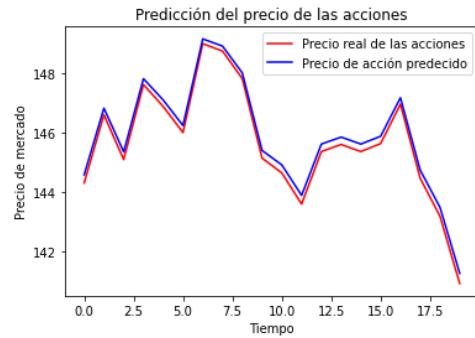


Gráfica 136. Test 2-Redes Neuronales 1 -Walmart

Test 3:

Atributos: epochs = 750

Resultados: error: 0.24826549553799138



Gráfica 137. Test 3-Redes Neuronales 1 -Walmart

### **-Prueba con programa de Redes Neuronales 2:**

Test 1:

Atributos: epochs = 200

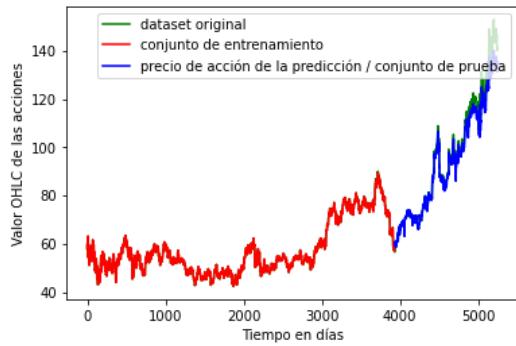
Resultados:

Entrenamiento RMSE: 0.63

Prueba RMSE: 3.76

Valor del último día: 131.44273376464844

Valor al día siguiente: 116.17157745361328



Gráfica 138. Test 1-Redes Neuronales 2 -Walmart

Test 2:

Atributos: epochs = 100

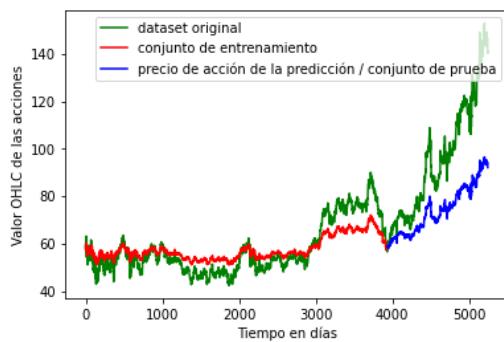
Resultados:

Entrenamiento RMSE: 5.89

Prueba RMSE: 25.80

Valor del último día: 92.10829162597656

Valor al día siguiente: 45.11955642700195



Gráfica 139. Test 2-Redes Neuronales 2 -Walmart

Test 3:

Atributos: epochs = 50

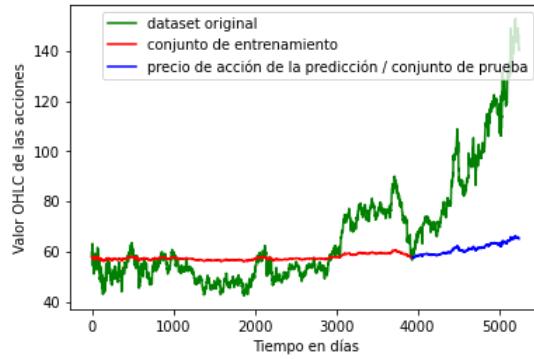
Resultados:

Entrenamiento RMSE: 9.54

Prueba RMSE: 40.94

Valor del último día: 65.14729309082031

Valor al día siguiente: 14.031723976135254



Gráfica 140. Test 3-Redes Neuronales 2 -Walmart

#### 14.11 Anexo K. Ejemplo de Base de datos (Apple)

Tabla 15. Ejemplo de Base de datos (Apple)

Date	Open	High	Low	Close	Adj Close	Volume
12/12/1980	0.128348	0.128906	0.128348	0.128348	0.101073	469033600
15/12/1980	0.12221	0.12221	0.121652	0.121652	0.0958	175884800
16/12/1980	0.113281	0.113281	0.112723	0.112723	0.088768	105728000
17/12/1980	0.115513	0.116071	0.115513	0.115513	0.090965	86441600
18/12/1980	0.118862	0.11942	0.118862	0.118862	0.093603	73449600
19/12/1980	0.126116	0.126674	0.126116	0.126116	0.099315	48630400
22/12/1980	0.132254	0.132813	0.132254	0.132254	0.104149	37363200
23/12/1980	0.137835	0.138393	0.137835	0.137835	0.108544	46950400
24/12/1980	0.145089	0.145647	0.145089	0.145089	0.114256	48003200
26/12/1980	0.158482	0.15904	0.158482	0.158482	0.124803	55574400
29/12/1980	0.160714	0.161272	0.160714	0.160714	0.126561	93161600
30/12/1980	0.157366	0.157366	0.156808	0.156808	0.123485	68880000
31/12/1980	0.152902	0.152902	0.152344	0.152344	0.119969	35750400
02/01/1981	0.154018	0.155134	0.154018	0.154018	0.121288	21660800
05/01/1981	0.151228	0.151228	0.15067	0.15067	0.118651	35728000
06/01/1981	0.144531	0.144531	0.143973	0.143973	0.113377	45158400
07/01/1981	0.138393	0.138393	0.137835	0.137835	0.108544	55686400
08/01/1981	0.135603	0.135603	0.135045	0.135045	0.106347	39827200
09/01/1981	0.142299	0.142857	0.142299	0.142299	0.112059	21504000
12/01/1981	0.142299	0.142299	0.141183	0.141183	0.11118	23699200
13/01/1981	0.136719	0.136719	0.136161	0.136161	0.107225	23049600
14/01/1981	0.136719	0.137277	0.136719	0.136719	0.107665	14291200

#### 14.12 Anexo L. Resultados general de modelo de correlación lineal 1

1. Apple

*Tabla 16. Correlación lineal 1-Apple*

<b>test_size</b>	<b>random_state</b>	<b>Score of the Linear Regression Model</b>
0.25	42	0.4574994775948822
0.35	50	0.47377626032801745
0.45	60	0.4577790125994289

2. Amazon

*Tabla 17. Correlación lineal 1-Amazon*

<b>test_size</b>	<b>random_state</b>	<b>Score of the Linear Regression Model</b>
0.25	42	0.553508491147729
0.35	50	0.5448903078838201
0.45	60	0.553248747941792

3. Berkshire Hathaway

*Tabla 18. Correlación lineal 1-Berkshire Hathaway*

<b>test_size</b>	<b>random_state</b>	<b>Score of the Linear Regression Model</b>
0.25	42	0.8274888429037911
0.35	50	0.8276653121217437
0.45	60	0.8261837220491746

4. Facebook

*Tabla 19. Correlación lineal 1-Facebook*

<b>test_size</b>	<b>random_state</b>	<b>Score of the Linear Regression Model</b>
0.25	42	0.9362192549227766
0.35	50	0.9345774054625563
0.45	60	0.9323610655283233

5. Alphabet

*Tabla 20. Correlación lineal 1-Alphabet*

<b>test_size</b>	<b>random_state</b>	<b>Score of the Linear Regression Model</b>
0.25	42	0.8466574526173501
0.35	50	0.8422815261157491
0.45	60	0.8391468413123021

6. Johnson & Johnson

*Tabla 21. Correlación lineal 1-J&J*

<b>test_size</b>	<b>random_state</b>	<b>Score of the Linear Regression Model</b>

0.25	42	0.774811124242707
0.35	50	0.7843288053584068
0.45	60	0.7793808080439903

7. Microsoft

Tabla 22. Correlación lineal 1-Microsoft

test_size	random_state	Score of the Linear Regression Model
0.25	42	0.540241869964517
0.35	50	0.542474832689299
0.45	60	0.5427033518608348

8. Procter & Gamble

Tabla 23. Correlación lineal 1-P&G

test_size	random_state	Score of the Linear Regression Model
0.25	42	0.8113327832532735
0.35	50	0.817961967262511
0.45	60	0.8130507175936332

9. Visa Inc.

Tabla 24. Correlación lineal 1-Visa

test_size	random_state	Score of the Linear Regression Model
0.25	42	0.8834415430908416
0.35	50	0.8830302496719609
0.45	60	0.882787932183045

10. Walmart

Tabla 25. Correlación lineal 1-Walmart

test_size	random_state	Score of the Linear Regression Model
0.25	42	0.8473734625027491
0.35	50	0.8489476818179207
0.45	60	0.8502902356009842

#### 14.13. Anexo M. Resultados general de modelo de correlación lineal 2

1. Apple

Tabla 26. Correlación lineal 2-Apple

test_size	Score of the Linear Regression Model
0.2	0.9808247323817118

0.4	0.9838453740255115
0.6	0.9829777290335552

2. Amazon

*Tabla 27. Correlación lineal 2-Amazon*

<b>test_size</b>	<b>Score of the Linear Regression Model</b>
0.2	0.9898938203996032
0.4	0.9886059789112015
0.6	0.9888760755268515

3. Berkshire Hathaway

*Tabla 28. Correlación lineal 2-Berkshire Hathaway*

<b>test_size</b>	<b>Score of the Linear Regression Model</b>
0.2	0.9930827864362123
0.4	0.9935796496287518
0.6	0.99337607951435

4. Facebook

*Tabla 29. Correlación lineal 2- Facebook*

<b>test_size</b>	<b>Score of the Linear Regression Model</b>
0.2	0.956004766795847
0.4	0.9611791254656632
0.6	0.9583746021108157

5. Alphabet

*Tabla 30. Correlación lineal 2-Alphabet*

<b>test_size</b>	<b>Score of the Linear Regression Model</b>
0.2	0.9810711368647088
0.4	0.9800258139653405
0.6	0.9824098581663854

6. Johnson & Johnson

*Tabla 31. Correlación lineal 2- J&J*

<b>test_size</b>	<b>Score of the Linear Regression Model</b>
0.2	0.9676117450005048
0.4	0.9676158790399497
0.6	0.9678352811191353

7. Microsoft

Tabla 32. Correlación lineal 2-Microsoft

test_size	Score of the Linear Regression Model
0.2	0.9789609871413563
0.4	0.9760197380709916
0.6	0.979424802031503

8. Procter & Gamble

Tabla 33. Correlación lineal 2-P&G

test_size	Score of the Linear Regression Model
0.2	0.9576459914678895
0.4	0.9585575806028583
0.6	0.9587024510109063

9. Visa Inc.

Tabla 34. Correlación lineal 2-Visa

test_size	Score of the Linear Regression Model
0.2	0.986053471223436
0.4	0.9846726429020765
0.6	0.9866743272875365

10. Walmart

Tabla 35. Correlación lineal 2- Walmart

test_size	Score of the Linear Regression Model
0.2	0.8473734625027491
0.4	0.8489476818179207
0.6	0.8502902356009842

#### 14.14. Anexo N. Resultados general de modelo de redes neuronales 1

a) Apple

Tabla 36. Redes Neuronales 1-Apple

Epochs	Error
200	1.7961129347195295
500	2.1701886389070557
750	1.3253522436824308

b) Amazon

Tabla 37. Redes Neuronales 1-Amazon

Epochs	Error
200	15.911142729426215
500	4.729788087593539
750	7.745100328710279

c) Berkshire Hathaway

Tabla 38. Redes Neuronales 1- Berkshire Hathaway

Epochs	Error
200	3701.2358287416564
500	695.8421687274326
750	1051.5591830138992

d) Facebook

Tabla 39. Redes Neuronales 1- Facebook

Epochs	Error
200	1.1020328090080538
500	0.8062193233275442
750	0.695361994798917

e) Alphabet

Tabla 40. Redes Neuronales 1- Alphabet

Epochs	Error
200	30.752271886562593
500	10.598201891930568
750	15.43327313451624

f) Johnson & Johnson

Tabla 41. Redes Neuronales 1- J&J

Epochs	Error
200	0.774811124242707
500	0.7843288053584068
750	0.7793808080439903

g) Microsoft

Tabla 42. Redes Neuronales 1- Microsoft

Epochs	Error

200	0.5679991031777459
500	0.9288485089356294
750	1.3904130732022832

h) Procter & Gamble

Tabla 43. Redes Neuronales 1- P&G

Epochs	Error
200	1.370565465087694
500	0.1260535385257277
750	0.4098822940982482

i) Visa Inc.

Tabla 44. Redes Neuronales 1- Visa

Epochs	Error
200	2.312904965668219
500	1.2313488619765085
750	0.3607054663621247

j) Walmart

Tabla 45. Redes Neuronales 1- Walmart

Epochs	Error
200	0.8213996964735799
500	1.0547385927819648
750	0.24826549553799138

## 14.15. Anexo N. Resultados general de modelo de redes neuronales 2

1. Apple

Tabla 46. Redes Neuronales 2- Apple

Epochs	Entrenamiento RMSE	Prueba RMSE	Valor del último día	Valor al día siguiente
200	1.44	11.62	99.15734100341797	72.0592041015625
100	7.94	44.25	27.780553817749023	5.622982978820801
50	8.87	49.49	15.459165573120117	1.7005761861801147

2. Amazon

Tabla 47. Redes Neuronales 2- Amazon

Epochs	Entrenamiento RMSE	Prueba RMSE	Valor del último día	Valor al día siguiente
200	91.71	1412.56	392.9415588378906	45.910160064697266
100	98.96	1522.81	169.2087860107422	8.289729118347168
50	100.08	1540.19	133.28379821777344	5.071072578430176

### 3. Berkshire Hathaway

Tabla 48. Redes Neuronales 2- Berkshire Hathaway

Epochs	Entrenamiento RMSE	Prueba RMSE	Valor del último día	Valor al día siguiente
200	745.18	18217.11	303950.3125	265474.75
100	771.61	21677.40	296962.65625	253212.9375
50	1506.32	18489.72	305707.25	268843.3125

### 4. Facebook

Tabla 49. Redes Neuronales 2- Facebook

Epochs	Entrenamiento RMSE	Prueba RMSE	Valor del último día	Valor al día siguiente
200	1.93	8.20	248.95394897460938	230.29002380371094
100	14.42	33.03	212.9302978515625	164.91575622558594
50	41.55	88.60	136.53993225097656	61.30923843383789

### 5. Alphabet

Tabla 50. Redes Neuronales 2- Alphabet

Epochs	Entrenamiento RMSE	Prueba RMSE	Valor del último día	Valor al día siguiente
200	20.99	143.29	1547.2540283203125	1229.76904296875
100	87.95	409.96	1144.970458984375	670.5430297851562
50	175.17	819.51	442.14617919921875	88.26483154296875

### 6. Johnson & Johnson

Tabla 51. Redes Neuronales 2- J&J

Epochs	Entrenamiento RMSE	Prueba RMSE	Valor del último día	Valor al día siguiente
200	0.29	5.91	146.1570281982422	128.91632080078125
100	0.29	6.82	143.1105499267578	123.32766723632812
50	0.28	5.01	148.38250732421875	132.95484924316406

### 7. Microsoft

Tabla 52. Redes Neuronales 2- Microsoft

Epochs	Entrenamiento RMSE	Prueba RMSE	Valor del último día	Valor al día siguiente
200	0.42	11.66	190.08883666992188	151.1627197265625
100	7.18	43.49	120.4112777709961	60.83268356323242
50	12.86	77.81	33.11177062988281	4.565129280090332

## 8. Procter & Gamble

Tabla 53. Redes Neuronales 2- J&J

Epochs	Entrenamiento RMSE	Prueba RMSE	Valor del último día	Valor al día siguiente
200	0.54	2.98	123.88084411621094	114.30765533447266
100	2.03	7.18	117.08013916015625	100.76918029785156
50	9.98	30.34	80.74797821044922	39.78200149536133

## 9. Visa Inc.

Tabla 54. Redes Neuronales 2- Visa

Epochs	Entrenamiento RMSE	Prueba RMSE	Valor del último día	Valor al día siguiente
200	0.63	12.53	179.0460968017578	157.151611328125
100	6.80	34.67	152.31031799316406	112.9681625366211
50	23.43	105.40	65.7435073852539	18.22549819946289

## 10. Walmart

Tabla 55. Redes Neuronales 2- Walmart

Epochs	Entrenamiento RMSE	Prueba RMSE	Valor del último día	Valor al día siguiente
200	0.63	3.76	131.44273376464844	116.17157745361328
100	5.89	25.80	92.10829162597656	45.11955642700195
50	9.54	40.94	65.14729309082031	14.031723976135254