

Proyecto Lenguajes de Programación

Alejandro Hahn Gallegos

Daniel Armando Nuñez Delgadillo

27/11/2020

Instituto Tecnológico y de Estudios Superiores de Monterrey

Introducción

Para el proyecto de la clase de Lenguajes de Programación se definió crear una IA para realizar una tarea que el alumno decida. Se usará el lenguaje de Python y varias herramientas para trabajar con Machine Learning que se especifican en las siguientes secciones.

Definición de Proyecto

Existen distintas opciones en las que se pueden aplicar las bondades del aprendizaje automático, las opciones que están siendo consideradas son las siguientes:

- Aplicación a un sistema de recomendaciones que analice las preferencias de un usuario y clasifique si una película le gustará o no.
- Aplicación a la predicción de precios y recomendaciones en relación al mercado de acciones en base a métricas de las compañías y precios históricos.
- Aplicación al análisis de las partidas de ajedrez para predecir el siguiente movimiento del jugador.

Tecnologías Empleadas

Python

Es un lenguaje de programación frecuentemente usado en machine learning y creación de IA, debido a ser uno de los lenguajes de programación más populares y su facilidad de uso, gran variedad de bibliotecas y su facilidad de probar código sin implementarlo.

Anaconda

Es una distribución abierta para Python y R que se usa en machine learning para procesar una grandes volúmenes de datos, análisis predictivo y cómputos científicos.

Jupyter

Es una organización que desarrolla software de código abierto para la informática científica y la ciencia de datos. Soporta principalmente a los lenguajes Julia, Python, y R. Ha creado ya varios productos como Jupyter Notebook, JupyterHub, JupyterLab y Jupyter Kernels.

Numpy

Numpy es una librería programada para Python que funciona como una extensión del lenguaje para añadir soporte a operaciones con vectores y matrices, ya que Python no fue diseñado para el uso de cálculos numéricos.

Pandas

Pandas es una librería programada para Python, que funciona como una extensión de NumPy que provee utilidades y operaciones para tablas numéricas y series temporales. Es principalmente utilizado para el análisis de datos, en especial para importar datos de archivos, seleccionar datos y limpiarlos.

Matplotlib

Matplotlib es una librería programada para Python, que se encarga de generar gráficos a partir de series los datos. Se pueden generar gráficos de líneas, histogramas, gráficos de dispersión, gráficos 3D, entre otros.

Tensorflow

TensorFlow es una biblioteca de código abierto para aprendizaje automático a través de un rango de tareas, y desarrollado por Google para satisfacer sus necesidades de sistemas capaces de construir y entrenar redes neuronales para detectar y descifrar patrones y correlaciones, análogos al aprendizaje y razonamiento usados por los humanos.

Keras

Keras es una librería de redes neuronales de código abierto escrita sobre Python diseñada para ser ejecutada sobre TensorFlow, Microsoft Cognitive Toolkit o Theano. Sus fuertes se centran en ser amigable para el usuario, modular y extensible.

Scikit-Learn

Scikit-learn (anteriormente scikits.learn) es una biblioteca para aprendizaje automático de software libre para el lenguaje de programación Python. Incluye varios algoritmos de clasificación, regresión y análisis de grupos entre los cuales están máquinas de vectores de soporte, bosques aleatorios, Gradient boosting,

K-means y DBSCAN. Está diseñada para interoperar con las bibliotecas numéricas y científicas NumPy y SciPy.

Tema o área a desarrollar

Se escogió analizar la predicción de precios y la clasificación de las acciones que cotizan en la bolsa de valores de Nueva York donde se busca revisar los movimientos realizados del 2010 al 2016 y que forman parte del indicador Standard & Poors 500, que contiene a las 500 empresas de mayor capitalización de Estados Unidos.

Dataset seleccionado

Se seleccionó el dataset “New York Stock Exchange Standard & Poors 500 companies historical prices with fundamental data” (Gawlik, 2017). Este conjunto de datos recorre las cotizaciones diarias al cierre del mercado desde el 2010 hasta el 2016, incluyendo máximos y mínimos, así como los volúmenes. El dataset ya se encuentra preprocesado, incluyendo los precios ajustados a los splits que sufrieron las acciones. También cuentan con información técnica que puede ser de utilidad para mejorar la precisión de la predicción.

Referencia repositorio de extracción del dataset

Gawlik Dominik. (2017, febrero). New York Stock Exchange S&P 500 companies historical prices with fundamental data, Version 3. Fecha de consulta: 2 de noviembre del 2020 desde <https://www.kaggle.com/dgawlik/nyse/metadata>.

Descripción de las columnas en el dataset

Las columnas del dataset son las siguientes:

- Columnas del dataset original.
 - date: Fecha del registro.
 - symbol: Abreviación de la empresa.
 - open: Precio por acción al inicio del día.
 - close: Precio por acción al final del día.
 - low: Precio por acción mínimo durante el día.
 - high: Precio por acción máximo durante el día.
 - volume: cantidad de acciones vendidas.
 - eps: Cálculo de las ganancias por acción.
- pe: Ratio entre el precio por acción y la ganancia por acción.
- yield1: Porcentaje de cambio del precio por acción a un mes.
- yield3: Porcentaje de cambio del precio por acción a 3 meses.
- yield6: Porcentaje de cambio del precio por acción a 6 meses.
- yield12: Porcentaje de cambio del precio por acción a un año.
- ma50: Media de precio por acción a 50 días.

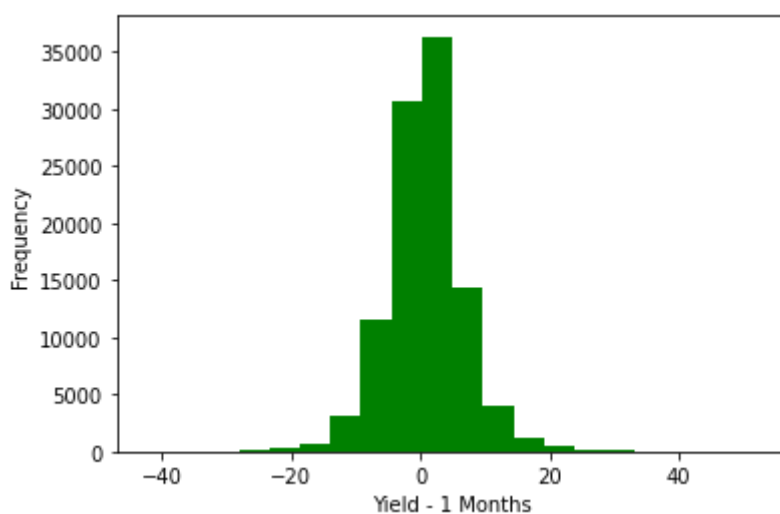
- ma200: Media de precio por acción a 200 días.
- ma300: Media de precio por acción a 300 días.
- rec: Recomendación de comprar una acción.
- term: Periodo de validez de la recomendación en meses.
- gain: Ganancia esperada en porcentaje.

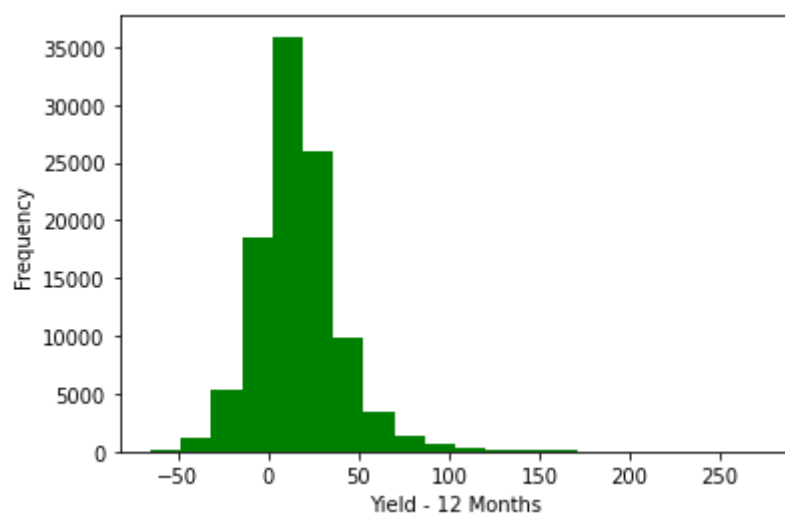
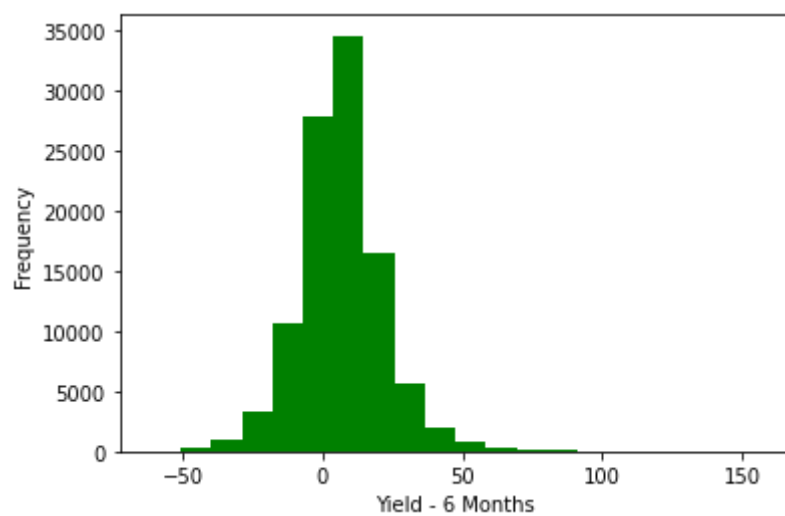
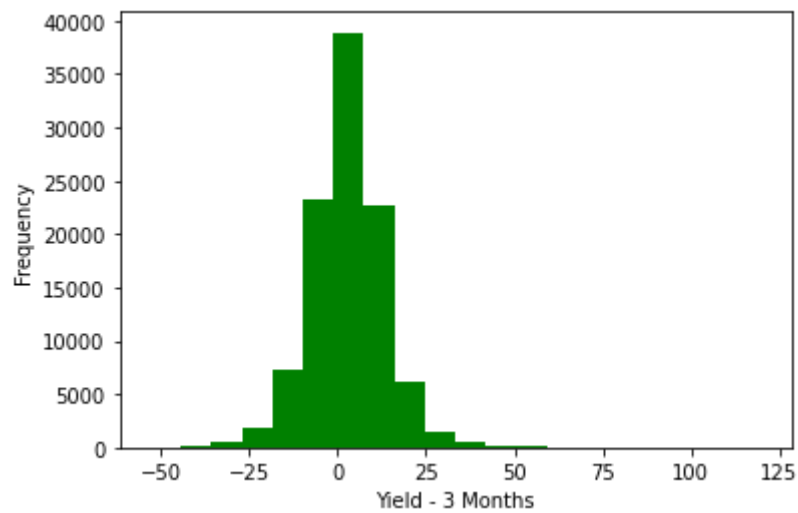
Proceso de limpieza del dataset

Se descartaron empresas del dataset que presentaban poca o nula información así como aquellos registros que por cuestiones de fechas era imposible utilizarlas para llevar a cabo el análisis. Se creó la columna del ratio del precio por ganancia (pe) que se obtiene dividiendo el precio por acción entre el ganancias por acción (eps). Después se calcularon los porcentajes de cambio del precio por acción (yield) en periodos de 1, 3, 6 meses y un año, usando el precio al cierre de la acción dividido en el precio al cierre de la fecha determinada y multiplicando por 100 para obtener el porcentaje, y después, se le restó 100 para obtener el cambio ya sea positivo o negativo. Seguidamente se calculó la media del precio por acción tomando la media del precio de cierre en un periodo de 50, 200 o 300 días anteriores al de la acción. se asignó un periodo de 3, 6 o 12 meses a cada acción en la columna term y a su vez se le asignó la posible ganancia esperada del 5%, 10%, 15% o 20% en la columna gain. La columna rec tiene el valor binario que indica si es recomendable comprar (1) o vender (0) la acción en base al periodo y al porcentaje esperado de ganancia indicados en la fila de la acción.

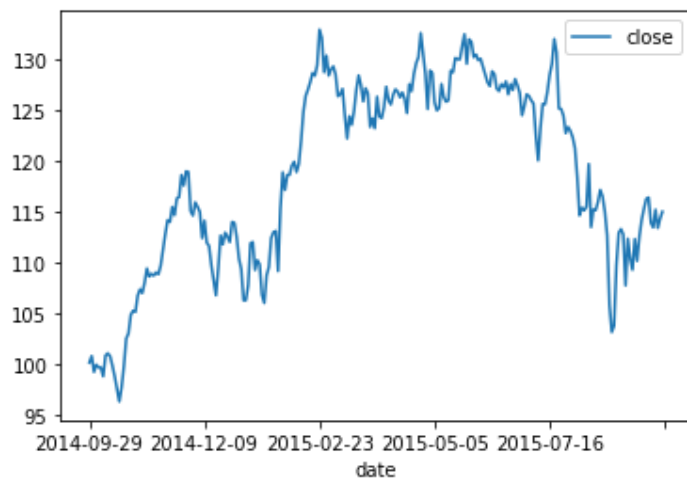
Análisis Exploratorio de los Datos

A continuación se presentan una serie de gráficos obtenidos del dataset procesado y limpiado. Las siguientes 4 gráficas representan el cambio porcentual de todas las acciones del dataset con respecto a 1 mes, 3 meses, 6 meses y 1 año respectivamente:

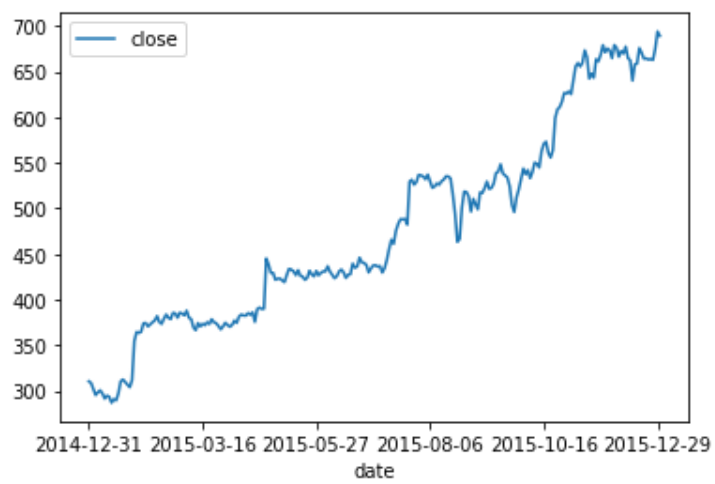




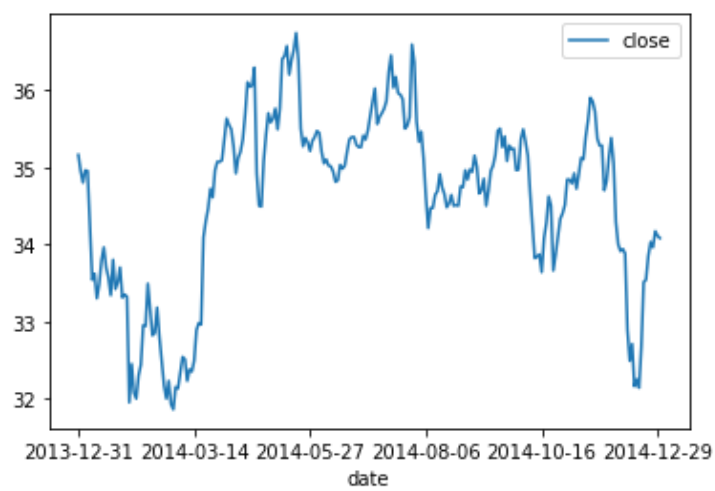
Las siguientes gráficas presentan el cambio del precio por acción con respecto al tiempo de las empresas Apple, Amazon y AT&T.



Apple



Amazon



AT&T

Descripción del error

De acuerdo con las métricas arrojadas por el proceso de entrenamiento de la red neuronal, el modelo tiene una precisión de 90.90%, un error cuadrático promedio de 0.4328 y un error absoluto promedio de 0.5. Todos los anteriores dentro de los estándares para un buen modelo de inteligencia artificial.

Conclusiones

En conclusión el uso de inteligencia artificial puede ser implementado en muchos ámbitos de nuestra vida, en este caso se optó por ver su uso en el mercado de acciones, se analizaron miles de acciones de cientos de empresas para poder predecir si es viable comprar o vender una acción determinada en un periodo determinado de tiempo si no cumple con tus ganancias esperadas. Estas técnicas de análisis destacan por poder encontrar los patrones que transcurren en la normalidad, patrones que solucionan problemas y patrones que nos ayudan a vivir mejor. Podemos decir que el uso de la inteligencia artificial está transformando al mundo en el que vivimos y será usada como una herramienta para mejorar no solo el análisis de acciones, sino nuestra vida diaria.

Referencias

Anaconda (distribución de Python). (2020, 7 de julio). *Wikipedia, La enciclopedia libre*. Fecha de consulta: 03:06, septiembre 19, 2020 desde [https://es.wikipedia.org/w/index.php?title=Anaconda_\(distribuci%C3%B3n_de_Python\)&oldid=127535376](https://es.wikipedia.org/w/index.php?title=Anaconda_(distribuci%C3%B3n_de_Python)&oldid=127535376).

Gawlik Dominik. (2017, febrero). New York Stock Exchange S&P 500 companies historical prices with fundamental data, Version 3. Fecha de consulta: 2 de noviembre del 2020 desde <https://www.kaggle.com/dgawlik/nyse/metadata>.

Keras. (2020, 26 de agosto). *Wikipedia, La enciclopedia libre*. Fecha de consulta: 02:28, noviembre 17, 2020 desde <https://es.wikipedia.org/w/index.php?title=Keras&oldid=128778152>.

Matplotlib. (2020, 27 de abril). *Wikipedia, La enciclopedia libre*. Fecha de consulta: 03:08, septiembre 19, 2020 desde <https://es.wikipedia.org/w/index.php?title=Matplotlib&oldid=125552725>.

NumPy. (2020, 16 de abril). *Wikipedia, La enciclopedia libre*. Fecha de consulta: 03:05, septiembre 19, 2020 desde <https://es.wikipedia.org/w/index.php?title=NumPy&oldid=125248885>.

Pandas. (2019, 22 de noviembre). *Wikipedia, La enciclopedia libre*. Fecha de consulta: 03:09, septiembre 19, 2020 desde <https://es.wikipedia.org/w/index.php?title=Pandas&oldid=121498623>.

Proyecto Jupyter. (2020, 14 de septiembre). *Wikipedia, La enciclopedia libre*. Fecha de consulta: 03:06, septiembre 19, 2020 desde https://es.wikipedia.org/w/index.php?title=Proyecto_Jupyter&oldid=129268399.

Python. (2020, 14 de septiembre). *Wikipedia, La enciclopedia libre*. Fecha de consulta: 03:06, septiembre 19, 2020 desde <https://es.wikipedia.org/w/index.php?title=Python&oldid=129272218>.

Scikit-learn. (2020, 8 de noviembre). *Wikipedia, La enciclopedia libre*. Fecha de consulta: 02:33, noviembre 17, 2020 desde <https://es.wikipedia.org/w/index.php?title=Scikit-learn&oldid=130746122>.

TensorFlow. (2020, 17 de septiembre). *Wikipedia, La enciclopedia libre*. Fecha de consulta: 02:44, noviembre 17, 2020 desde <https://es.wikipedia.org/w/index.php?title=TensorFlow&oldid=129340039>.