
Springboard Data Science Career Track

Jesse Mailhot

Mentored by Tony Paek

Capstone Project 2

Milestone Report 1

Predicting Microsoft Stock Price Movement With LSTM

Introduction

Forecasting stock price movement is an important part of financial management. Using predictions from a machine learning model can help financial advisors make more informed decisions, thereby creating more value for the company and the client. The goal of this project is to build a Long Short Term Memory (LSTM) recurrent neural network to predict the daily, weekly, monthly, and quarterly movement in Microsoft stock price, then compare it to yearly predictions using Facebook's Prophet additive regression model.

Goals

1. Gather data using the Alphavantage API
2. Clean and prepare the data for use in the LSTM and Prophet models
3. Train the models and generate predictions

Dataset

The dataset will be comprised of Microsoft adjusted closing price, and daily high obtained using the Alphavantage API.

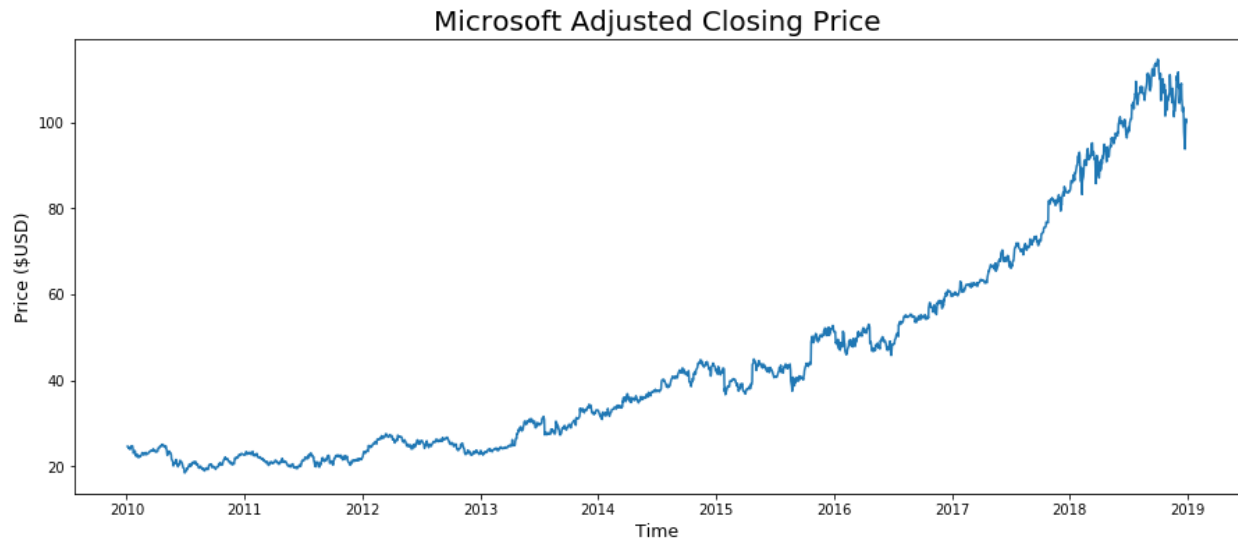
Data Wrangling

A custom function was created to obtain the stock price data using the Alphavantage API. These functions use the requests package to extract the data in json format, load it into a pandas dataframe and return the data for the specified time frame. Once the data has been sliced into the desired time frame, it is loaded into a CSV file and stored in the raw data folder for exploratory data analysis. The dataset contains 2263 rows of Microsoft data from 1/4/2010 - 12/28/2018.

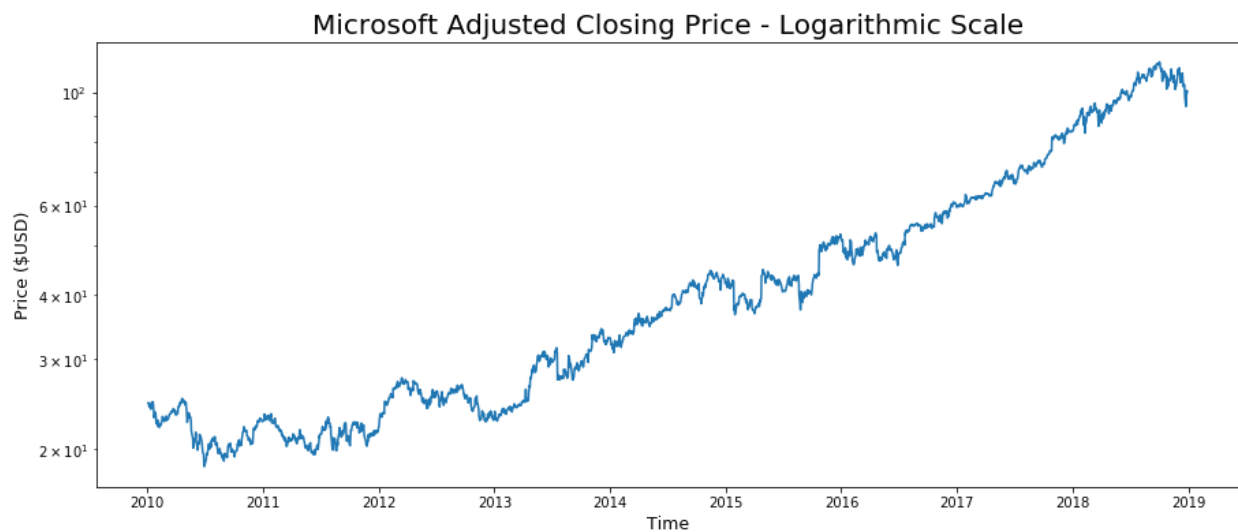
Data Analysis

Raw Data

The raw price data for Microsoft adjusted closing price has a clear long term upwards trend, but short term trends such as daily and weekly movement are much more random. Such movement is normal for stock prices as the markets fluctuate.



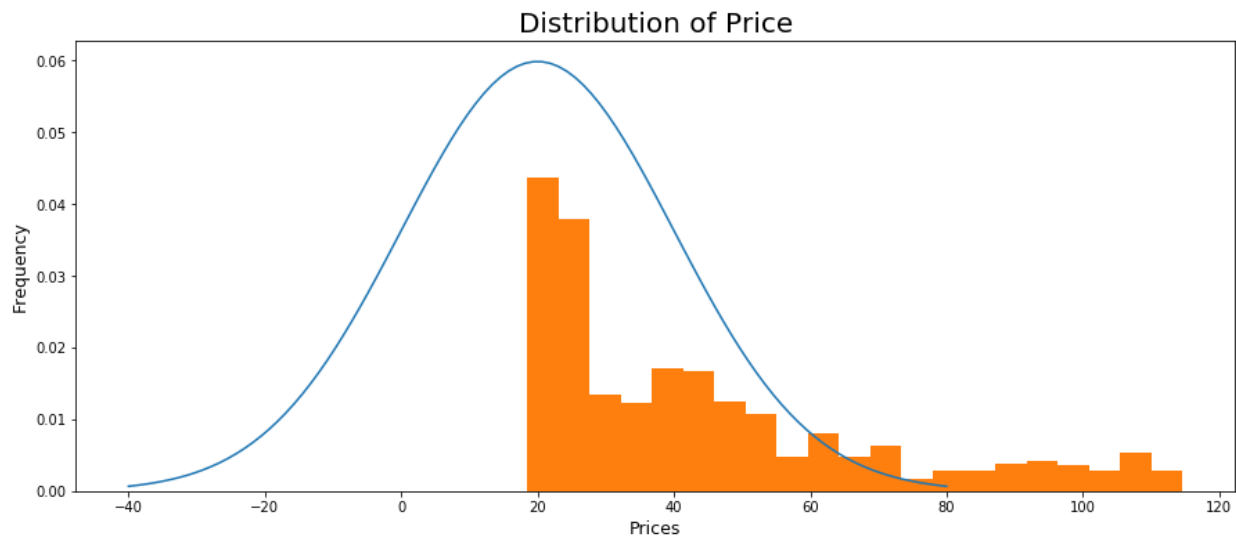
To get a clearer picture of the long term trend, it is advantageous to view the data on a logarithmic scale.



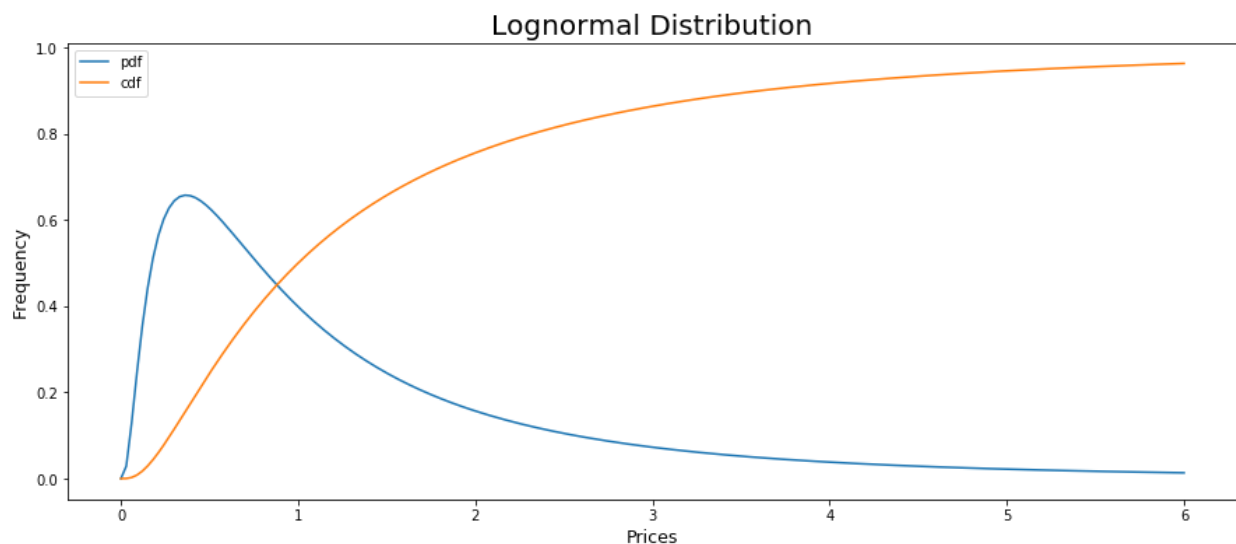
The long term trend of price increase is clearer; however, the short term movement is still very jagged.

Distribution of Price

Examining the distribution of prices, it is clear the data is not normally distributed.



Additionally, price alone is meaningless because it derives its current value from the value of prior periods. Instead of using historical prices to predict future prices, I will use return periods; however, return periods are also not normally distributed. To solve this problem, I will use logarithmic returns instead of traditional holding period returns.



Using logarithmic returns means the data will resemble a lognormal distribution, which more accurately describes daily price movement for financial time series since most daily movement is relatively small. The intuition behind using logarithmic returns is that it is equivalent to infinitely compounding returns. The logarithmic return also represents the instantaneous rate of change of return, which describes the current direction the stock is moving in, which is ultimately what the model will try to predict.

Next Steps

1. Normalize the data by using lognormal returns
2. Prepare the data for use in an LSTM model
3. Train the models and analyze the results