ML Pipeline (sklearn, mongo, quandl, airflow)

Jeffery Scully

June 17, 2021

Contents

Quandl api key - 6UfyTfL7MDo26ypfYzVd Include pydoc Include pip freeze output Include try blocks

1 Train LinearRegression Model

Using Quandl, an API that provides stock price data. The python libarary returns a pandas dataframe.

Getting Data for Google stock prices from DATE - DATE. Will train model using LinearRegression, and use it to predict future stock prices.

Data: Open: High: Low: Close: Volume: Ex-Dividend: Split Ratio: Adj. Open: Adj. High: Adj. Low: Adj. Close: Adj. Volume:

2 Define Variables

2.1 DB table column - $\mathbf{HL_{PCT}}$

2.2 DB table column - $\mathbf{PCT_{change}}$

- This is the percent change of closing price verses the opening opening price

$$\text{PCT\_change} = \text{Closing\_Price - Opening\_Price}\frac{}{Opening\_Price \times 100}$$

2.3 Python Variable - X

2.4 Python Variable - $\mathbf{X_{lately}}$

2.5 Python Variable - y

2.6 Python Variable - $\mathbf{X_{train}}$

2.7 Python Variable - $\mathbf{y_{train}}$

2.8 Python Variable - $\mathbf{X_{test}}$

2.9 Python Variable - $\mathbf{y_{test}}$

3 How LinearRegression Works

4 Save Model, using Pickle

5 Create