Capstone Project 2

CME Futures Markets EDA of tick real-time stream, with feature development and possible investment/trading strategizing

[Inception, ideas, goals, constraints, definitions- with appended after-the-fact Results and pointers]

Explore interdependencies and orderly correlating structures within market daily activities in rea-ltime datasets of the most liquid Futures markets in the world: S&P 500 and NQ (NASDAQ). Develop indicators, features, correlates useful for trading/investment.

Criteria for success:
[the Results had been appended, post initial Capstone project definitions, for quick gauge of accomplished]
- Adopting datasets for real-time processing from tick data-stream. Results: *Work with time series datasets, filed derivatives, new column like OHLC candles at various intervals (from tick data: to: 1 sec, 2 sec, 1min, 3 min, 10, min), work with, from up to 2 Million rows, reducing it to daily minutes of 1,440 rows, or 86,400 seconds. Further NaN, and inf values work, with cleaning, clearing and reducing/converting datasets with meaningful addoptions of derived features for further model processing.*
- Finding correlations between derived features. Results: *Many correlates found, the process identified BA streams that suggest Market direction.*
- Finding Islands/Clusters of Order. Results: *BA tick data, when standardized, become Oscillator-like indicator, reflecting hidden Market tendencies. Cluster BA spikes, discovered in proprietary BA oscillator suggest future tops/bottoms and further directional bias, especially when combined with another Correlate and Time. TBC.*
- Running the datasets through several ML models, like Linear Regression, Decision Tree, NN, CNN, etc. Results: *Completed! Actual results are not*

suited for investment purposes, however, although offer solid leads toward further development of 'Market Edge'.

- TBD: Projecting Dialable Time Intervals in a given Direction onto a plot (chart) real time, updatable with stream data, including all functions to be capable of uptake and process/show/project results of model. Make Decisions (Up/Down) within a constraint of: 50% less Drowdown then Predicted Profit. <u>Results:</u> Huge undertaking, a *problem suited for a team of developers with a 1 year completion horizon. However, strong correlates were discovered; the initial real-time charting engine has been set in working condition, charting metrics and visually charting a unique proprietary Oscillator reflecting streaming data activity, reflecting order book activities with strongly suggestive properties of Market directional bias.*

Scope of solution space:
2-3 month deadline to finish. <u>Results:</u> Completed within deadline (3 months).

EDA ideas and general Model directions of interest:

1. ES (S&P500) Futures Markets (CMA) daily data set analysis. 4 features EDA.
(data sets: available)
Features:

    1. Time: HH MM SS MM; Hour, Minutes, Seconds, MicroSeconds
    2. Volume
    3. Price
    4. BidAsk (containing 4 types: BelowBid, Bid, Ask, AboveAsk) . Normalize?
Standardize or Normalize the set? <u>Results:</u> *Done!*

2. Stratified Analysis of 10-20 different days: [<u>Results:</u> Found many different derivatives, many of which are not useful for investment purposes, some of which show tremendous potential, and 1 of which shows direct reflective bias of Market Direction.]
        - Find day types?
        - Discover Correlations, Interdependencies <u>Results:</u> Within Jupyter nb.

3. Create 5 minute independent feature, as well as 10 min (of user choosable) Interval for IntervalQuery. Project it in the Future and make it a Target Dependent Y feature. Move it in time, and project it onto plot, with a function that runs/updates each Dialable Interval (eg: 30 sec). TBD: this is more suited either for the 2 Capstone Project or for a team of many developers. Dataframes were, however, adopted for such further work in real-time, in any set interval and functions and procedures defined and executed for such interval conversions.

Find out: Characteristics between the 4 variables,
- General classifications of the Intervals Query Clusters analysis:
- Correlations between each 2 of the 4 / 5 features
- Clustering between each 2 of the 4 / 5 features
  Results: Within Jupyter nb.

4. The final Model should also be able to update with a time-series updating real-time market data stream and plot its findings on an updating plot. Results: *Realtime streaming chart visualizations from tick dataset to Pandas df, with indicators has been accomplished. Making trades, strategies for developing a winning investment strategy is a complex work requiring teams of developers and time. But solid ground foundations have been built around proprietary Oscillator developed during this Capstone.]*

5. Try 2-3 trading Strategies:
    1.
        - Parlor strategy Negative (adding to losers under certain constraints) TBD (complex strategy for future development]
        - Parlor strategy Positive (adding to winners under certain constraints)
What is the Outcome? [in better models, current simple models offer not good results. Metrics available in Jupyter nb's]

    2.   Analyze Market's behavior near 100 SM, 200 SM, VWAP.
Results: [*TBD:] [During this Capstone it has been found that this type of strategizing needs better than simplistic models, current ML models offered not investment good results. However, datasets had been worked with, prepped, cleaned, processed, split, passed through pipeline or models and metrics derived. Available in repositories at Github. Metrics and visuals also available in*

Jupyter nb's at GH. Looking forward to complex deep NN's or Reinforcement Model development, hopefully in the last Capstone 3 or in future projects.]
   Results: Within Jupyter nb.

6. All visualizations available in Jupyter nb's at GitHub repo.

Constraints:

Drawdown of Predictions has to be 50% less then predicted Moves
SHould work and update realtime and react within a 30 sec delay of the real market data stream.

Stakeholders:

Myself, Mentor, School

Data sources:

CME, private market data subscription with API mining in C#, solid data transfer, [discovered later that occasional data gaps do happen with the need to reload datasets]. Csv file reading real-time from API streaming tick data.

Ideas for Capstone Project 3 or further development of the Market analysis work:

1. ES (S&P500) Futures Markets (CMA) daily data set analysis. 4 features EDA. *(data sets: available)*
   Stratified Analysis of 10-20 different days:
   - Find day types?
   - Discover Correlations, Interdependencies, Entropy metrics
   - 5 minute clusters analysis/ 10 min clusters analysis: (new features of   5/10 min interval)
   - Correlations between 4 / 5 features
   - The column of open - high - low + close should describe the central tendency of retrace. Make 1,3,5 min.
     <u>Results:</u> *See Jupyter nb's.*