EDA Insights Summary

Exploratory analysis of data showed that there are a number of essential features of the data that led to the configuration of the feature and the design of the model. The most notable trend between the data sets was the high relationship between price-based indicators, such as short and long-term moving averages, average true range and the Width of the Bollinger Bands. These characteristics had almost linear correlation not only with close price, but also among themselves, indicating extensive redundancy, and indicating that they would tend to record similar underlying trend information as opposed to independent information. MACD and the trend line that accompanies it were no exceptions and they were following the same trend in terms of being constructed on the basis of the exponential moving averages.

Indicators that are based on momentum like the RSI system, and the rate-of-change indicators showed lower correlations with price and volatility characteristics. These indicators showed short-range changes in the strength and overbought or oversold status, which give supplementary data and not a duplicate of trend signals. Their actions indicated that they were useful as regime conditions and not the main market direction drivers. Valumes and volume derivatives ratios showed low results and were not correlated with both price and technical data. Although there were spikes in volumes that were accompanied by sharp price changes, these relationships were not very strong, and it supports the understanding of volume as a form of secondary confirmation and not a predictor by its own.

The study of individual stock price dynamics revealed significant heterogeneity of the market behavior. Others showed good and sustainable positive trends, whereas others showed oscillating trends or range. A few of the stocks had sharp downfalls and then part recovery showing the existence of more than one market regime in the dataset. This variety aided in the conceptualization of this problem as a category classification exercise as opposed to a directional forecasting.

This conclusion was also supported by RSI behavior. There were always swings in the traditional 30 to 70 scale, with long upsurges over 70 in line with rising trends and downwards slaps in line with contractions before recoveries. Nevertheless, RSI failed to identify accurate turning points all the time and it is important to note that it is (Precise) turning points that are weaknesses of this timing tool with its greater strengths in terms of a situational indicator. The analysis of sector level distribution of returns showed that there were significant variations in volatility where energy stocks had the largest dispersion and consumer oriented sectors had a comparatively stable behavior. Lastly, the target labels were reasonably well balanced amongst uptrend, downtrend, and sideways classes, which alleviated the issue of class imbalance and facilitated the reasonable multi-class assessment. Taken altogether, these results attest to the fact that the dataset can be largely used when regimes-based modeling is to be performed under the influence of the strength of trends and volatility predominantly.