Sentiment Analysis Data Preparation

* News articles for Tesla were not available for every trading day
* Filled missing sentiment analysis with hysteresis prior to regression analysis

Monte Carlo Regression Analysis

A monte carlo analysis was performed on 2 sentiment analysis data frames (All Sentiments and Sentiments by News Source) to determine which feature (sentiment analysis) has the best correlation to Tesla’s daily returns. The analysis was performed using two different monte carlo functions which iterated over 20 days.

* Model 1: Lasso Regression
  + <https://towardsdatascience.com/ridge-and-lasso-regression-a-complete-guide-with-python-scikit-learn-e20e34bcbf0b>
* Model 2: Lasso Regression using LogisticRegression
  + <https://towardsdatascience.com/feature-selection-using-regularisation-a3678b71e499>

Monte Carlo Results

The charts below show the overall sentiment analysis winner overlaid with Tesla’s daily returns for each analysis ran. It appears that using Model 1 on the News Source data frame results in the best correlation to Tesla’s daily returns. However, the correlation is not strong enough to make trading decisions.

|  |  |  |
| --- | --- | --- |
|  | Model 1  Lasso Regression | Model 2  Lasso Regression using LogisticRegression |
| All Sentiments | A screenshot of a cell phone  Description automatically generated  Coeff: 0.086 | A screenshot of a cell phone  Description automatically generated  Coeff: 0.457 |
| Sentiments by News Source | A screenshot of a cell phone  Description automatically generated  Coeff: 0.112 | A screenshot of a social media post  Description automatically generated  Coeff: 0.577 |

The charts below identify the winning feature (sentiment analysis) per iteration for each of the monte carlos performed. From analyzing the data, we confirmed that the overall winner shown on the hvplots isn’t necessarily the feature that won the most times over the 20 day iteration. For the All Sentiment data frame, the textblob sentiment analysis has the most correlation to Tesla’s stock returns. For the Sentiment by News Source data frame, Model 1 predicts that Motley Fool has the best correlation to Tesla’s stock returns, whereas Model 2 predicts Zacks has the best correlation.

|  |  |  |
| --- | --- | --- |
|  | Model 1  Lasso Regression | Model 2  Lasso Regression using LogisticRegression |
| All Sentiments |  |  |
| Sentiments by News Source |  |  |

Enhancement Opportunities

* Run analysis over a longer period of time (current analysis is only for 9 months)
* Data mine for additional articles on stock for more sentiment analysis data
* Further customization of sentiment analysis on given article
  + Determine if article was written for Tesla vs Tesla being mentioned in the article
* Integrate volume of sales data and sell ‘type’ (industry vs retail) into regression analysis