Version 1.12

AI Analysis of EDGAR

AI Analysis of the Norges Bank Investment Management Positions in EDGAR

1/31/2021

Dave Brown
Founder, SignalPop LLC

Contents

Overview

SignalPop has collected and analyzed, the Norges Bank Investment Management (NBIM) US equity positions held over a ten-year period as reported by the US Securities and Exchange Commission's EDGAR database. The data collected includes nearly all 13F filings from mid-2010 through 2019, as well as all 10K and 10Q filings that were available for each of the companies held by NBIM over this period.

Our motivation for this research was driven by our desire to demonstrate how SignalPop's AI software can analyze extensive data sets to better understand complex and otherwise indiscernible relationships within the portfolio to improve the overall investment decisions. Our vision at SignalPop is that AI is here to empower humans, not replace them. In that spirit, our detailed report shows how AI analysis can be an integral part of the investment process.

Throughout this document the data and process used are discussed in detail, showing how our models take into account the the thousands of financial line items making up the EDGAR data schema, and our proprietary process allows us to learn from each company regardless of industry or size.

This document is organized in the following sections:

Results; reveals objective results from data driven SignalPop analysis of NBIM holdings.

Data; describes extensive data sets sourced for the analysis.

Process; describes how AI was used to analyze and get the results found.

Risks; provides a summary of AI risks and how they can be mitigated.

DISCLAIMER

This document and its contents are provided for informational purposes only and are not intended, nor should be construed or relied on, as investment advice in any way. We are an AI research company and provide this report solely to show the potential of using AI. We do not provide financial or investment advisers.

The analysis contained in this document was produced using data from public sources, such as the EDGAR database. We have not independently verified any such data.

SignalPop disclaims all liability arising from any use of this document or its contents.

AI Results Summary

After training the SignalPop AI models on around 40,000 10Q/10K filings covering over a decade of time, our models are able to predict when NBIM will increase a position with an accuracy of over **70%** throughout this time period.

Reverse engineering these models reveals finanical line items highly correlated to the decisions made. For example, the 'Operating Lease' related financial line items produce a strong signal correlating to increasing a position – we believe that this may be related to oil and gas positions. In another example, the 'Number of Restaurants' also shows a strong signal related to position increases. These are just a few of the thousands of financial line items showing correlations with decisions made as reported in the 13F filings.

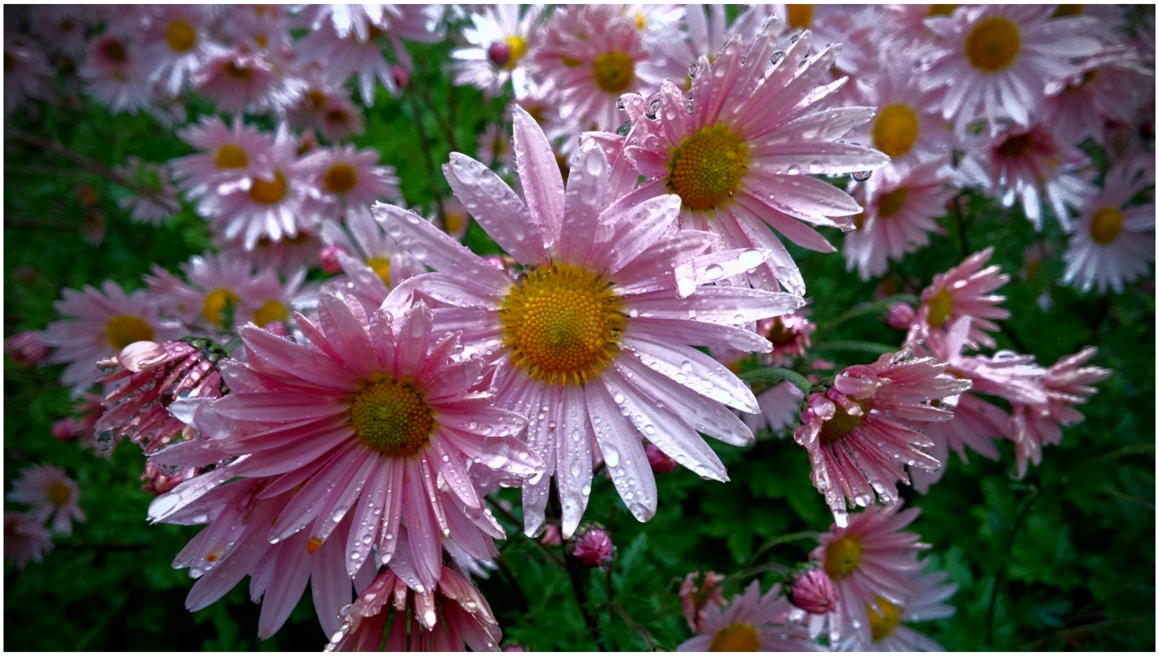
SignalPop believes that these models may be useful in the following main ways:

1.) The SignalPop AI model results provide objective, data driven analysis that may supplement, enhance some traditional methods for position change decisions. In our process, we follow the data, where it leads us, while striving to eliminate any biases.

2.) Our models find hundreds of key financial line items within each 10K/10Q that are highly correlated to the decisions reflected in the 13F filings. Such correlations provide very targeted information at the financial line item level that may provide valuable insights not available through other means.

3.) The resulting correlative analysis between the data sets and position change decisions offer opportunities to refine the investment process by adding alternative equity selections otherwise not considered or eliminating positions that would have otherwise been overlooked.

The following sections describe the data and process used to obtain the AI results achieved to date.



Data

As the basis for our analysis, SignalPop used the equities listed under the 13F filings from 6/30/2010 through 12/31/2019 (filings during the 2020 period were unavailable at the time of this report). Out of the equities listed during this time, we found over 2,700 that had associated 10Q and 10K filings. The remaining equities were either from external funds or other entities that did not have 10Q/10K filings, or those filings were not available.

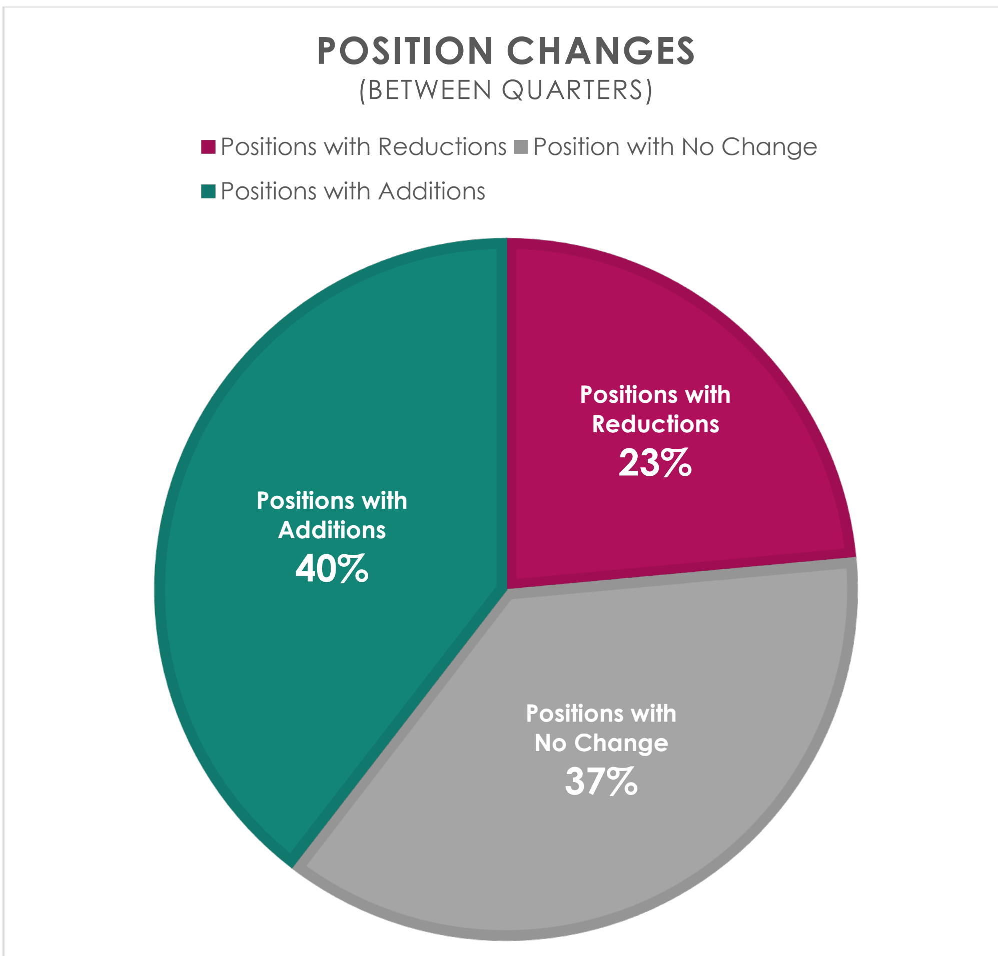
From over 2,700 equities, most 10Q/10K filings were collected for a total of over 57,000 filings, each of which are used to produce data inputs used in the analysis. For each filing, the current quarter filing was compared against the previous one, or two quarters, and then validated and normalized. Validation was performed to ensure that each filing for each equity had enough meaningful data to analyze and normalization was performed to ensure that we could compare one equity to another regardless of their size. Since the validation process naturally reduced our pool of data, we were able to increase the pool by also adding a comparison of the current quarter with the average of the previous two quarters of data. Each training set includes 80% of the items randomly selected from the dataset and the remaining 20% were assigned to the separate and distinct testing set.

Signals for each model were produced by calculating the percentage change in shares held between quarters, and between a given quarter and the average number of shares held over the previous two quarters.

Data Statistics

SignalPop's collection of position changes (from the 13F filings) between quarters for over 2,700 equities produced a set of over 57,000 data points where each data point comprises the differences between quarters. However, a portion of these data points did not have enough common data within their 10Q and/or 10K filings and were therefore not considered. This reduced the initial dataset to nearly 45,000 data points.

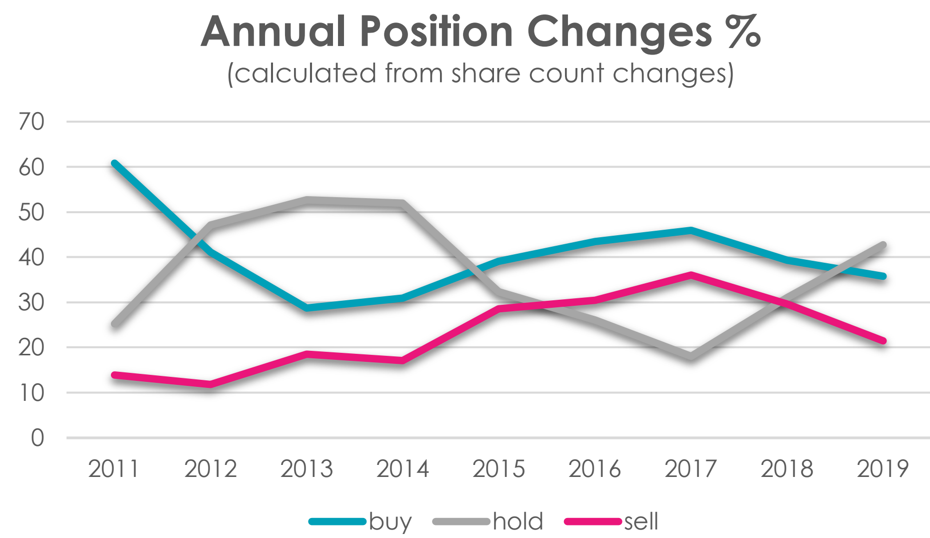
These initial data points had the following break-down when grouped by position change between quarters (e.g., position reduction, no change, position increase).



As can be seen above, over the period between 2010 and 2019, there was a growth in the overall portfolio, as position additions were nearly double to position reductions.

NOTE*: These percentages, and most other metrics, will differ from your own analytics for we are only able to look at a sub-set of the overall positions because a few positions had no associated 10Q/10K data and were therefore not included in our analysis. Also* *note, that to normalize out position appreciation, we only consider the number of shares held in each position and not its overall value.*

When looking at the data changes over time, we can observe which years had more aggressive growth and turn-over.



As shown above, the years 2014 to 2017 appear to be accumulation years with churn, as the percentage of positions with shares added not only increased but maintained nearly 1.5x the percentage of positions where shares were reduced. For example, in 2017 our observed data shows 46% of the position changes as increasing to the previous position, whereas only 36% of the positions were decreasing.

Underlying this seemly simple chart, is a key to AI models in general – a large dataset that allows for the analysis in the first place. More important is the amount of data per category/label that allows the AI model to learn discernable patters and correlations.

The above chart suggests the dataset is well balanced and thus capable of teaching the AI model and revealing its behavior.

Process

All 2TB of data downloaded from the EDGAR database was parsed and placed into a structured database. The normalized values for each filing were rendered into a set of images that were each labeled based on the signal technique used (described in detail below). This process was performed using SignalPop proprietary finance plug-ins to The SignalPop AI Designer. Next, each model was trained for over 500K iterations to produce the results. The MyCaffe AI Platform [1] (a C# complete re-write of CAFFE [2]) was used for all AI operations. For more information on using The SignalPop AI Designer in finance, please see the SignalPop Technology Presentation.

Dataset Collection and Preparation

After downloading and parsing nearly all 13F, 10Q and 10K filings available over the 10-year period of analysis, we were able to collect over 12 million data values for over 2,700 equities held over 57,000 positions of the over 90,000 positions reported in the 13F filings over this time.

During the dataset creation process, only data points validated using an internal, non-biased, selection criteria were used to ensure each data point had enough data to contribute to the overall learning. The data values within each data point, comprised of all data values extracted from a given 10K/10Q filing, were used to create a single image corresponding to the 10K/10Q filing. This selection process noted above reduced the datasets to around 49,000 images for the position change set and 40,000 images for the price change set.

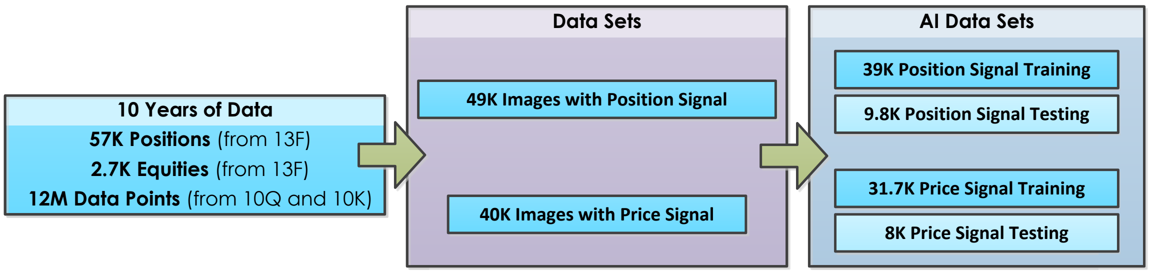


Figure 1 Dataset Creation

Data sets were then each separated into distinct training data sets and testing data sets. Training data sets were used to train each AI model and testing data sets were only used to test the model's overall accuracy.

Dataset Labeling for Position Changes

Two different labeling methods were explored to create signals that we sought to learn. In the first method, the labels were assigned based on the actions determined from the 13F filings themselves. For example, when a position in a given equity decreased (e.g., number of shares decreased), we would assign that data point with the label 0 for 'sell', and when an equity position increased (e.g., the number of shares increased), we assigned that data point with the label 1 for 'buy', and when the position did not change, we assigned the data point with the label 2 for 'hold'.

In the second method, the labels were assigned based on what the stock price of a given equity did either immediately after the 10Q/10K filing date, or immediately after the end of the quarter for which the 10Q/10Q was filed. For example, if the normalized price change over a given time period decreased, the data point was assigned a label = 0, and subsequently if the normalized price change over a given period increased over a given time period, the data point was assigned a label = 1. To determine the normalized price change over a time period, a normalized slope of the daily closing price regression line was used.

NOTE: ***In this report we only focus on the position change analysis.***

Next, with each labeling scheme, a subset of the dataset items was run through the TSNE [3] algorithm, which produces a visual representation of the dataset showing the labeled dataset's organization by label and grouping. Datasets that are very easy to learn typically show very distinct groupings around each label. With the price and position datasets we were not so lucky to have such an organization, indicating that these data sets would be more challenging to learn.

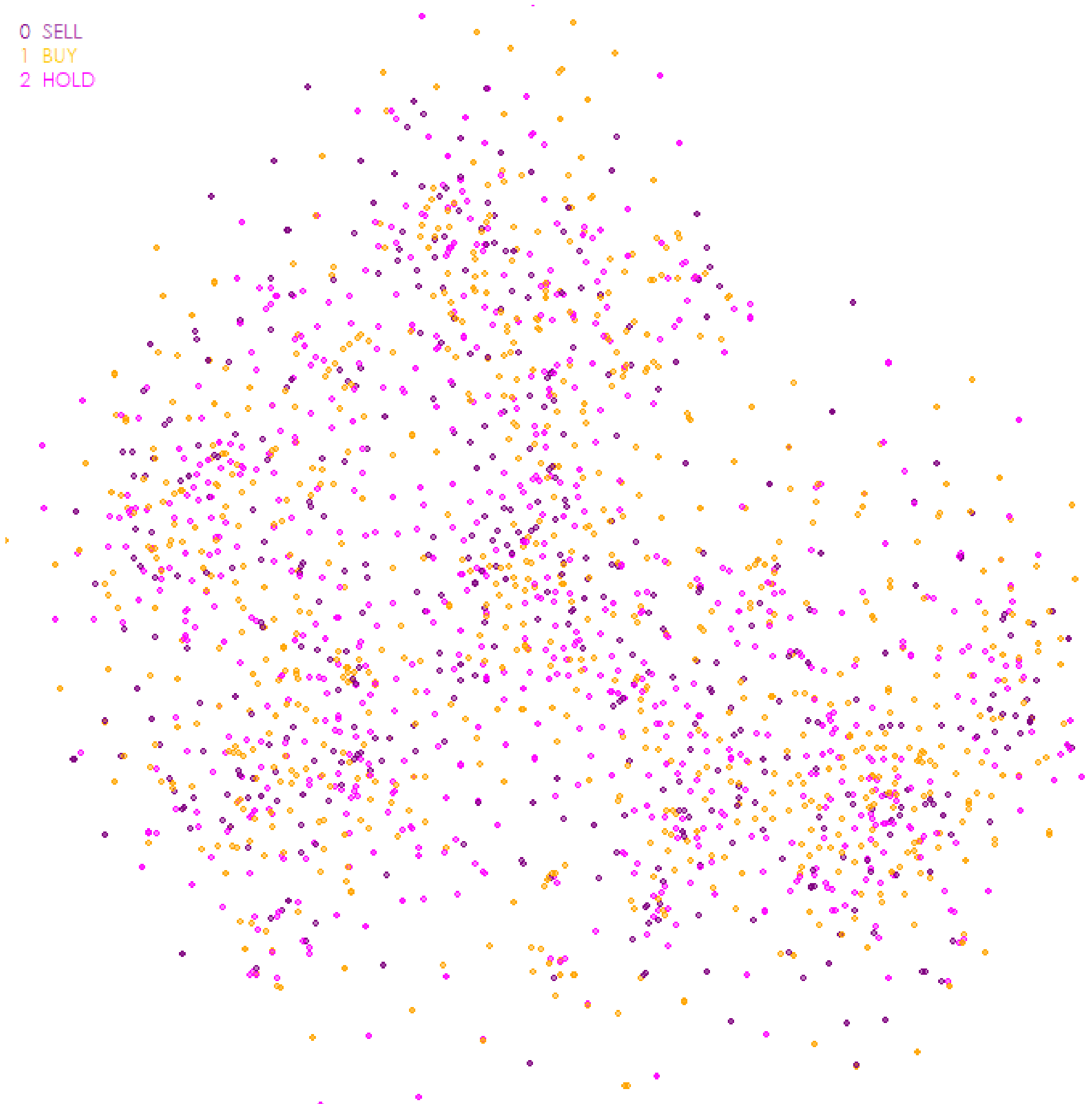


Figure 2 TSNE Analysis of SELL-BUY-HOLD Position Change Dataset

Each dot in the image above represents one data point in the dataset labeled as 0 = Sell (Purple); 1 = Buy (Orange); or 2 = Hold (Fuchsia). Data points did tend to cluster, with a slight hint of grouping around one label or another, but not in a very clear manner.

In our next test on the data learning potential, we sought to discover the data separation between labels. For this task, we subtracted each data value within each data point from the dataset mean image and placed the resulting values into one of several bucketed number ranges, where bucket collections were organized by label. The bucketed values were then compared across labels 0-1, 0-2 and 1-2, and graphed so that we could verify data differences between the data points assigned to each label.

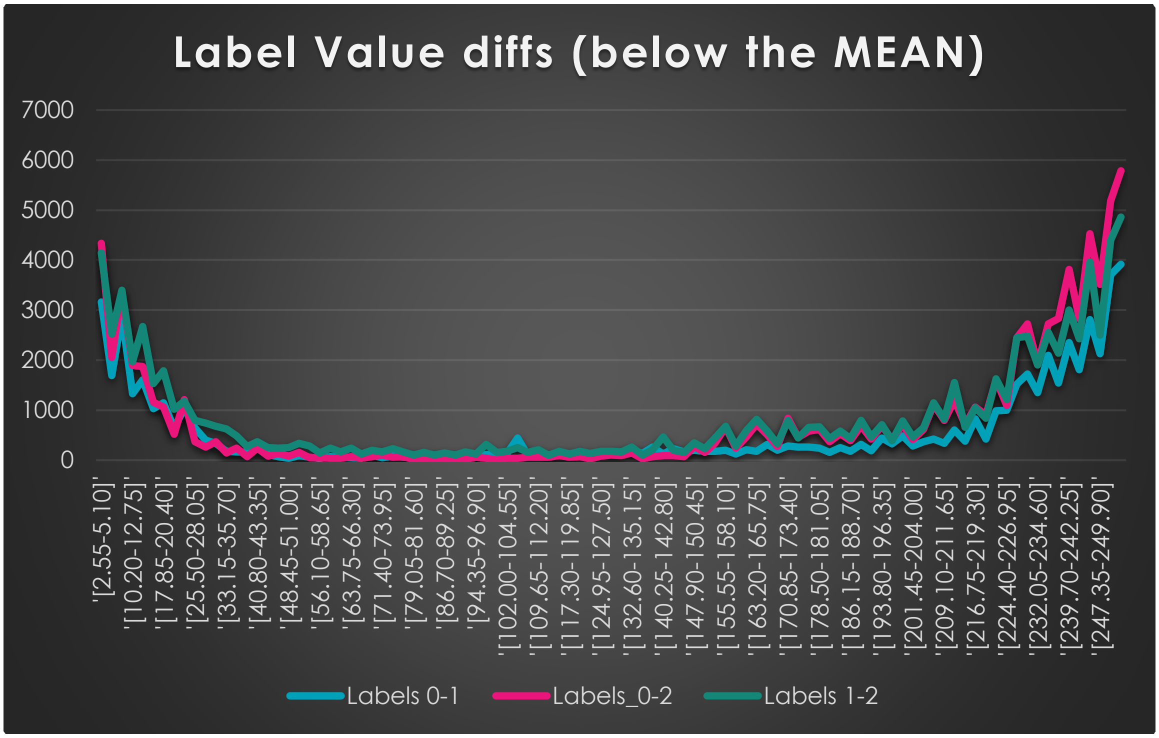


Figure 3 Data item value differences (below the MEAN) compared across labels.

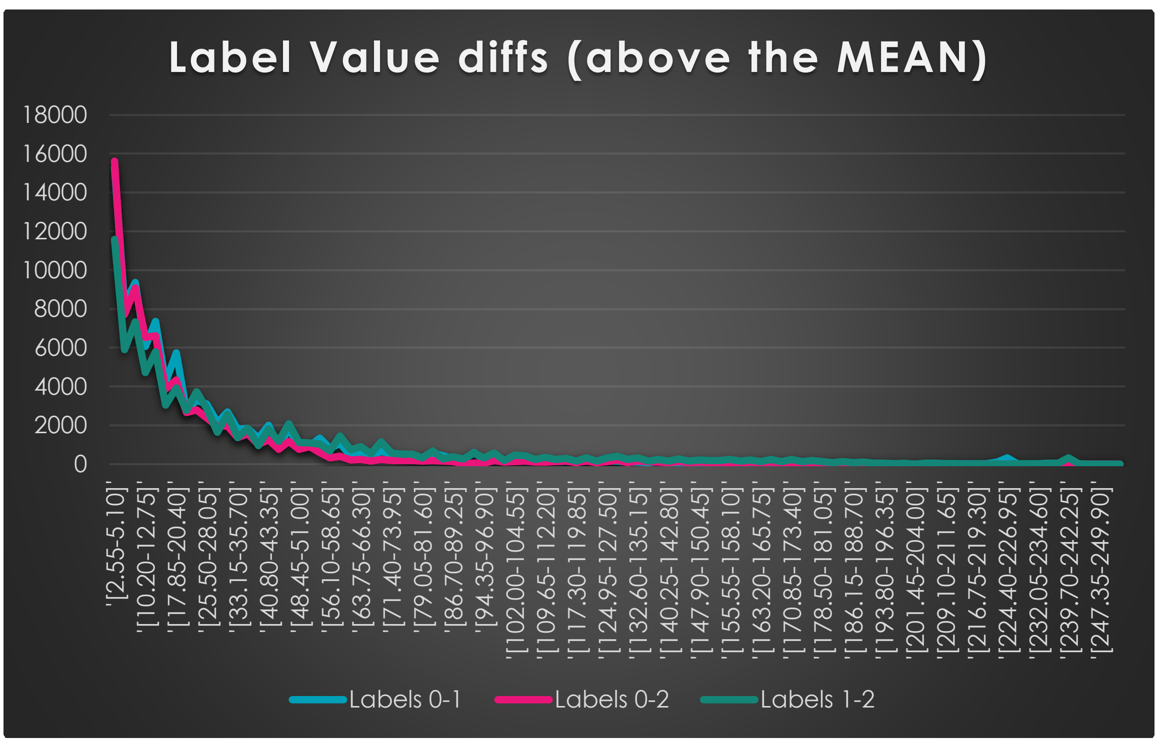


Figure 4 Data item value differences (above the MEAN) compared across labels.

These graphs and associated correlation analysis give us two important details that helped us better learn the data:

1.) The values below the mean appear have a better distribution of differences which may help speed up our learning when included, meaning that should not center the data in our preprocessing.
2.) The labels 0 (sell) and 2 (hold) are highly correlated both below (0.97) and above (0.99) the mean.

Calculating the correlation of the counts associated with each label further verifies the differences between the data items assigned to each label.

Correlations of Values below the MEAN

Label A	Label B	Correlation Coefficient
0 (sell)	1 (buy)	0.93
0 (sell)	2 (hold)	**0.97**
1 (buy)	2 (hold)	0.93

Correlations of Values above the MEAN

Label A	Label B	Correlation Coefficient
0 (sell)	1 (buy)	0.91
0 (sell)	2 (hold)	**0.99**
1 (buy)	2 (hold)	0.89

Given the high correlation between labels 0 (sell) and 2 (hold), the two labeled items were grouped together to create a final dataset where all sell/hold items were grouped into label = 0, and all buy items remained as label = 1. Doing this gave us the additional advantage of creating a binary result, which helped improve the overall learnability.

Final Position Labeling

For the final position change labeling, we grouped the sell/hold items together as one label = 0 and labeled all buy items with label = 1. We then compared and graphed the differences observed between the two labels.

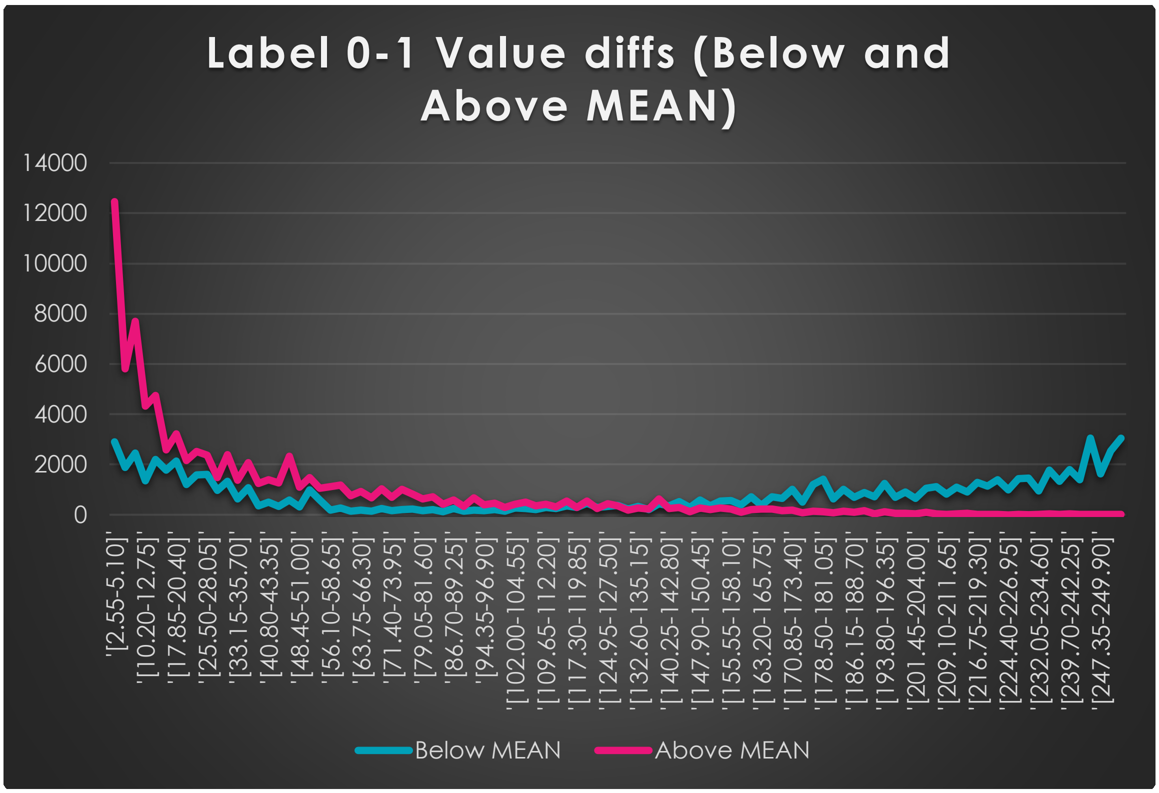


Figure 5 SELL-HOLD vs BUY Label Difference Comparison

Given the wider distribution of differences in values below the mean and relatively high correlation between changes above the mean, the dataset was not centered to make sure values below the mean were included in the learning[1].

Correlations of Values below the MEAN

Label A	Label B	Correlation Coefficient
0 (sell/hold)	1 (buy)	0.79

Correlations of Values above the MEAN

Label A	Label B	Correlation Coefficient
0 (sell/hold)	1 (buy)	**0.98**

[1] Models tested use the RELU layer, which cuts off negative values which in effect may ignore values below the mean on centered data.

AI Model Results Details

Using the data sets and labeling previously discussed, we have successfully trained several convolutional based models to learn whether the fund will Buy vs Sell/Hold an equity on a given quarter to **+70% accuracy** with a balanced +70% accuracy reached on each label 0 and 1. Furthermore, greater refinement of the dataset may increase our model learning accuracy to pinpoint specific sectors and equities therein for investment consideration. Our models were learned after training over 500,000 iterations, using various learning rate strategies.

Model Reverse Engineering

After achieving a consistent accuracy, our next question was why? What in the data triggers each buy or sell/hold decision? To find this information, we reverse engineered each model by creating an image showing the impact each pixel within the image had on firing either the buy or sell/hold label.

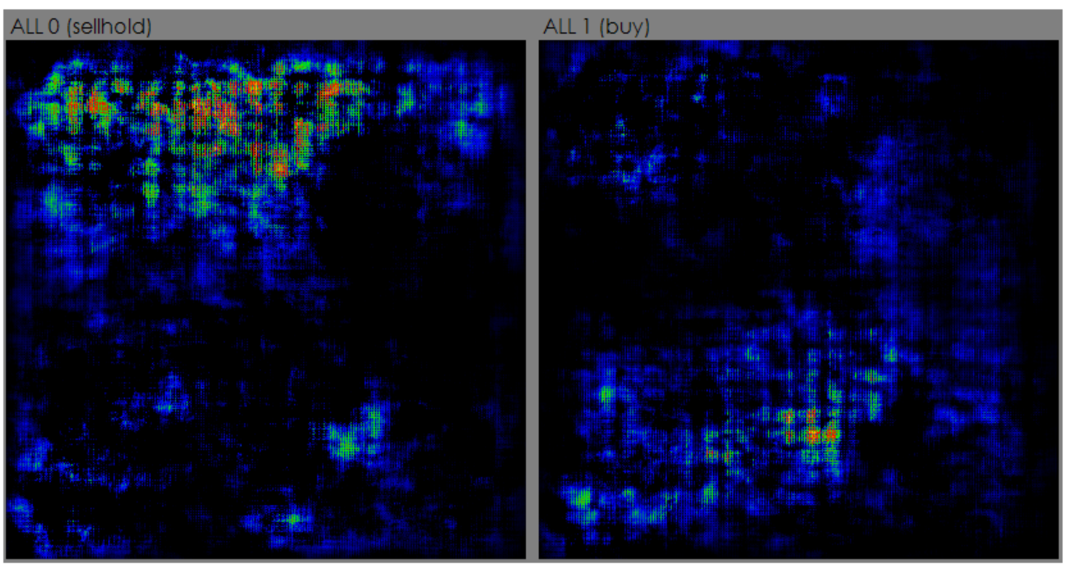


Figure 6 Model Reverse Engineering

The image on the left shows the relative strength each pixel had on firing the Sell/Hold label (0), and the image on the right shows the relative strength each pixel had on firing the Buy label (1).

Given that each pixel on the image corresponds to information extracted from each 10Q/10K filing, we took this a step further and produced an extensive Excel spreadsheet that shows the correlative relative strength of each pixel and reveals the specific corresponding line-item within the10K/10Q filing from the EDGAR database.

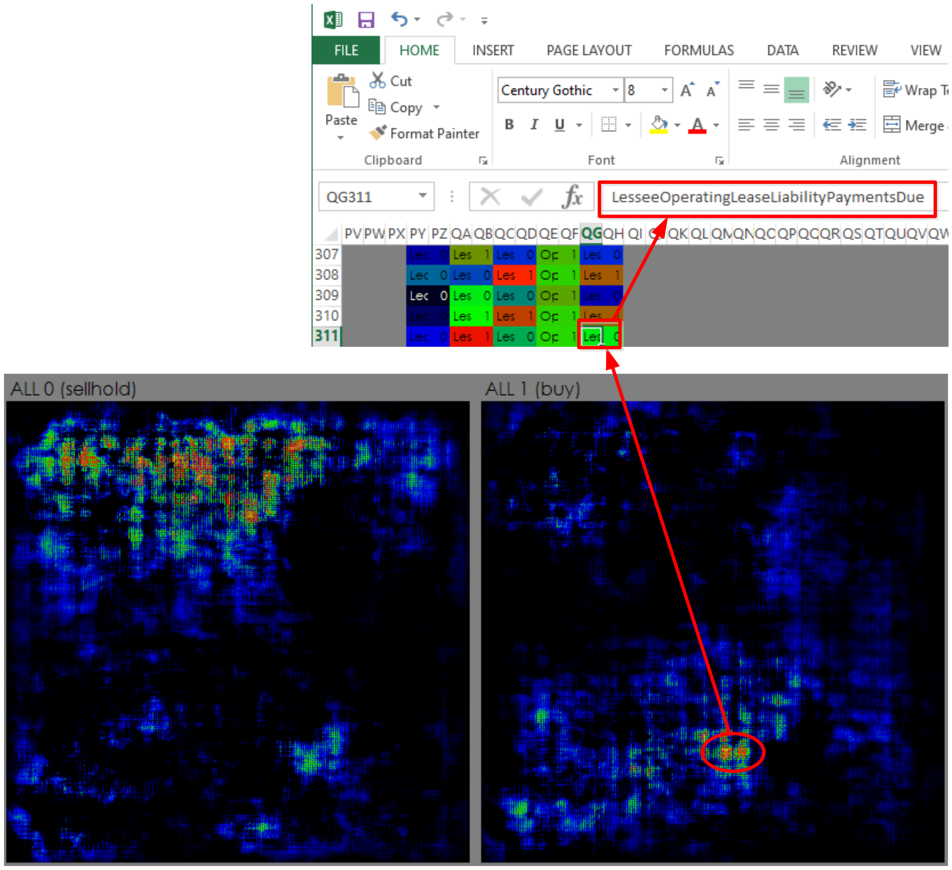


Figure 7 Model Reverse Engineering

For example, the circled hot spot shown in the ALL 1 (buy) image (right side above) represents Lease information that we presume relates to lease information from Oil and Gas company 10Q/10K filings. As shown above, the corresponding item selected on the Excel spreadsheet relates to **LesseeOperatingLeaseLiabilityPaymentDue** which according to the EDGAR database is defined as:

LesseeOperatingLeaseLiabilityPaymentDue – Amount of lessee's undiscounted obligation for lease payment for operating lease.

Each Excel spreadsheet contains thousands of similar items that show how strongly each data value contributes to the firing of a given label.

The above image hot-spot mappings were created showing how buy vs sell/hold decisions were made for all equities. However, such mappings can be created to focus on a company, sector, or industry, which may provide meaningful insights on how the objective data relates to investment decisions made.

The following hot-spot map shows the data points that triggered the **sell/hold** label (learned using nearly 40,000 position changes over 10 years).

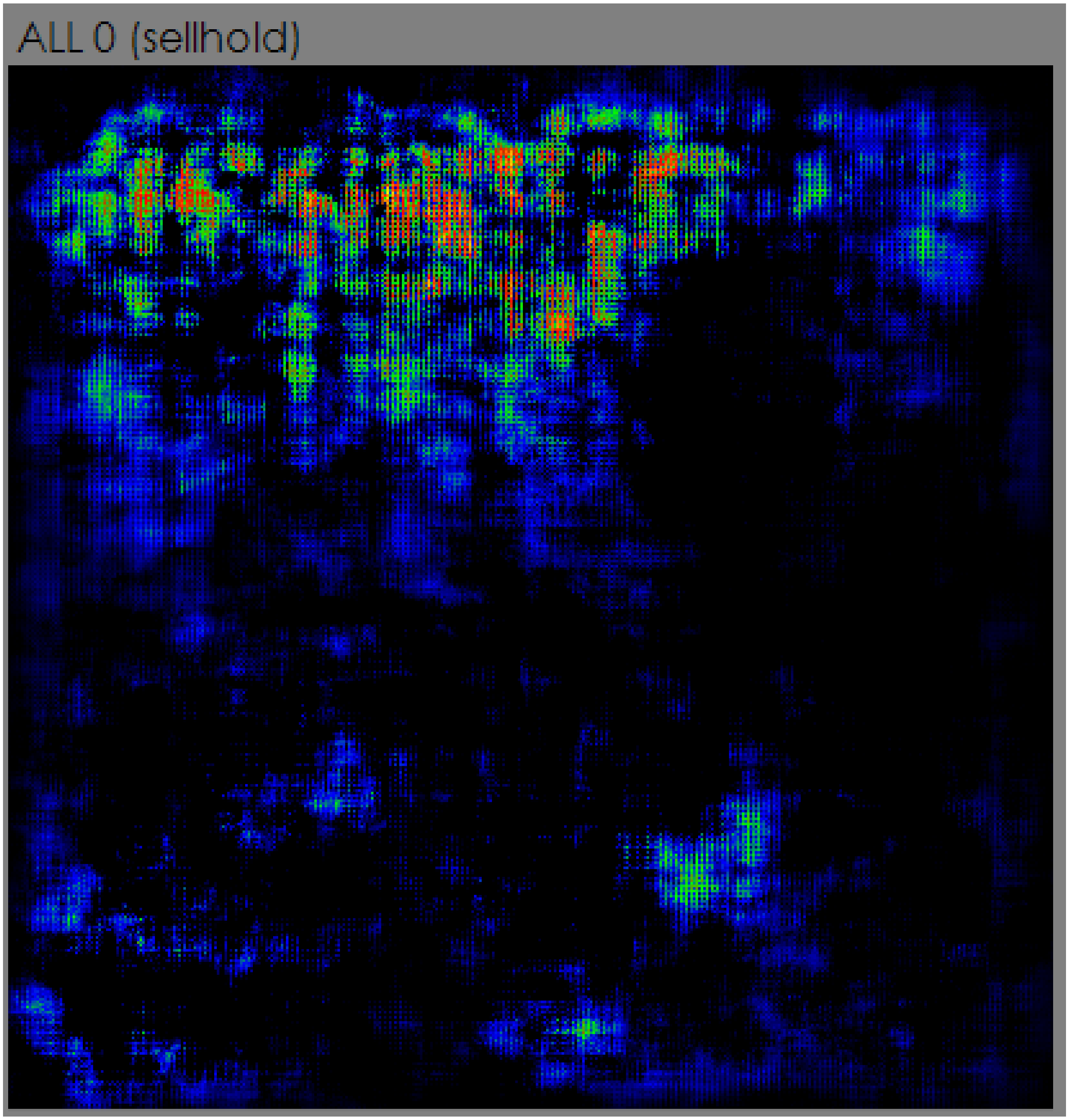


Figure 8 Sell/Hold Label Hot-Spots

Several notable hot-spot items impacting the decision to sell/hold a position are:

PaymentsToAcquireLoansReceivable
LineOfCreditFacilityInterestRateDuringPeriod
IncreaseDecreaseInDerivativeLiabilities
CapitalizedCostsOfUnprovedPropertiesExcludedFromAmortization
DebtSecuritiesAvailableForSaleRealizedGainLoss

The following hot-spot map shows the data points that triggered the **buy** label (learned using nearly 40,000 position changes over 10 years).

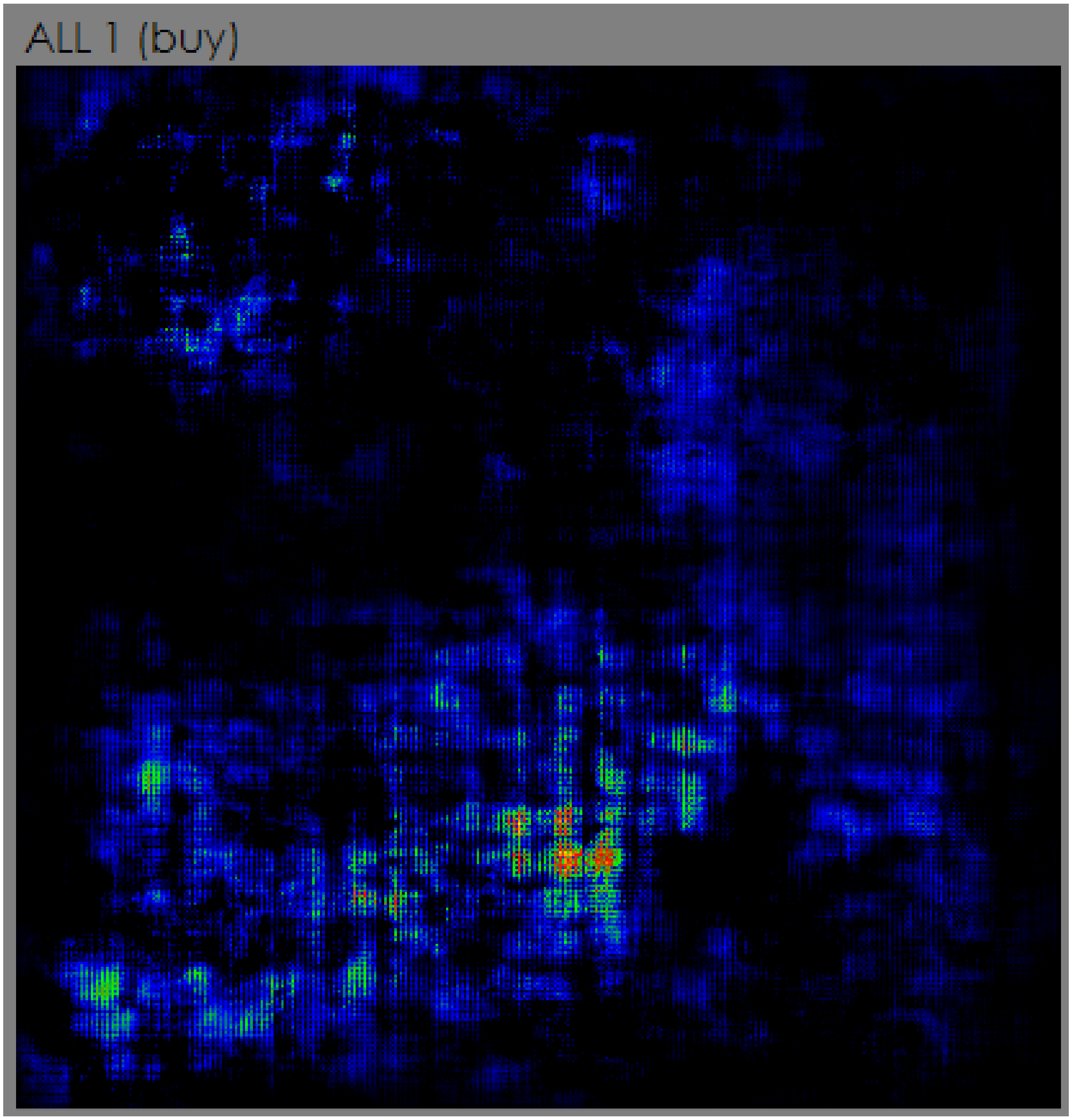


Figure 9 Buy Label Hot-Spots

Several notable hot-spot items impacting the decision to increase a position are:

IncreaseDecreaseInRiskManagementAssetsAndLiabilities
WeightedAverageNumberOfSharesContingentlyIssuable
NumberOfRestaurants
SharebasedCompensationArrangementBySharebasedPaymentAwardOptionsNonvestedNumberOfShares
LesseeOperatingLeaseLiabilityPaymentsDueYearFour

Risks

AI must be used in a responsible way to empower the decisions we make as humans and that, as humans, we must constantly double check and view each result with a skeptical eye, for there are numerous things that can go wrong in any AI solution. The following describes several potential risks and what we do to mitigate those risks.

1. **Limited Scope (Generalization)**; AI models relate closest to the data that matches what they were trained to detect. New data will not always be like the training set, so AI models should be retrained frequently with new data. This is especially true in the finance markets where data continually evolves. The models and analysis that we have provided in this report may work today, but there is no guarantee that they will work in six months. Continual re-training on new data helps each model evolve with changes in the environment of focus.
2. **Accepting Results Verbatim**; Regardless of how good an AI model appears; each result must be tested independently for verification. In the software development world, software developers are separated from software testers for good reason – the tester's job is to verify the developer's work. AI is no different. To verify each model, we ran a statistical analysis on each result to verify that they fell within the scope of the data itself.
3. **Explainable AI (XAI)**; There is always a reason each an AI solution finds the result that it does. It is our job to understand why, so that we can communicate the reason with others. Visualization features of the The SignalPop AI Designer were designed to help better understand each AI model and how it works.

Moving Forward

Several research areas related to the positions taken by funds such as the NBIM still remain to be explored. Some of these areas include the following.

1.) Deploying Single-Shot Multi-Box technology the AI models can detect specific line items within the EDGAR data set that correlate strongly to buy/sell/hold decisions focused on a single company, group of companies or sector.

2.) Triplet Net technology can improve accuracy of the data selections across a subset of companies already detected as buy/sell candidates for further analysis to refine the selection and decision-making process.

3.) SignalPop AI models can detect correlated relationships between the equity price action and the release of the financial filings for many companies. These correlations may be used as an AI input bias that help generate short-term trading decisions. Current short-term methods already employ short-term option and equity price data streams.

4.) Future modeling may be applied to larger equity pools outside the NBIM universe to analyze companies on a sector-by-sector basis and improve overall accuracy and generalization which may help find new opportunities.

Regardless of the direction of research, our AI software is up for the task.

References

[1] D. W. Brown, "MyCaffe: A Complete C# Re-Write of Caffe with Reinforcement Learning," Cornell University, 4 October 2018. [Online]. Available: https://arxiv.org/abs/1810.02272.

[2] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama and T. Darrell, "Caffe: Convolutional Architecture for Fast Feature Embedding," *arXiv,* vol. arXiv:1408.5093, 20 June 2014.

[3] L. van der Maaten and G. Hinton, "Visualizing Data using t-SNE," *Journal of Machine Learning Research,* no. 9, pp. 1-27, 2008.

[4] S. Lohr, "Netflix Competitors Learn the Power of Teamwork," The New York Times, 27 July 2009. [Online]. Available: https://www.nytimes.com/2009/07/28/technology/internet/28netflix.html?_r=1&ref=business.

SIGNALPOP LLC
PO Box 2522, White Salmon, WA, USA
www.signalpop.com