Student name: Chan Ching Kit
Student number: 20242725

# Deep dive into the ISM Report: Application on investing

Ching Kit Chan

## Abstract

This project aims to automate the process of analyzing The ISM's Report on Business, which comprises both quantitative and qualitative data from various sectors in the US. Instead of solely analyzing the quantitative numbers provided in the report, I will employ a sentiment classification model to analyze the qualitative data. This approach will enable a more comprehensive analysis of automation. Furthermore, the generated data will be compared against the performance of the corresponding market exposure to gain valuable insights for investment purposes.

## Introduction

The ISM's Report on Business is widely regarded as one of the most significant leading indicators in the financial world. Analysts rely on both quantitative and qualitative data from the report to shape their forecasts for various sectors in the US.

There are three types of ISM reports: Manufacturing PMI, Services PMI, and Hospital PMI. Among them, Manufacturing PMI holds the longest history and enjoys the highest level of recognition. These reports are issued by the ISM Manufacturing and Services business survey committees. Every month, these committees conduct surveys with hundreds of companies across different industries to gain insights into industry trends. Given the insider information provided in the report, it is reasonable to assume that a thorough analysis can lead to better investment decisions.

In this report, I have developed a Python program to fully automate the process of obtaining the original source data and analyzing it using web scraping and a text classification model developed by Huang et. al. (2022)[1].

The main objective is to gain quantitative insights into the state of different sectors, allowing for easy interpretation alongside other data. Ultimately, the aim is to gain an edge in investment decision-making using the ISM data. If the measurement and backtesting are

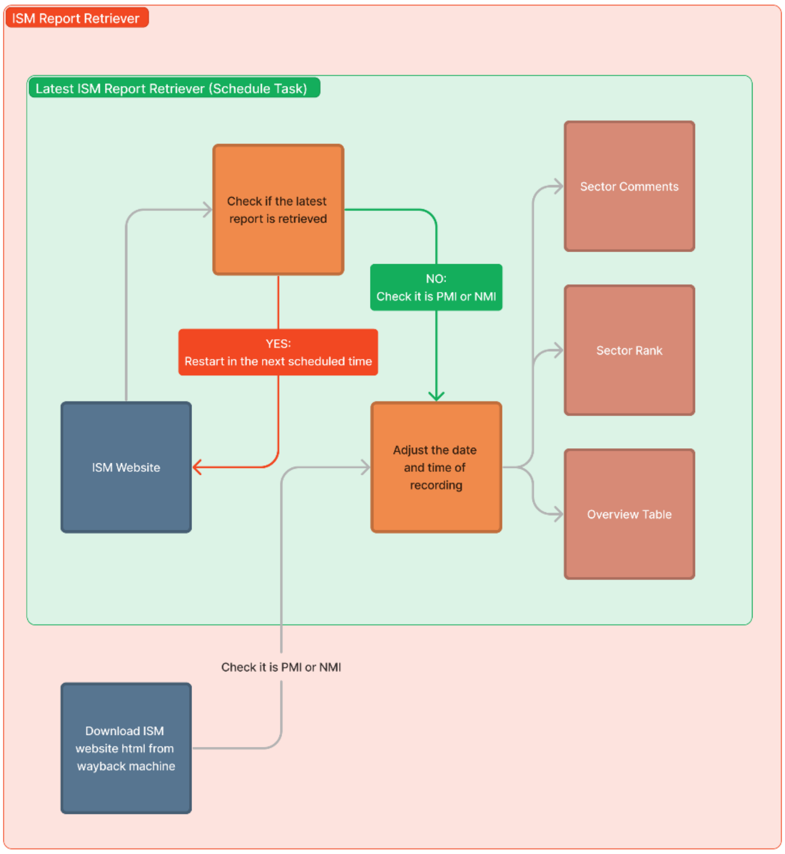
[1] On Hugging Face: https://huggingface.co/yiyanghkust/finbert-tone

conducted fairly and the model for analyzing qualitative data yields accurate results, a successful outcome could even confirm the existence of an edge provided by the ISM data. In the worst-case scenario, where no significant insights are discovered, the framework still offers an automated and quantitative means of interpreting the data.

In addition to discussing the project's objective and corresponding results, this report will also review the project's implementation, including its methodology, limitations, potential solutions, and other areas for future expansion.

## Research Methodology

The research and corresponding programs will be divided into three main parts. The first part involves creating retrievers to scrape data from the ISM website[2], specifically the Manufacturing PMI and Services PMI. The Hospital PMI will not be scraped due to the lack of qualitative data in the report and its limited coverage of sectors compared to the Manufacturing PMI and Services PMI.




In the first flow chart, the primary objective is to scrape three types of data, which include:

1. Quantitative overview table with metrics such as New Orders, Production, Employment
2. Qualitative comments from the industry practitioners
3. A ranking of the different industries based on the different metrics mentioned in (1)

Notice there is an alternative data source beside ISM Website. It is the Wayback Machine provided by the website Internet Archive[3]. In fact, I have created two retrievers for two sources, and most of the data retrieved

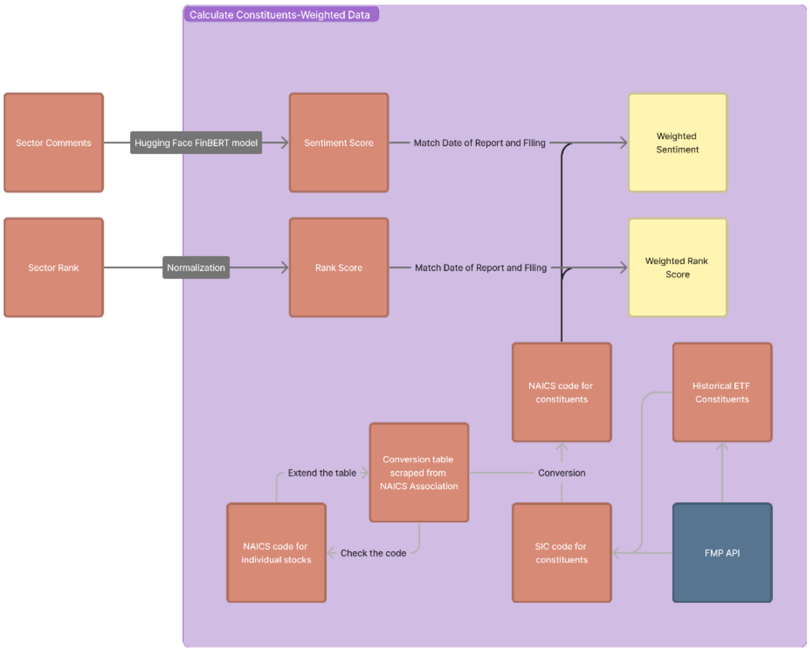
[2] ISM Report: https://www.ismworld.org/supply-management-news-and-reports/reports/ism-report-on-business/
[3] Wayback Machine: https://archive.org/web/

in this project is from this source other than the original ISM website. That is because the ISM website only provided the last two months of reports. Lack of data available is going be detrimental to the analysis if only the data from the original source is available. The Wayback Machine provides historical snapshots of the URL given. As of 12 Dec 2023, ISM data back to June 2020 are still available in their snapshots.

In the design of the first part, most of the part would be the same since the snapshots from Wayback Machine are in the same HTML structure as the original source with only a few extra HTML of Internet Archive embedded. Therefore, little to no modification is needed to create an extra retriever for another source. However, A checking mechanism is added to the original source retriever as to make it a schedule task in the system. The checking mechanism would read the Json file where the ISM data stored and check whether the release date of the latest report is in the file. If the date is already recorded, the program will be terminated and executed only until the next scheduled time. Please note that since the release date of Manufacturing PMI and Services PMI are slightly different where they are being released in the 10 a.m. ET of the first business day and third business day of the month respectively. Therefore, the program must account for this difference in the indexing. There will be cases where the program only scraped the Manufacturing report and did not scrape the Services report, usually happens between the release date of Manufacturing PMI and Services PMI.

## Data Compilation and Findings

The second part is to clean the data and apply text classification model and normalization on the raw data to extract more meanings from the numbers. For the machine learning model, I have used the FinBERT on Hugging Face, developed by Huang et. al. (2022)[4]. In their paper FinBERT: A Large Language Model for Extracting Information




[4] FinBERT: https://huggingface.co/yiyanghkust/finbert-tone

from Financial Text, they explain the methodology and accuracy of the model in details. They have labeled financial data such as corporate report, earnings call transcript to build up from the Google LLM BERT to achieve more accurate prediction than the peer models they reviewed. This model will be used to generate sentiment score for the ISM industry comments data.

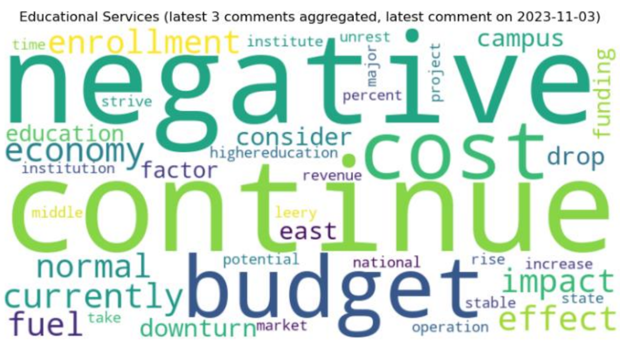
The normalization is done on the sector ranking data where the program assign a ranking number for the sector based on the order it is appeared in the text. A text of new orders in the Manufacturing PMI are shown as an example below, Where the first industry appeared indicated the most growth or decline according to the text. With the normalization, the integar type of ranking data is transformed into float type scores bounded within 1 to 0 which can be better analyzed in a time series as it allowed for apple-to-apple comparison.

## NEW ORDERS

ISM®'s New Orders Index contracted for the 15th consecutive month in November, registering 48.3 percent, an increase of 2.8 percentage points compared to October's reading of 45.5 percent. "Of the six largest manufacturing sectors, only Food, Beverage & Tobacco Products reported increased new orders. New order levels contracted at a slower rate compared to October as a result of continuing sluggishness in three capital-focused industries (Computer & Electronic Products; Machinery; and Fabricated Metal Products) that are among the seven biggest by share of manufacturing GDP. The index registered its second-highest reading since August 2022, when the index recorded 50.4 percent," says Fiore. A New Orders Index above 52.7 percent, over time, is generally consistent with an increase in the Census Bureau's series on manufacturing orders (in constant 2000 dollars).

The two manufacturing industries that reported growth in new orders in November are: Food, Beverage & Tobacco Products; and Plastics & Rubber Products. Thirteen industries reported a decline in new orders in November, in the following order: Electrical Equipment, Appliances & Components; Textile Mills; Wood Products; Furniture & Related Products; Paper Products; Computer & Electronic Products; Machinery; Nonmetallic Mineral Products; Fabricated Metal Products; Chemical Products; Primary Metals; Miscellaneous Manufacturing; and Transportation Equipment.

Beside these two data transformation, wordclouds are created during the processing of the text data. For each sector, the technique of text cleaning such as stopwords and lemmatization are being leverage to generate well formatted wordclouds. The comments from the industry from the last 3 months are extracted for the quick understand of what are the focus of the different industry in a quarter.





For the two wordclouds above, we can quickly understand there might be a continue decline in the education service sector due to the cost, and the art, entertainment & recreation business is seems to getting busy.

Student name: Chan Ching Kit
Student number: 20242725

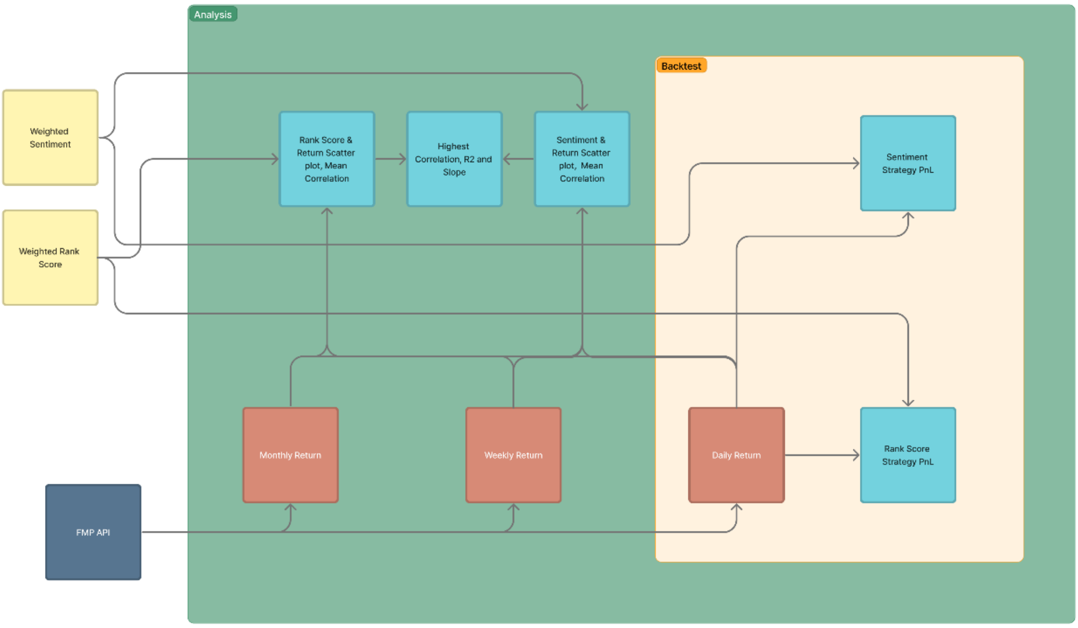
After the processing in the data, the sentiment score and rank score for each industry in the ISM report are extracted and should be ready for analysis. However, these data from the industry in the ISM report cannot be simply analyzed. Due to the limit of the market where currently there is no direct exposure for the industry since the classification of the industry in the ISM report are NAICS classification system. The closest option avaliable is sector ETF in a different classification system such as GICS from the S&P. The main concern being some of the constituents in the ETF are different from what supposed to be in another classification system. In another word, the sentiment score and rank score is not suitable for the ETF since the constituents are different. Theoreotically, these data should not provide any insight for investing this ETF because of the difference within the roots.

As the problem raised, this second part of data processing would required more work for it aids in the analysis part in the end. A workaround solution is to create weighted sentiment and rank score for the ETF which would be more complicated. For this solution, the financial modeling prep (FMP) API was used for getting data. The solution is to request the historical ETF constituents data first, then assign the sentiment and rank score based on the sector of each constituent in the ETF, and finally calculate the weighted sentiment and rank score based on the percentage holding of the constituent in the ETF. During the assigning part, since the FMP API only provide the SIC industry classification code for the constituents, the code conversion table from NAICS Association[5] is retrieved to convert SIC code into NAICS code to match the classification of ISM industry. As the conversion table from NAICS Association is lacking some of the SIC codes, an extra program was created to extend the table from the individual stock search.

[5] NAICS Association conversion table: https://www.naics.com/sic-naics-crosswalk-search-results/

Student name: Chan Ching Kit
Student number: 20242725

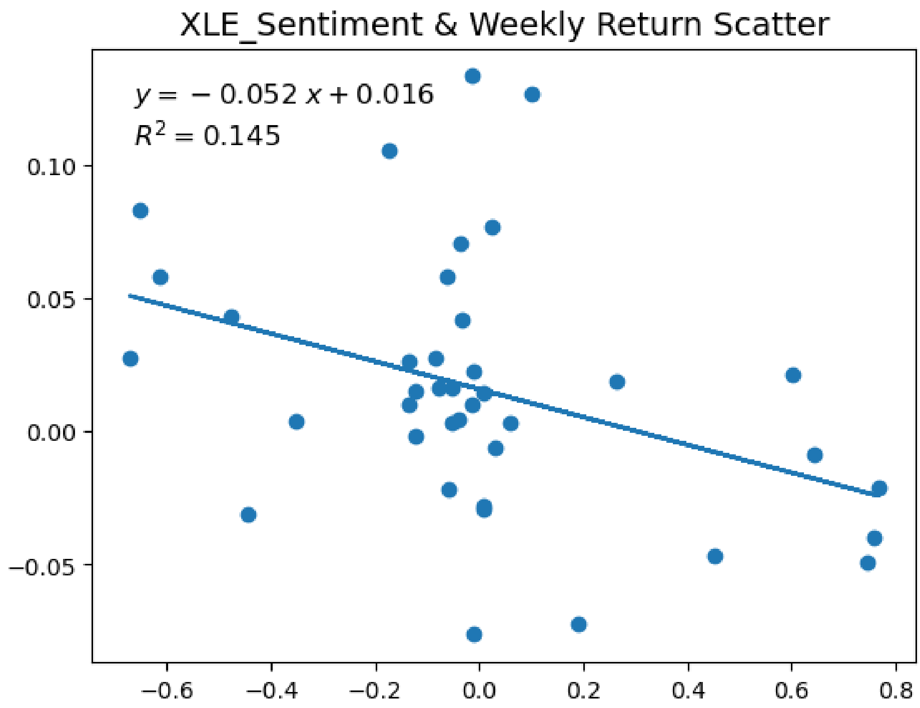
# Discussion, Analysis and Implications




For the third part of the program, the weighted sentiment and rank score would be analyzed along with the price data of the ETF. The price data was also requested from the FMP API. Then the price data was matched with the index of the weighted data since the objective is to find out the impact of the report once it is released from the website. After the indexing, price data was transformed into return in daily, weekly, and monthly frequency for measuring the correlation between the weighted sentiment and rank score. Since the original assumption is that a higher sentiment or rank score will contribute to a higher return. If the correlation found out is aligned with this assumption, we will have higher confidence in establishing a strategy with the logic of the assumption.

```
Mean Correlation for Sentiment and Monthly return is:  0.10431995150072983

Mean Correlation for Sentiment and Weekly return is:  0.09543514310597154

Mean Correlation for Sentiment and Daily return is:  0.009335728500408301

Mean Correlation for Rank score and Monthly return is:  0.020479860366667015

Mean Correlation for Rank score and Weekly return is:  -0.01630164283127933

Mean Correlation for Rank score and Daily return is:  0.1252960206508287
```

For the sentiment score, the highest correlations are between the monthly return and weekly return. And the highest correlation for rank score is with daily return. For these three pairs of data, a further study on their relationship will be done with an analysis on the scatter plot.




The figure on the right is an example of the plot from the highest correction pair mentioned above. The R square is only 0.145 even though it is already the pair with the highest R square. Moreover, the slope of the plot is -0.052 which is not aligned with the assumption. With the statistic not in favor of the assumption, still, we cannot just admit that the data has no insight provided. The result might be verifying another assumption that positive sentiment is having a negative impact for that sector.

To see if the data is providing an edge or not, back testing is also a good way as it directly stimulates the performance of investing from a decision-making process. Based on the assumptions, I created three strategies to back test on:

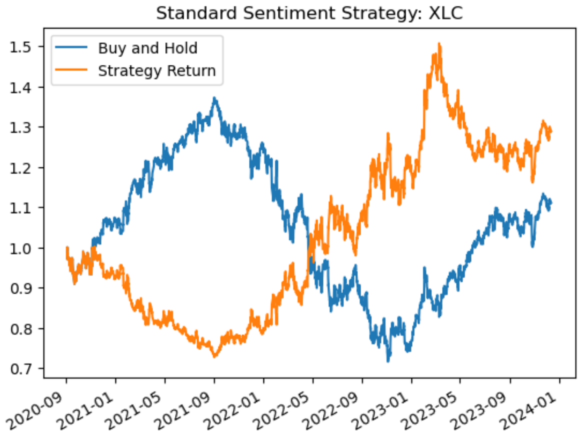
1. Standard Sentiment Strategy: Long when sentiment > 0.25 and short when sentiment < -0.25. Position limited to 1

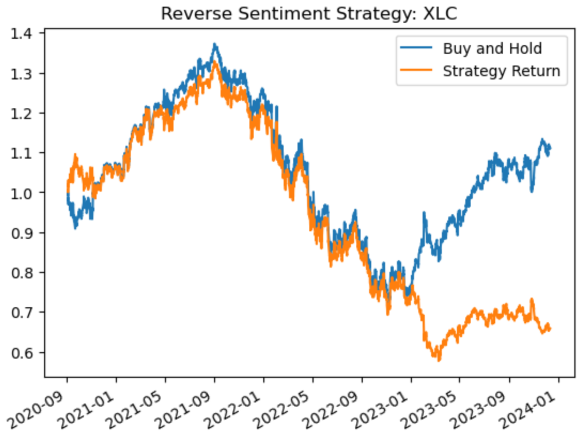
2. Reverse Sentiment Strategy: Long when sentiment < -0.25 and short when sentiment > 0.25. Position limited to 1

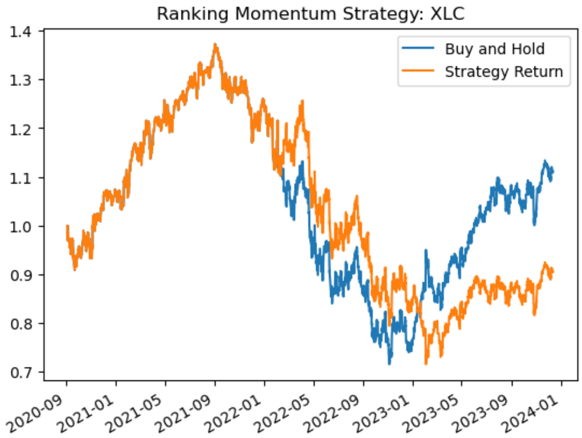
3. Ranking Momentum Strategy: Long when rank score > 0 two times in a row and short when rank sore < 0 two times in a row. Position limited to 1

Below are some of the results of the stimulation, benchmarked with the simple buy and hold.

Student name: Chan Ching Kit
Student number: 20242725




Although from this example of ticker XLC the result seems to be favoring the assumption that positive sentiment is having positive impact on the underlying, the results of other tickers shown that this might be only by random chance since the results of other tickers do not exhibit a similar and consistent characteristic.

If we count the number of tickers that outperformed the return of buy and hold for the whole period, they are:

1. 2/ 11 for Standard Sentiment Strategy
2. 1/ 11 for Reverse Sentiment Strategy
3. 5/11 for Ranking Momentum Strategy

Since the period of back testing is over three years. The number of tickers outperforming the buy and hold should at least more than half if the strategy has an edge. Even a small edge (if it is a consistent edge) would show notable result over a long term due to compound effect. To conclude, all the data did not display any sign of an edge on the market. It is likely that the weighted sentiment and weighted rank score cannot help one to make better investment decisions.

## Limitations of Research

Although the research conducted has failed to generate any meaningful insight, there are many limitations on the research and that gave the possibility that the research would be able to achieve its objective once the limitations are eliminated with the solutions provided in this report. From a high-level view, the limitations are mainly come from three categories:

1. The lack of accuracy of the data

   Most of the ISM reports analysed in this research are from a third-party source: Wayback Machine. There is even one month of data (Nov 2020 PMI) is not available on that source, where one more third-party source[6] is taken for the continuation of the study. With this kind of source, it is very hard to verify the originality of the data. It is possible for the provider to modify the data for their own benefit.

   For the transformation of raw data to weighted data, I have scraped the SIC to NAICS conversion table from NAICS Association to convert the SIC code of the constituents in the ETF. However, some of the SIC codes are not in the table as I noticed. There are still many missing SIC codes even though a specific program to extend the table has been made. This is causing some of the constituents to get NaN value for its NAICS during the data processing in my program. This problem will hugely affect the outcome of the weighted score since all the constituents with a NaN value will be removed and not contributing to the calculation of the weighted score. When the % holding is large for the constituents with a NaN value, the calculated weighted score will be very different and generate a different research result.

2. The faulty transformation of the raw data

   The FinBERT model was used for the transformation from text to sentiment score. However, it is likely that the model will make mistakes in the classification of industry comments since the model is not finetuned with industry comments data. It is trained with financial data which can be quite different from industry comments where some of the comments are related to the specific business logic in that sector. Therefore, the inaccuracy of the sentiment would give nothing but a random number that provide no

[6] https://www.prnewswire.com/news-releases/manufacturing-pmi-at-57-5-november-2020-manufacturing-ism-report-on-business-301181920.html

insight.

In the discussion, we talked about an assumption that a positive sentiment might have negative effect on some sectors. With the method of calculating weighted score, sentiments from different sectors are mixed due to the different industry classification the constituents having. So, it is impossible now to use this weighted score to find out which industry is having negative impact under a positive sentiment.

3. Survivorship bias

   The survivorship bias occurred before the calculation of the weighted score. Since the SIC code retrieved for the conversion to NAICS is the classification based on the latest business model of a company, it is possible that a company has changed their business model drastically and thus changing their SIC code. When that is the case, the calculated weighted score is once again wrong.

## Conclusion and Recommendations

The weighted sentiment and rank score calculated from the raw ISM data in this project are likely not insightful. Although the ISM's Report on Business theoretically should give some kind of investment edge with the insider information it gives, to automate the whole analysis process and generate an investment edge would require a framework that can 1) make sure the originality of the source, 2) classify accurate text sentiment, 3) seek out accurate exposure, 4) avoid survivorship bias in sector classification.

For a recommendation to tackle the caveats listed above:

1. The research can be conducted using the ISM website as the only data source. It would require a period to record a significant amount of data. This would drastically increase the quality of the data.

2. For the machine learning model, a variety of different models on Hugging Face can be tested to see the accuracy of the classification. To take this further, a proprietary model can be built using industry comments data to finetune a pretrain model. For the problem of not being able to verify the assumption that positive sentiment causing negative impact, we can build up portfolio that consisted of only one NAICS industry to isolate the impact of that sentiment score.

3. Another API with NAICS data for individual stock can be used to avoid the NaN data happening during the conversion between SIC and NAICS code. Without the unexpected NaN value, the weighted method reflects exactly the sentiment according

to the composition of the ETF and thus creates a suitable exposure that is accurate to research on.

4. The survivorship bias can be eliminated if historical data of stock’s NAICS code is found. In the worst-case scenario where none of the data vendor distribute this kind of data, this problem is still acceptable compared to the other problems since most of the listed companies will not change their business model so much that even the classification code would change. That is only happening in a few outliers. Therefore, it is a minor problem that can be overcome with a larger amount of data.

Overall, the project has established a framework to apply the data from the ISM Report on Business to generate investment insight and edge. This kind of framework can be used for more rigorous research on the same topic, or it can be extended for other data to proceed a workflow of web scraping, precise indexing for data transformation and an analysis for both the data outcome and the research itself.

# References

Huang, Allen H., Hui Wang, and Yi Yang. (2022). "FinBERT: A Large Language Model for Extracting Information from Financial Text.". *Contemporary Accounting Research*.

*Internet Archive*. (n.d.). Retrieved from Wayback Machine: https://archive.org/web/

*ISM Manufacturing and Services business survey committees*. (n.d.). Retrieved from ISM Report on Business: https://www.ismworld.org/supply-management-news-and-reports/reports/ism-report-on-business/

*NAICS Association*. (n.d.). Retrieved from SIC to NAICS conversion table: https://www.naics.com/sic-naics-crosswalk-search-results/

Postgraduate Diploma in Financial Analytics and Algo Trading (Module 5: Web Scraping and Text Analytics in Quantitative Finance)

Student name: Chan Ching Kit
Student number: 20242725

# Appendices

## Full Flow Chart