DSKUS Global Lab:
Global Data Science Exchange Program between Korea and U.S.

Supply Chain Analysis of the Automotive and Semiconductor Industries

Mitigating COVID-19 Disruptions in the United States and South Korea

Carlos Guzman
Euidam Kim
Cody Le
Janghoon Yu

DEPAUL UNIVERSITY | HANYANG UNIVERSITY

Advisors:
Ilyas Ustun, Ph.D.
Casey Bennett, Ph.D.

14 June 2022

Abstract

In this study, we investigate the impacts of the COVID-19 pandemic on the global supply chain by examining trade, production, and manufacturing capacity of the semiconductor and automotive industries in the United States and South Korea. To do so, we collect and explore the trade and manufacturing data of the semiconductor and automotive industries for both countries and COVID-19 data such as infections, testing, vaccinations, stringency, and government response to perform a cross-comparison between the two countries. The key purpose of this analysis is to select the salient features using random forest for the target variables, Industrial Production Index (IP), Industrial Capacity Index (ICAP), and Production Price Index (PPI), which could be well-predicted with these salient feature variables. With those feature and target variables, we build the forecasting model using an autoregressive Gated Recurrent Unit (GRU) recurrent neural network which predicts the future response of the target variable under the change of a certain feature variable. We aim to provide these forecasts to potential beneficiaries; our solution calls for the investment and development of SCDash (Semiconductor Diagnostic Accelerator and Supply-Chain Hub), a dashboard that automatically downloads and updates trade, production, manufacturing, and COVID-19 data. SCDash synchronizes the data into a central hub, which will include an alert and notification system that will notify the user of potential disruptions at a given time based on real-time data and provide the projections and forecast potential disruptions to end-users. Based on the result of the feature selection, we conclude that vaccinations are important for semiconductor production and manufacturing capacity in the United States and both vaccinations and testing are important in South Korea. In contrast, COVID-19 related features are not salient for predicting production of motor vehicles in either country but instead, relied more heavily on the production and capacity of semiconductors. Therefore, both the United States and South Korea should continue their advocacy for vaccinations, promote the use of COVID-19 tests, and further funding and investments in the manufacturing of semiconductors.

Contents

1 Introduction

In March 2022, after maintaining COVID-19 low-level transmission, The Washington Post published that South Korea had one of the highest infection rates in the world driven by the omicron variant. For most of the pandemic, South Korea was able to control the low transmission of the virus through widespread testing and contract tracing and avoided stringent lockdown. However, as the Korean Herald reports, the country retreated from its pandemic response strategy and relaxed anti-virus measures resulting in higher numbers of infections since the start of the pandemic. This surge came a few months after the United States experienced its omicron-driven surge and as of today, the World Health Organization (WHO) confirms that the United States has reached one million deaths and has the highest number of infected cases.

The COVID-19 pandemic has disrupted various industries around the world which has affected the supply of raw materials, intermediate goods, and finished products (Z. Xu et al.). Several industries had commodities and supply chains disrupted with shortages in high-tech products, automotive parts, medicines and medical devices, and food, and higher demand for drugs, medicines, and personal protective equipment and ventilators (Z. Xu et al.). In January 2022, the McKinsey Global Survey identified supply chain disruptions as the greatest threat to economic growth for both companies and their countries' economies, greater than the pandemic and labor shortages. The Economist Intelligence Unit surveyed 175 supply-chain managers in Asia, North America, and Europe, and found that 51.7% of managers in the automotive industry stated significant disruptions to their supply chains, the highest out of all industries, with the manufacturing industry coming in second at 42.3%. The top reasons cited for supply chain disruptions in the automotive industry include production stoppages, trade restrictions, and access to raw materials or primary inputs. The semiconductor shortage facing the automotive industry has surfaced as a key topic in public discussion. In the United States, the White House Biden-Harris Administration Supply Chain Disruptions Task Force identified that semiconductors are vital to every sector of the economy including energy, healthcare, agriculture, consumer electronics, manufacturing, defense, and transportation. In addition, the United States reliance on imported chips has resulted in their vulnerabilities into the semiconductor supply chain. In January 2022, the Biden-Harris Administration announced major plans to bring semiconductor manufacturing back to the United States with 80 billion dollars in new investments through 2025. The Census Bureau's small business pulse survey revealed that the manufacturing sector had the most amount of domestic and foreign supplier disruptions during COVID-19. Moreover, the pandemic disrupted the automobile and manufacturing industries supply chains and simultaneously revealed the importance and higher demand for semiconductors.

While the pandemic has paralyzed and disrupted various supply chains and many industries that are affected by the disruptions, the response to mitigating the spread of the virus in the United States and Korea has been quite different. According to the Center for Disease Control (CDC) and Think Global Health, the United States enacted a nationwide lockdown in March 2020, while South Korea did not. Conversely, South Korea promoted mask, testing and quarantine requirements very soon after the pandemic, while the United States did not. Therefore, the two countries had different policies to combat the pandemic with the United States having inconsistent policies due to differences state by state, whereas South Korea had a uniform response focused on public health containment. These differences in COVID-19 related policies were reflected in the way the industries in the two countries reacted to the pandemic.

Under these circumstances, what is the extent of the disruptions in the supply chain in the motor vehicle and semiconductor industry since the onset of COVID-19? What extent did changes in the manufacturing of semiconductors affect the motor vehicle industry? This project seeks to explore data on COVID-19 infections, testing, vaccinations, and lockdown style restrictions to determine their effects on the trade and manufacturing of motor vehicles and semiconductors in the United States and South Korea. Unlike previous studies, this study will explore real time empirical data, analyzed as a time-series, and gathered from government databases to perform a cross-comparison between the two countries. Since both countries are considered major manufacturers of semiconductors but implemented vastly different policies and approaches towards mitigating COVID-19, the analysis will provide key insights in determining if differences in COVID-19 related restrictions and policies had negative impacts in both countries. In addition, the analysis will also look at the two industries and how they may have affected each other prior to and during the pandemic. Previous studies have focused on model simulation both economic and agent-based modeling to model future disruptions. In this analysis, we will perform forecasting using the COVID-19 data to predict trends and potential future spikes, which represents new variant surges in the ongoing pandemic to model supply chain resilience in the two countries. The results of the analysis and forecasting will be used to create a web application to allow industries and governments on a global scale to monitor trade and manufacturing in real time, use analytics for decision making, and implement policies that bolster both industries supply and demand worldwide.

2 Literature Review

More than two years into the pandemic, disruptions continue in the global supply chain affecting various industries especially the manufacturing sectors. Several types of research have been conducted to explain the problem, model the current situation, and evaluate the impact of the global supply chain disruptions caused by COVID-19. Many articles, studies, and research publications have focused on theories related to supply chain resilience, supply chain risk management, or supply chain recovery through an economic framework. Although recent surveys and news publications have emphasized that the motor vehicle and semiconductor sectors are found to be the most negatively impacted industries by the COVID-19 pandemic, very few empirical studies have been conducted on the two industries and the effects of COVID-19 on their supply chains. Since the COVID-19 pandemic is still ongoing, little research has been published on specific aspects of COVID-19 including infections, testing, vaccinations, or policies related to the mitigation effects of COVID-19 and its effects on the global supply chain. Recently published articles have focused on how disruptions in the global supply chain have affected the economy, how lockdowns or closures affected the economy, or how lockdowns and closures affected manufacturing or foreign trade. Most publications have focused on economic model simulation, agent-based modeling, or supply chain resiliency, but not on quantitative data analysis based on observed COVID-19 data and its effect on the industry. This analysis seeks to add to the literature by cross-comparing two industries and exploring current quantitative data on different aspects of COVID-19 and its effect on both real time trade and manufacturing between the semiconductor and motor vehicle industries. In addition, this analysis applies deep learning techniques for forecasting production and manufacturing capacity in the two industries in the wake of the pandemic using techniques that have been shown to be effective at forecasting the spread of the virus.

2.0.1 COVID-19 and Global Supply Chains

First, Goel et al. focused on the impacts of the supply chain disruptions on economic growth in their publication, "Supply chain performance and economic growth: The impact of COVID-19 disruptions" (2021). Their main hypothesis was that better supply chain performance increases economic growth. The data explored in this study is from the World Bank and focused on an unbalanced panel of 136 countries observed from 2007 to 2017 and their logistics performance index. The target variable was the real per capita annual GDP growth rate across all countries and the feature variable was the index capturing the overall supply chain logistics performance. Since the dataset only has data up to 2018, they used a correlation matrix to show that improvements in the supply chain performance precede improvements in economic growth which is shown in the year 2008-2009 around the time of the recent financial crisis. They used the recent financial crisis as an example of a shock to the supply chain; they believe COVID-19 would have the same negative ripple effect on economic growth. The method for the analysis was an ordinary least squares model in which the authors ensured the model was not affected by heteroskedasticity by using robust standard errors and not affected by non-normality by using quantile regression. The results showed that improvements in overall logistics performance boost economic growth. The study concludes that the results suggest that improving the logistical performance of the supply chain will mostly be beneficial to low-growth countries and that growth impacts of logistics performance vary across nations with different growth rates. This study provides a foundation for our study since the authors confirm that a global crisis such as COVID-19 could be a possible threat to economic growth and global supply chain, but the study is limited by only focusing on economic forecasting using the financial crisis instead of real quantitative data related to COVID-19.

Second, Guan et al. assessed the short-run supply chain effects of different sets of idealized containment and lockdown strategies and industrial sectors to evaluate how the pandemic-related economic losses will be distributed along the global supply chain by using economic disaster simulation model in their article, “Global supply-chain effects of COVID-19 control measures” (2020). Their model contained four modules: the production module which characterizes the production processes, the allocation module which describes how suppliers allocate products to their clients, the demand module which characterizes how firms and households issue orders to their suppliers, and the simulation module. The authors modeled four different sets of pandemic scenarios, three representing different spread extents and containment responses in China, Europe, the United States, and Global. The authors used their models to assess the potential impact of different COVID-19 control policies on the supply chain and examine the externalities of control measures. Using these models, the authors found that relaxing lockdown restrictions gradually over a long period of time (new normal) results in substantially lower supply chain effects than lifting restrictions quickly, then potentially having another round of strict lockdown. Overall, their results reveal that economic losses will be minimized by stricter initial lockdown, provided that such strictness reduces the duration of the measures. This study revealed how the COVID-19 lockdowns and closures impacted the overall economic status of the country, and thus, informs us that the strictness of lockdown imposed by governments is a meaningful characteristic impacting the global supply chain. This article uses model simulation to focus on the impact of COVID-19 lockdown on an economical level but does not apply the effects to specific aspects of the manufacturing industries.

Third, Cai and Luo analyzed the response of the manufacturing industry to the COVID-19 pandemic from the supply chain perspective in their article, "Influence of COVID-19 on the manufacturing industry and corresponding countermeasures from supply chain perspective" (2020). In this study, they found how the lockdowns in China resulted in massive delays and shutdowns in the rest of the world and severely damaged the supply and demand for products. They also found that raw material production is at high risk for complete shutdowns and found that logistics is the weak point for the manufacturing industry. The authors highlighted the growing risk of bankruptcy for small-to-medium sized enterprises, and they analyzed their response and resilience to the COVID-19 pandemic. Overall, the authors found that their recovery lagged significantly to larger enterprises, and that small enterprises were at higher risks of shutdowns and delays due to lockdowns and labor shortages. From this article, we were able to confirm how the quarantine and lockdown policies due to COVID-19 affects the manufacturing industry and the global supply chain, and what data from COVID-19 should be considered as a feature variable in our models. However, this article focuses mainly on the overall influence of COVID-19 on the overall manufacturing industry through qualitative data and does not use quantitative data specifically focused on the semiconductor and motor vehicle industries.

Forth, Belhadi et al. compared the impact of COVID-19 on the automobile and airline supply chain by evaluating the short-term and long-term response strategies of the two sectors using a combination of qualitative and quantitative techniques in three phases in their publication, "Manufacturing and service supply chain resilience to the COVID-19 outbreak: Lessons learned from the automobile and airline industries" (2021). The automotive industry was chosen to represent the manufacturing sector and the airline industry was chosen to represent the service sector. The authors use a Supply Chain Resilience (SCR) theoretical framework with a direct and indirect approach to quantify SCR, with the indirect approach accessing resilience strategies: proactive strategies (technology-driven), and reactive strategies (real-time data-driven). Results showed that the automobile sector had a higher impact on supply chain disruptions and manufacturing shutdowns whereas the airline sector had a higher shortage of working capital, sales, and overall impact. Hierarchical cluster analysis was used to segment responding firms into three types according to their response strategies: well-prepared, partially prepared, or ill-prepared and an ANOVA test validated that the clusters were significantly different. Overall, the authors found that the automotive industry perceived that the best methods to mitigate COVID-19 risks were to develop localized supply sources and use advanced industry 4.0 technologies. In contrast, the airline industry perceived that the immediate need was to get ready for business continuity challenges, by defining their operations both at the airports and within fights. This study is one of the few studies that quantitatively compared the impact of COVID-19 on the supply chain of the automotive industry. However, this study does not use real time data on the manufacturing of motor vehicles and instead focuses on survey data based on responses and interviews from managers in the automobile industry.

Lastly, in their analysis of the impacts of COVID-19 on global value chains, Hayakawa and Mukunoki investigated the impacts of COVID-19 on global value chains by analyzing monthly trade data of finished machinery products from January to June 2019 and 2020 in their article, "Impacts of COVID-19 on global value chains" (2021). The authors used the number of COVID-19 cases and deaths as measures of the impact and examined how those impacts affected trade in three scenarios: countries that import finished machinery products, countries that export finished machinery products, and countries that export machinery parts to countries exporting finished machinery products. By doing this, they aimed to assess the impacts of COVID-19 on the demand, output, and supply chain of the produced goods. In conclusion, they found that improving inventory management and providing exceptions to lockdown policies are key inputs for mitigating the negative effects of COVID-19. This is the first study to use actual observed trade and COVID-19 data to examine how COVID-19 disrupted the trade of final goods through input and output linkage. In the perspective that this study used the real time data of the COVID-19 and focuses on a specific type of manufacturing industry to evaluate the impact of COVID-19 on the industrial field, our study can be considered as an extension of this study. Like the authors, our study will also focus on real time data of COVID-19 but will look at both trade and manufacturing outputs in two different but closely related industries: semiconductors and motor vehicles.

2.0.2 Forecasting COVID-19 with Deep Learning

Given the continued spread of the virus and its mutations and variants, there is a need to develop more accurate models to predict the spread of the virus using artificial intelligence (Ayris et al. 2022). Being able to accurately forecast the spread of the virus will provide governments with better tools to enact more effective mitigation strategies and enforce health containment efforts (L. Xu et al. 2022). Recent studies have been conducted using COVID-19 data from John Hopkins University (JHU) and shown that deep learning methods can accurately predict the spread of the virus. A long short term memory (LSTM) model is shown to be effective in forecasting new cases of COVID-19 by L. Xu et al. in their study, “Forecasting COVID-19 new cases uses deep learning methods,” which compares the effectiveness of CNN, LSTM, and CNN-LSTM hybrid models for predicting new cases in Brazil, India, and

Russia. Their study incorporates government policies including face coverings, no gatherings, closure of public transportation, and stay at home measures, which are all measured through a stringency index from Out World in Data and used as a feature in their forecasting models. The authors' research showed that an CNN-LSTM hybrid model performed better than a CNN model. The authors observed that all their models performed with less accuracy for Brazil with the LSTM model yielding the best results. The authors used their LSTM model for forecasting to predict 7 days into the future by using the previous day's prediction as input data. The results show that the model had good forecasting performance in India but did not perform as well with Brazil. Moreover, their study shows that LSTM can be used to forecast new cases of COVID-19 for regions that do not have rapidly changing case numbers.

The approach by L. Xu et al. is limited to a specific country, Ayris et al. proposes two new deep learning approaches that can predict the spread of the virus for any country in their article, "Novel deep learning approach to model and predict the spread of COVID-19" (2022). The first approach is called Deep Sequential Prediction Model (DSPM) which is a stacked LSTM network consisting of four stacked LSTMs. The second approach is called Non-parametric Regression Model (NRM) which uses an additive regression time-series algorithm with a decomposed time series model to handle non-linear predictions. The two techniques were tested on global COVID-19 data from JHU for all countries in the dataset and compared to a baseline model which used Support Vector Machines (SMV) to predict the spread of the virus. Both the DSPM model and NRM model performed better with countries that had a high number of cases compared to the baseline model which failed at predicting countries with high case counts. The DSPM model performed well for Brazil, which shows that their model is more accurate than the model presented by L. Xu et al. which could not accurately predict for Brazil. However, both the DSPM and NRM models only consider the number of COVID-19 cases and do not consider other factors such as restrictions, recoveries, or health containment measures. In addition, the authors only explored model prediction and did not apply their models to forecasting to determine if their models can be used to forecast the spread of COVID-19. Like the authors, our study will use deep learning methods for forecasting but will apply the techniques to forecast production and manufacturing in the semiconductor and motor vehicle industries using COVID-19 related factors as predictive inputs.

In summary, it has been confirmed that international crises such as COVID-19 pose a serious threat to the global supply chain and logistics, and in particular, each country's policies mitigating COVID-19 have a significant impact on it. In terms of industries, the manufacturing industries and in particular the motor vehicle industry have been significantly affected by COVID-19. This was not only due to the direct effect of the COVID-19 but also the indirect effect of the bottleneck caused by other industries, such as the semiconductor industry. Since now, there have been limited studies regarding the impact of the COVID-19 to manufacturing industries, but few have studied the motor vehicle industry and hardly any have focused on the semiconductor industry. In addition, the current literature has not studied cross-comparison of trade and manufacturing data during the COVID-19 between two industries and between different countries. Furthermore, most recent studies have focused on model simulation or qualitative data instead of examining real time quantitative data. Therefore, our project aims to derive policies related to the global supply chain to deal with the global crisis such as the COVID-19, by analyzing how two industries may affect each during COVID-19 in two different countries using observed COVID-19 and industry data through a quantitative analysis. The analysis will apply deep learning techniques to forecast production and manufacturing capacity in the two industries as one solution to prevent future supply chain disruptions.

3 Data Prepossessing

3.1 Data Source

The datasets were curated using various government sources focusing on trade and manufacturing data for semiconductors and motor vehicles in the United States and South Korea (see Appendix A.1). The semiconductor and motor vehicle sectors were selected for this analysis since data existed for both countries for these sectors and both countries play key roles in the production of goods in these two sectors on a global scale. There were initial challenges in the data collection process since public data directly related to supply chain logistics or transport logistics were not readily available for both countries without licenses and/or fees. For the

United States, government agencies hosted separate databases which were independent from each other. In contrast, in South Korea, the government has a central hub for all statistical data related to all government agencies. This hub is hosted on a website that displays information in both English and Korea, but the website was challenging to navigate due to translation issues, different labels for similar information, different subgroups, and sub-labels, and in general, both countries using different industry classifications when classifying industrial sectors or products. The manufacturing data for South Korea is an index that looks at the relative comparison with the base year as 2015, whereas for the United States the data looks at the relative comparison with the base year of 2017. With this main difference in the metrics for the manufacturing data, two datasets were curated, one for each country, which will contain the same features for the trade and manufacturing data but measured with slightly different metrics.

3.2 Data Description

3.2.1 United States Data

For the United States, foreign trade data including total import and exports for semiconductors and motor vehicles were obtained through the United States Census Bureau, USA Trade Online, an international trade database updated monthly. Manufacturing data including industrial capacity and capacity utilization were obtained through the Federal Reserve System, Industrial Production and Capacity Utilization database updated monthly and seasonally adjusted. The production price indexes including Producer Price Index (PPI) were obtained through the Bureau of Labor Statistics updated monthly. Each row represents a month between January 2012 through March 2022 and the recorded trade, manufacturing, or price index per month for the respective sector. In total, 16 features were extracted with 8 features for each industry and the date as an additional feature. The date is represented in various formats for each data source and was transformed to a standardized format of YYYY-MM.

#	Name of Feature	Data Type	Description for the United States
1	Date	Ordinal	The month and year from January 2012 through March 2022 in YYYY-MM format.
2	Exports_Semicon	Numeric	Total value of goods in dollars exported in the U.S. for semiconductors and electronic parts.
3	Imports_Semicon	Numeric	Total value of goods in dollars imported as appraised by the U.S. Customs and Border Protection (excludes U.S. import duties, freight, insurance, or other import fees) for semiconductor and electronic parts.
4	Exports_MotorV	Numeric	Total value of goods in dollars exported in the U.S. for motor vehicles and motor vehicle parts.
5	Imports_MotorV	Numeric	Total value of goods in dollars imported as appraised by the U.S. Customs and Border Protection (excludes U.S. import duties, freight, insurance, or other import fees) for motor vehicles and motor vehicle parts.
6	IP_Semicon	Numeric	Industrial Production Index for Semiconductors and electronic parts. Measures the real industry output and is expressed as a percentage of real output in a base year, 2017.
7	IP_MotorV	Numeric	Industrial Production Index for Motor Vehicles. Measures the real industry output and is expressed as a percentage of real output in a base year, 2017.
8	ICAP_Semicon	Numeric	Industrial Capacity Index for Semiconductors. Captures the sustainable maximum output within the framework of a realistic work schedule, after factoring in normal downtime and assuming sufficient availability of inputs to operate the capital in place. Expressed as a percentage of a real output in a base year, 2017.
9	ICAP_MotorV	Numeric	Industrial Capacity Index for Motor Vehicles. Captures the sustainable maximum output within the framework of a realistic work schedule, after factoring in normal downtime and assuming sufficient availability of inputs to operate the capital in place. Expressed as a percentage of a real output in a base year, 2017.
10	CAPUTL-Semicon	Numeric	Capacity Utilization Rate for Semiconductors. The rate is equal to the output index (seasonally adjusted) divided by the capacity index. Represents the percentage of resources used to produce the outputs.
11	CAPUTL-MotorV	Numeric	Capacity Utilization Rate for Motor Vehicles. The rate is equal to the output index (seasonally adjusted) divided by the capacity index. Represents the percentage of resources used to produce the outputs.
12	PPI_Semicon	Numeric	Producer Price Index for semiconductors and related device manufacturing. Measures the average change over time in the selling prices received by domestic producers for their outputs.
13	PPI_MotorV	Numeric	Producer Price Index for motor vehicle manufacturing. Measures the average change over time in the selling prices received by domestic producers for their outputs.
14	EPI_Semicon	Numeric	Export Price Index for semiconductors and related device manufacturing. Percentage change in prices paid to foreign producers for their outputs within the U.S.
15	EPI_MotorV	Numeric	Export Price Index for motor vehicle manufacturing. Percentage change in prices paid to foreign producers for their outputs within the U.S.
16	IPI_Semicon	Numeric	Import Price Index for semiconductors and related device manufacturing. Percentage change in prices received by the U.S. producers for outputs sold overseas.
17	IPI_MotorV	Numeric	Import Price Index for motor vehicle manufacturing. Percentage change in prices received by the U.S. producers for outputs sold overseas.

Table 1: Description of United States Trade and Production Data Features

3.2.2 South Korea Data

For South Korea, the foreign trade and manufacturing data, as well as the production price indices were all obtained from the Korean Statistical Information Services (KOSIS) which is a one-stop database offering statistics produced by over 120 statistical agencies covering more than 400 subject matters. KOSIS acts as a central hub and tabulates data from various agencies in South Korea and displays the data through a web application. The trade data including imports and exports is tabulated from the Ministry of Oceans and Fisheries. The manufacturing data comes from the Statistics Korea Industrial Trend Division. Production price indices come from the Bank of Korea. The data contains information by month from January

2012 to March 2022. In total, 16 features were extracted with 8 features for each industry and the date as another feature.

Several pre-processing and transformations were performed on the data to match the data features between the two countries. First, the data is obtained in Korean and all written descriptors from the data source in Korean were translated into English. Industry sectors are labeled differently between the two countries but after careful research, correct data closely aligning to corresponding sectors between the two countries were extracted. In South Korea, imports and exports for motor vehicles include both vehicles and vehicle parts, whereas in the United States, the two are separated. These two values were combined so that the feature for the United States represents both vehicles and parts. In addition, imports and exports were displayed in thousands of dollars in the original data source for South Korea, and these features were updated to display exact dollar amounts to match the United States. For the manufacturing and production price indices in South Korea, the sector used is labeled as transport equipment, which represents automobiles and automobile parts whereas in the United States, the sector used is labeled as motor vehicles representing the motor vehicle manufacturing industry. Lastly, the industrial production, industrial capacity, and capacity utilization rate is displayed as a percentage for the United States but only as an index for South Korea, with both representing real output relative to a base year. All names of features extracted were renamed in both datasets to match.

#	Name of Feature	Data Type	Description for South Korea
1	Date	Ordinal	The month and year from January 2012 through March 2022 in YYYY.MM format.
2	Exports_Semicon	Numeric	Cargo transport performance for semiconductors for exports (original unit: thousands of dollars)
3	Imports_Semicon	Numeric	Cargo transport performance for semiconductors for imports (original unit: thousands of dollars)
4	Exports_MotorV	Numeric	Cargo transport performance for vehicles and their components for exports (original unit: thousands of dollars)
5	Imports_MotorV	Numeric	Cargo transport performance for vehicles and their components for imports (original unit: thousands of dollars)
6	IP_Semicon	Numeric	Index of Industrial Production for Semiconductors (2015 = 100) compared to a base year, 2015.
7	IP_MotorV	Numeric	Production Index for Transport Equipment (2015 = 100) compared to a base year, 2015.
8	ICAP_Semicon	Numeric	Manufacturing production capacity index for Semiconductors (2015 = 100). Production performance during production under a normal operating environment (facilities, manpower, operating hours, etc.) compared to base year, 2015.
9	ICAP_MotorV	Numeric	Manufacturing production capacity index for Transport Equipment (2015 = 100). Maximum production performance during production under a normal operating environment (facilities, manpower, operating hours, etc.) compared to base year, 2015. Transport equipment includes automobiles, trailers, and automobile parts.
10	CAPUTL-Semicon	Numeric	Manufacturing capacity utilization rate index for Semiconductors (2015 = 100). Calculated as the operating rate at comparison to the standard operating rate x 100. Represents production capacity to production performance ratio.
11	CAPUTL-MotorV	Numeric	Manufacturing capacity utilization rate index for transport equipment (2015 = 100). Calculated as the operating rate at comparison to the standard operating rate x 100. Represents production capacity to production performance ratio.
12	PPI_Semicon	Numeric	Statistics that measure price fluctuations of goods and services supplied to the domestic market by domestic producers and used as indicators of economic trends and GDP deflators. Semiconductor categories have individual devices and integrated circuits as sub-elements.
13	PPI_MotorV	Numeric	Statistics that measure price fluctuations of goods and services supplied to the domestic market by domestic producers and used as indicators of economic trends and GDP deflators. Transport equipment has automobiles, trailers, automobile parts, and motorcycles as sub-elements.
14	EPI_Semicon	Numeric	Export Price Index in dollars for Semiconductors (2015 = 100). Statistics that measure the price fluctuations in dollars of export goods to identify changes in export profitability, measure trade conditions through comparison of import-export indexes, and use them as an import-export deflator to calculate real GDP.
15	EPI_MotorV	Numeric	Export Price Index for Transport Equipment (2015 = 100). Statistics that measure the price fluctuations in dollars of export goods to identify changes in export profitability, measure trade conditions through comparison of import-export indexes, and use them as an import-export deflator to calculate real GDP.
16	IPI_Semicon	Numeric	Import Price Index for Semiconductors (2015 = 100). Statistics that measure the price fluctuations in dollars of import goods to identify changes in import cost burden, measure trade conditions through comparison of import-export indexes, and use them as an import-export deflator to calculate real GDP.
17	IPI_MotorV	Numeric	Import Price Index for Transport Equipment (2015 = 100). Statistics that measure the price fluctuations in dollars of import goods to identify changes in import cost burden, measure trade conditions through comparison of import-export indexes, and use them as an import-export deflator to calculate real GDP.

Table 2: Description of South Korea Trade and Production Data Features

3.2.3 COVID-19 Data

For both the United States and South Korea, data associated with the COVID-19 pandemic were obtained through the Our World in Data COVID-19 dataset (OWDC) and the Oxford COVID-19 Government Response Tracker (OxCGRT), both originated from projects from major research universities. From the OWDC dataset, data on confirmed cases, confirmed deaths, ICU hospitalizations, and vaccinations were extracted. The curator of the dataset obtained the confirmed cases and confirmed deaths through the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). OxCGRT project collects systematic information on policy measures that governments have taken to combat COVID-19. The researchers produced several policy indexes that aggregate the data into a single number from 0-100 and measures how many relevant indicators a government has acted upon and to what degree. Currently, the project uses 21 indicators representing containment and closure policies, economic policies, health system policies, and vaccine policies. The four indices produced are the government response index, containment and health index, stringency index, and economic support index. From these two datasets, COVID-19 features fall into three categories: COVID-19 infections and the spread of the virus, COVID-19 testing and vaccinations or methods to combat the spread of the virus, and government response to policies related to the pandemic.

Pre-processing and aggregation were performed on both datasets. Both datasets contain entries for all countries, only data for the United States and South Korea were filtered and extracted. For the OWDC dataset, the entries represent daily counts. For this project, the daily counts were aggregated into monthly sums and transformed from daily entries into monthly entries from January 2020 to March 2022. For the OxCGRT dataset, the entries represent daily index numbers, which for this analysis the average of the indices per month was used to represent January 2020 to March 2022. After aggregation, both datasets were first combined with each other to create the COVID-related features dataset. Additional dates in months were added to this dataset so that the date range for the dataset starts from January 2012, with entries for each month prior to January 2020 as zeroes. These entries are zero as it represents the time frame prior to the COVID-19 pandemic. Lastly, this dataset is combined with the trade and manufacturing dataset for the United States and South Korea separately, resulting in the finalized dataset for the two countries.

#	Name of Feature	Data Type	Description for Both Countries
18	total cases	Numeric	Total confirmed cases of COVID-19. Counts may include probable cases, where reported.
19	new_cases	Numeric	New confirmed cases of COVID-19. Counts may include probable cases, where reported.
20	total_deaths	Numeric	Total deaths attributed to COVID-19. Counts may include probable deaths, where reported.
21	new_deaths	Numeric	New deaths attributed to COVID-19. Counts may include probable deaths, where reported.
22	icu_patients	Numeric	Number of COVID-19 patients in intensive care units (ICU) on a given day.
23	total_tests	Numeric	Total number of tests for COVID-19.
24	new_tests	Numeric	New tests for COVID-19 (only calculated for consecutive days)
25	positive_rate	Numeric	The share of COVID-19 tests that are positive, given as a rolling 7-day average.
26	total_vaccinations	Numeric	Total number of COVID-19 vaccination doses administered.
27	people_vaccinated	Numeric	Total number of people who received at least one vaccine dose.
28	people_fully_vaccinated	Numeric	Total number of people who received all doses prescribed by the initial vaccination protocol.
29	total_boosters	Numeric	Total number of COVID-19 vaccination booster doses administered.
30	new_vaccinations		New COVID-19 vaccination doses administered (only calculated for consecutive days).
31	StringencyIndex	Numeric	This index records the strictness of 'lockdown style' policies that primarily restrict people's behavior. It is calculated using all ordinal containment and closure policy closure indicators and an indicator recording public information campaigns.
32	GovernmentResponseIndex	Numeric	This index records how the response of governments has varied over all indicators, becoming stronger or weaker over the course of the outbreak.
33	ContainmentHealthIndex	Numeric	This index combines the 'lockdown' restrictions and closures with measures such as testing policy and contract tracing, short term investment in healthcare, as well as investment in vaccines.
34	EconomicSupportIndex	Numeric	This index records measures such as income support and debt relief. It is calculated using all ordinal economic policies indicators.

Table 3: Description of COVID-19 and Government Response Data Features

There are 17 features related to COVID-19 with 5 features for infection and spread of the virus, 8 features for testing and vaccinations, and 4 features for government response indices. In total, each dataset for each country contains 33 features with 123 instances. The datasets are time-series data with each entry representing a time recording the output for a particular month for each feature. The date variable is the only original variable formatted as a date period. All industry related features are numeric types with foreign trade features representing dollar amounts for both countries, manufacturing features representing percentage rate for the United States or index amount for South Korea, and production price indices representing index amounts for both countries. COVID-19 related features are numeric types with features related to cases, deaths, patients, tests, and vacations representing counts, positive rate representing percentage rate and policy response indices representing indices for both countries.

4 Objective

This analysis seeks to understand what key factors influenced the disruptions in the automotive semiconductor supply chain and how these factors can be used to prevent future disruptions and improve supply chain resilience. There are two main objectives in this analysis divided into two main parts. The first objective is to determine if COVID-related features can predict trade, production, or price indices in the semiconductor and motor vehicle industries and use machine learning techniques to determine which features are most salient for prediction. Since the United States and South Korea enacted different public health strategies for mitigating the spread of COVID-19, our assumption is that government response indices such as stringency index which monitors lockdown restrictions will be salient for predicting the trade and production of semiconductors in the United States, but not in South Korea. For South Korea, testing and health containment features will be salient for predicting semiconductor trade and production. In addition, we also assume that COVID-related features did not impact trade or production of motor vehicles in either country since our assumption is that the main disruption occurred in the semiconductor industry, which exasperated the supply chain issues in the motor vehicle industry.

The second objective of this analysis is to use the salient features uncovered from the first part of the analysis to forecast and obtain future projections using a recurrent neural network. Recurrent neural networks have been shown to be able to forecast the spread of COVID-19, and so to mitigate supply chain issues, creating a forecasting model that combines salient COVID-19 features that directly affect trade and production in the semiconductor or motor vehicle industries will be useful for planning, stockpiling, and inventory logistics. The proposed solution would be to use the forecasting model and apply the projections to a user-interface for government or industry use so that better business strategies can be implemented and policies can be enacted that prevents future supply chain disruptions.

5 Data Analysis

5.1 Methodologies

For this analysis, the datasets are split into two categories: the full dataset, which includes all instances, and the COVID-19 subset, which includes instances from January 2020 through March 2022 only, representing the onset of the COVID-19 pandemic. For the full dataset, there are 97 instances representing dates prior to COVID-19 and 26 instances representing dates since the onset of COVID-19, for a total of 123 instances. For the COVID-19 subset, there are a total of 26 instances. Targets were split by industry with a set of targets for semiconductors and a separate set for motor vehicles. Based on initial exploratory data analysis, IP, ICAP, CAPUTL, PPI, and IPI for semiconductors were chosen as targets and exports, imports, IP, ICAP, and PPI were chosen as targets for motor vehicles. Chosen targets showed strong correlation with COVID-related features with a correlation coefficient above 0.8 through evaluating correlation heatmaps (see Appendix A.2). For feature selection, support vector recursive and regularized regression were performed for semiconductor targets and random forest was performed for motor vehicle targets. The target and feature selection method that resulted in a root mean square error (RMSE) below 0.5, was then used in forecasting models using recurrent neural networks (RNN).

5.2 Feature Selection Methods

5.2.1 Ridge and Lasso Regression

For the semiconductor industry, features such as targets, regularized regression methods including Ridge and Lasso (Least Absolute Shrinkage and Selection Operator) were applied to perform feature selection. Ridge regression was chosen since during initial exploratory data analysis, it was determined through multicollinearity analysis, that many of the COVID-related features are highly correlated, and Ridge regression is a method that estimates coefficients in scenarios where linearly independent variables are multicollinear or highly correlated (see Appendix A.3). Ridge regression shrinks the regression coefficients so that features that have less contribution to the outcome, have coefficients closer to zero by introducing a L-2 norm penalty term to the regression model which is the sum of the squared coefficients. Although Ridge regression does not necessarily eliminate any features from the model, the method provides insights on salient features while preventing multicollinearity in the model. Lasso is another regularized regression method that introduces a L-1 norm penalty to the regression model which is the sum of the absolute coefficients. Lasso is a feature selection method since it shrinks the coefficient to zero and eliminates the features based on the penalty shrinkage. Lasso produces simpler and more interpretable models that incorporates a reduced set of predictors; however, the method has no way of distinguishing between a strong causal feature with predictive information and a high regression coefficient and a weak feature with no explanatory or predictive information that has a low regression coefficient. Moreover, with groups of features that are highly correlated, the method tends to arbitrarily select only one feature from the group. Both methods were implemented using cross-validation functions from the scikit-learn package with Ridge regression generally performing better than Lasso regression but selecting a higher number of features. Given that the dataset has COVID-19 related features that are highly correlated, Lasso regression resulted in more sparse models but less explainable models than Ridge regression.

5.2.2 Support Vector Recursive Feature Elimination

Among the numerous machine learning techniques, support vector machines (SVM) is a widely-used, well-known classification and regression model which works by mapping data to a high-dimensional feature space so that data points can be categorized, even when the data are not otherwise linearly separable. Although less widely implemented than SVM, support vector regression (SVR) has been proven to be an effective tool in real-value function estimation. As a supervised-learning approach, SVR trains using a symmetrical loss function, which equally penalizes high and low mis-estimates. One of the main advantages of SVR is that its computational complexity does not depend on the dimensionality of the input space. Additionally, it has excellent generalization capability, with high prediction accuracy. Meanwhile, Recursive Feature Elimination (RFE), is a popular feature selection algorithm. RFE is widely used because it is highly configurable and is effective at selecting features in a training dataset that are more or most relevant in predicting the target variable. In this perspective, we used the support vector machine regression based recursive feature elimination (SVM-RFE) technique suggested by Guyon et al., as one of the feature extraction methods to discover salient features. The SVM-RFE was done using the "SelectFromModel" and "SVR" function of the scikit-learn package, and the mean of the feature importance were used as the threshold of the feature selection for the semiconductor targets.

5.2.3 Random Forest

For the motor vehicle industry features as targets, we applied the wrapper-based method using the Random Forest model to perform feature selection. Among machine learning techniques, several models have the ability to find features salient for regression or classification in the process of learning/fitting to data. Examples of those several models include the L1 penalty or the Tree-based Model, especially since the Tree-based model repeats selecting the most appropriate feature to split the branch in the process of building a tree, that is, the most information gain / gini reduction, important features can be identified after learning. Since data such as semiconductor and motor vehicle production and capacity does not have many instances due to the frequency of records by month, so instead of using gradient-boosting models that carefully learn various parts of the data through weak-learners sequentially or decision tree model which is too overly simple, we chose the Random Forest of appropriate ensemble type as a model for feature selection. A wrapper-based method was attempted using the "SelectFromModel" of scikit-learn package, and the average value of importance/weight was used as the threshold for selecting features. We did not limit the maximum number of selected features, nor did we apply any hyper-parameter tuning to the Random Forest Regressor.

In order to validate the prediction performance with the features selected by the wrapper based method, Ordinary Least Squares (OLS), Ridge Regression, Lasso Regression, and Random Forest were set as models and the results were compared with each other. At this point, the Ridge Regression and Random Forest models showed better performance than OLS and Lasso Regression. The reason for this is that the feature selection through Random Forest could also have some non-important features or features that do not have a linear relationship with the target, which seems to have adversely affected the regression scores of OLS and Lasso Regression. However, since Ridge Regression excludes features that are relatively less important or have a nonlinear relationship more strictly than Lasso, Ridge performed better than OLS, Lasso, Random Forest with the circumstance that the relationship between features and targets is linear, otherwise Random Forest was shown to be superior for nonlinear relationships.

5.2.4 Comparison of Feature Selection Methods

As a result of comparing the performance of the feature selection method and models that performed well in each industry data, the wrapper based method using Random Forest scored the most general and best RMSE scores, so it was decided to use the technique as the final feature selection method. In addition, to consider both linear and nonlinear relationships between targets and features, Ridge Regression and Random Forest Regressor were used as prediction models. Moreover, Ridge Regression and Random Forest Regressor were applied to wrapper-based feature selection results for targets of semiconductor industry features. The expected results were similar to the motor vehicle targets, as the models also showed the best performance with Ridge or Random Forest more than Lasso and SVM-RFE. Interestingly, for the semiconductor industry in the United States, Ridge showed better performance than Random Forest on well-predicted targets, while Random Forest performed better in the South Korea semiconductor industry than Ridge.

5.3 Feature Selection Results and Discussion

5.3.1 Results for IP, ICAP, and PPI for Semiconductors

Semiconductor Method / RMSE	USA			KOR		
	IP	ICAP	PPI	IP	ICAP	PPI
RF with Ridge	**0.0914**	**0.0396**	**0.2524**	0.7656	0.8442	0.9432
RF with Random Forest Regressor	0.2128	0.1676	0.4582	**0.3890**	**0.0785**	0.8339
Lasso	0.1624	0.1337	0.2581	0.4193	0.1477	0.8991
SVR	0.1917	0.2292	0.2973	0.3858	0.6314	1.0924

Table 4: RMSE Values for Semiconductor Feature Selection for COVID Subset

The feature selection methods were evaluated by comparing the model's predictive performance for the selected targets with the selected features. The metric used for evaluation is RMSE, which is a frequently used measure of the differences between predicted value by a model and observed values. For the analysis, an RMSE of less than 0.5 was considered as a model that can predict the target variable accurately. Results show that the best feature selection method was Random Forest which was able to select features and create models with accurate predictability using the subset of the dataset containing dates from January 2020 through March 2022. As shown in Table 4, results for the United States, show that the models using Random Forest with Ridge Regression were able to predict IP, ICAP, and PPI for semiconductors. For South Korea, the models using Random Forest with Random Forest Regressor were able to predict IP and ICAP. From these results, it appears that production, capacity, and producer price index for semiconductors displayed a more linear relationship for the United States, compared to South Korea. PPI for semiconductors could not be predicted accurately for South Korea, which shows that in South Korea the changes in price for domestic producers were not necessarily impacted by COVID-related or motor vehicle industry features.

5.3.2 Results for IP, ICAP, and PPI for Motor Vehicles

Motor Vehicle Method / RMSE	USA			KOR		
	IP	ICAP	PPI	IP	ICAP	PPI
RF with Ridge	1.8569	**0.3718**	**0.0850**	0.7817	**0.4919**	**0.1110**
RF with Random Forest Regressor	1.1663	0.6509	0.1466	0.8964	0.7361	0.2239
RF with Lasso	1.1355	1.4151	1.4947	1.0940	1.3379	1.5264

Table 5: RMSE Values for Motor Vehicle Feature Selection for COVID Subset

In contrast, Table 5 above shows that for both the United States and South Korea, the models using Random Forest with Ridge Regression were able to predict for ICAP and PPI for motor vehicles. This shows that a linear relationship exists for capacity and producer price index for motor vehicles for both countries. However, for both countries, the models could not predict IP for motor vehicles, which shows that semiconductor and COVID-related features are not salient to the production of motor vehicles. Although, semiconductor and COVID-related features are salient to industrial capacity of motor vehicles, which is the number of resources present in a place that will enable an industry to produce goods, in this case, semiconductors are one resource and COVID-related features may affect the workforce which is another resource, but ultimately, these features are not salient to produce motor vehicles. Production of motor vehicles may have various other factors not examined by this dataset. Moreover, the results show that the models were unable to predict the production of motor vehicles for both countries but were able to accurately predict semiconductor production and capacity. For complete feature selection results, see Appendix A.4.

5.3.3 Features Selected for IP, ICAP, and PPI for Semiconductors

Features Selected by Random Forest for COVID Subset	USA Semiconductor			KOR Semiconductor		
	IP	ICAP	PPI	IP	ICAP	PPI
	8	10	6	12	9	4
Exports Motor Vehicles						
Imports Motor Vehicles						
IP Motor Vehicles						
CAPUTL Motor Vehicles						
ICAP Motor Vehicles						
PPI Motor Vehicles						
EPI Motor Vehicles						
IPI Motor Vehicles						
Total Cases						
New Cases						
Total Deaths						
New Deaths						
ICU Patients						
Total Tests						
New Tests						
Positive Rate						
Total Vaccinations						
People Vaccinated						
People Fully Vaccinated						
Total Boosters						
New Vaccinations						
Stringency Index						
Government Response Index						
Containment Health Index						
Economic Support Index						

Table 6: Features Selected by Random Forest with COVID Subset for Semiconductor Targets

From these results, IP and ICAP for semiconductors and ICAP for motor vehicles were chosen as targets to be used in forecasting models. As shown in Table 6 for IP and ICAP for semiconductors, besides total case counts and total deaths, vaccination features were salient features selected for both countries. Testing features were salient for South Korea but not the United States, which suggest that public health measures related to testing were more salient and affected the production and capacity in South Korea more than the United States. In contrast, as shown in Table 7, vaccination features were not selected for ICAP for motor vehicles in either country and instead, testing and hospitalizations were selected for both countries along with semiconductor industry features including production and capacity utilization for the United States, and imports and capacity for South Korea. This suggests that the manufacturing capacity for motor vehicles for both countries rely on semiconductor production and capacity. In addition, since testing and hospitalizations were salient features, the motor vehicle industry may be more affected by disruptions in the workforce in addition to disruptions in supply chain.

5.3.4 Features Selected for IP, ICAP, and PPI for Motor Vehicles

USA Motor Vehicle			KOR Motor Vehicle			Features Selected by Random Forest for COVID Subset
IP	ICAP	PPI	IP	ICAP	PPI	
5	10	10	8	9	14	
						Exports Semiconductors
						Imports Semiconductors
						IP Semiconductors
						CAPUTL Semiconductors
						ICAP Semiconductors
						PPI Semiconductors
						EPI Semiconductors
						IPI Semiconductors
						Total Cases
						New Cases
						Total Deaths
						New Deaths
						ICU Patients
						Total Tests
						New Tests
						Positive Rate
						Total Vaccinations
						People Vaccinated
						People Fully Vaccinated
						Total Boosters
						New Vaccinations
						Stringency Index
						Government Response Index
						Containment Health Index
						Economic Support Index

Table 7: Features Selected by Random Forest with COVID Subset for Motor Vehicles Targets

Lastly, government response indices such as stringency index, government response index, containment health index, and economic support index were not selected as salient features for predicting production and capacity of semiconductors or capacity of motor vehicles. This suggests that lockdown and closures may not have been salient factors in the semiconductor production and manufacturing capacity and government response policies may not have necessarily played a prominent role. For South Korea, the manufacturing industry did not experience a lockdown and in the United States, lockdown policies were different state by state, which means that various levels of production and manufacturing continued throughout the pandemic based on state and location. In addition, due to the limited instances used in the analysis, the models may be predicting positive trends in the data which means that features selected would contribute to positive relationship, especially in the production and capacity of semiconductors which have only been on a positive trajectory as both countries ramp up production in the industry since the onset of the pandemic.

5.4 Forecasting Methods

To prepare the data for the forecasting models, we first converted the monthly time series data into daily time series data by using 1-D interpolation to fill in the missing daily values between each month. This was done to all of our selected features and target variables in order to have more data to train a forecasting model. We used Scipy's interpolation package to have an accurate and efficient implementation and also decided to use linear interpolation to be able to fit the data as best as possible. We found cubic spline interpolation to compute values that would be impossible for certain features such as negative new COVID cases and opted for linear interpolation instead.

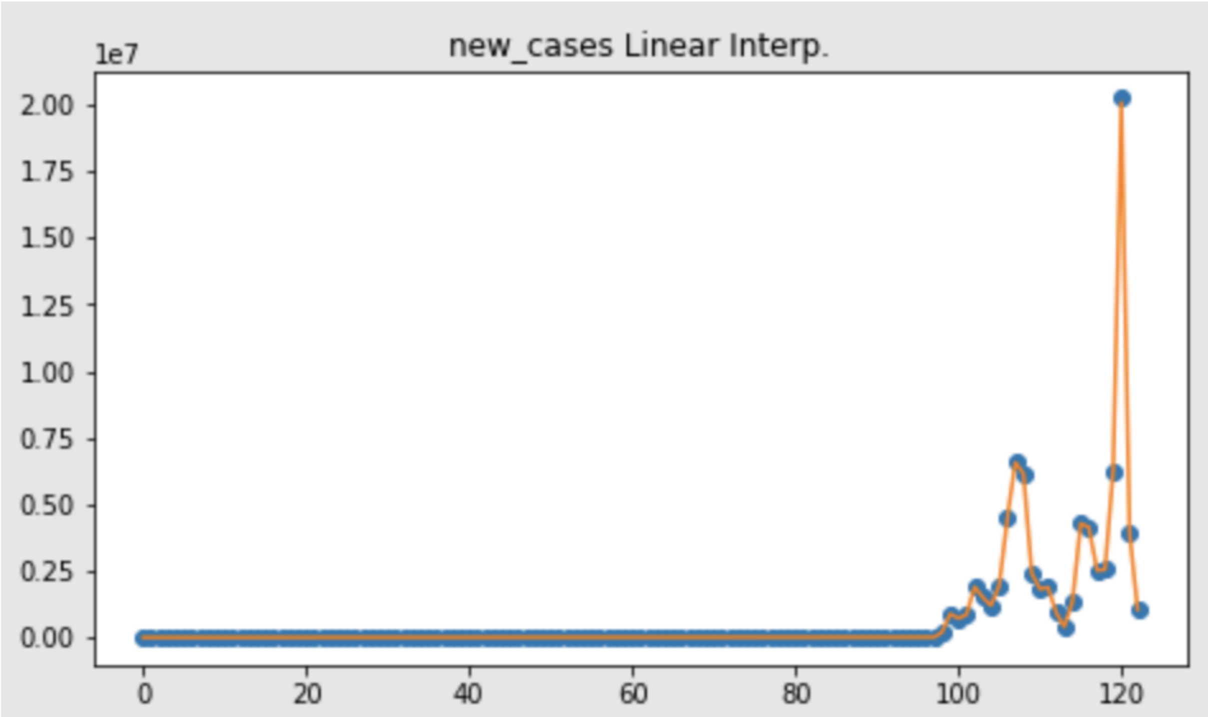

Figure 1: Example of linear interpolation of new_cases feature (USA). Notice that all values remain non-negative.

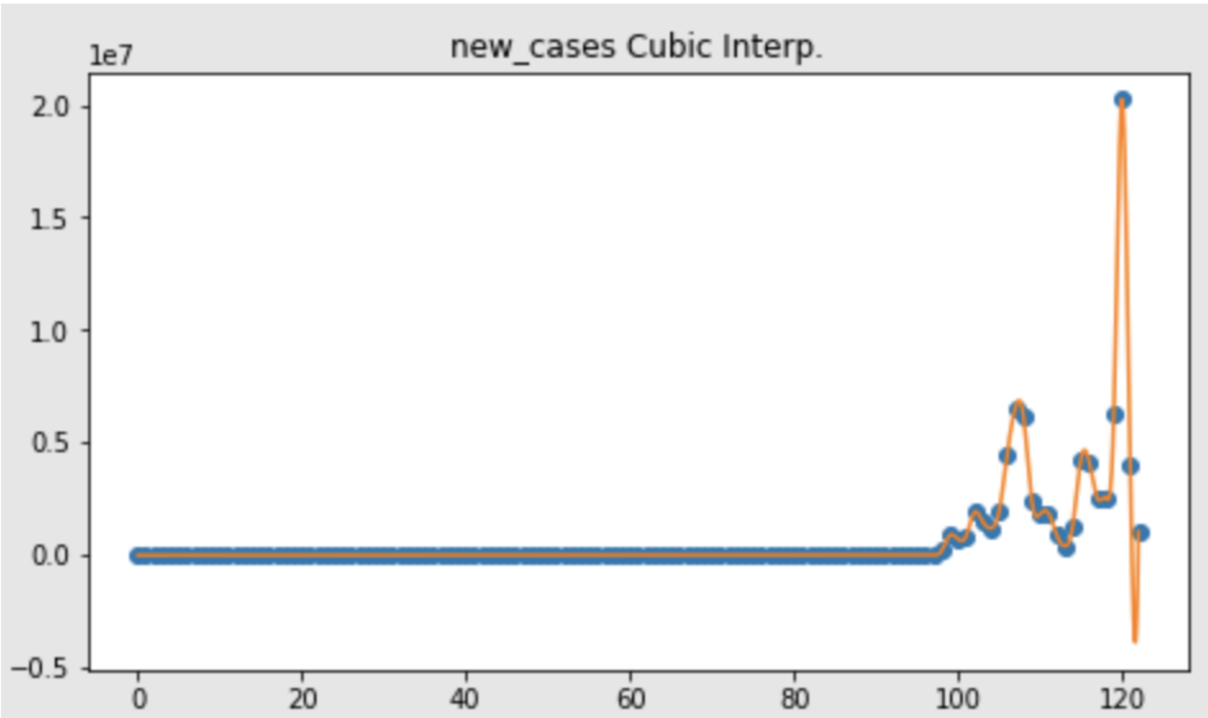

Figure 2: Example of cubic interpolation of new_cases feature (USA). Notice that some values are negative.

In order to prevent errors from using non-existent pre-2020 data for the COVID-19 related features, we reduced our dataset to use only dates from January 2020 to March 2022. As a result our dataset size went from having 123 months in our data set to only 27 months. Interpolating the monthly data into daily data then allowed us to have 820 days to be able to use for our forecasting model. Our resulting dataset was split into 90% training set and 10% testing set in order to ensure that we had enough samples to training the forecasting model. Furthermore, we standardized the dataset to have zero mean and unit variance by using the mean and standard deviation of the training set.

5.4.1 Long Short Term Memory (LSTM)

To forecast the target variables, we used the features selected from the Random Forest feature selection as the most salient for that particular target. We begin by using a Long Short-Term Memory (LSTM) neural network. An LSTM network is a recurrent neural network consisting of a LSTM cell. The LSTM cell consists of a forget gate, an input gate, an output gate, and a cell state. The forget gate decides whether to keep the information in the long run. The input gate decides which values to update and the output gate decides the value of the actual output of the neuron. The cell state acts as a highway in order to propagate the gradients of the network with respect to loss function for us to update the trainable parameters. We found that an LSTM network was the best option to use for our forecasting task because it allows us to use multiple features to forecast one target variable and it has been effectively used by others to forecast the COVID-19 pandemic.

To evaluate the training of the network we used a Mean Absolute Error (MAE) loss function. We used the MAE loss function based off of L. Xu et al., which used the same loss function for their LSTM models. We also used the Adam Stochastic Optimization algorithm as our optimizer to backpropgate the errors in order to update the network parameters; a learning rate of 1x10-3 was used for the Adam optimizer. We went with the use of an autoregressive LSTM network in order for the network to use its previous prediction for the calculation of the next prediction. This allows the model to be able to have a forecasting output of varying length as well.

Our forecasting window size consists of using the last seven days as our input to forecast the next seven days into the future. Our goal was to forecast up to 14-28 days into the future starting from April 1st, 2022 using the last seven days of the month of March 2022. Unfortunately, the LSTM model did not perform as well as expected given our lack of instances in our dataset. Furthermore, optimizing and training the network required many calculations given that we were not training for a long period of time. It was concluded that a different forecasting model such as a Gated Recurrent Unit (GRU) network would be better suited for our data instead.

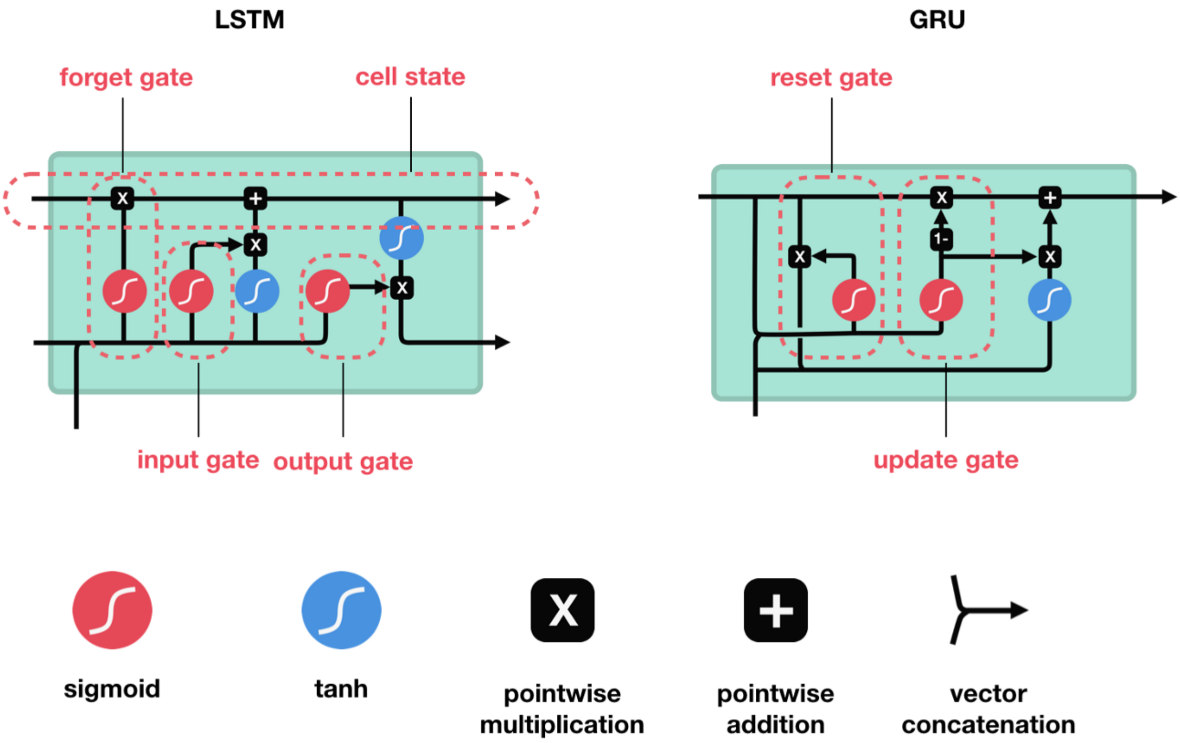

Figure 3: LSTM and GRU cell structure.

5.4.2 Gated Recurrent Units (GRU)

The GRU network is considered a simplified version of an LSTM network and has better performance for smaller amounts of data. A GRU neuron consists of a reset gate that performs similarly to the forget gate of the LSTM neuron and an update gate that decides whether we update the parameters of the neuron with a corresponding input. There is also a cell state that performs the same as the LSTM's cell state. Both the LSTM and GRU cells use sigmoid activation functions inside the gates and a tanh activation function for the output of the cell. Our GRU network also uses an autoregressive model and the MAE loss function to evaluate the training. Similarily the Adam optimization algorithm is used to train the network's parameters.

The forecasting models were built and implemented using Tensorflow. This allows us to have an efficient and accurate implementation of the networks as well as the use of their gradient tape and automatic differentiation to keep track of the gradients for the trainable parameters of each cell. In order for us to find the best mode architecture for each of the selected target variables, we ran a grid search to find the most optimal number of units for the network. The number of units refers to the total number of LSTM/GRU cells inside of a hidden cell in the network. The hidden cells refer to the number of time steps we use in order to forecast the next desired number of days. The grid search iterated through $[16, 32, 64, 128, 256, 512, 1024]$ output units. In all cases, we stopped training as soon as the MAE loss was minimized as much as possible.

5.5 Forecasting Results and Discussion

Figure 8 shows the results of our grid search to find the optimal number of units for our data. Each model was evaluated by calculating the MAE for the test set. The number of output units chosen to create Figures 4-9 correspond to the minimum MAE found in our grid search. For the United States, MAE values were all well below 0.30 but the networks required at least 256 output units to reach those levels. The forecasting models trained on the South Korean dataset required at most 128 output units but the highest MAE score it reach was 0.5547. As seen from our exploratory data analysis, the semiconductor industry in the United States had taken a far greater disruptions compared to South Korea and hence the need for more output units to be able to forecast the target. As seen in Tables 6 and 7, for most cases, more features were selected for the Korean targets than for the US target variables. This could also be a reason why the Korean target variables required less output units to forecast the variable. In all six forecasing cases, the GRU networks overfitted the training set but this is inevitable due to the nature of the model to easily overfit the training set.

Forecasting Results	USA			KOR		
	IP_Semicon	ICAP_Semicon	ICAP_MotorV	IP_Semicon	ICAP_Semicon	ICAP_MotorV
Mean Absolute Error	0.2907	0.1593	0.0805	0.3886	0.1990	0.5547
Output Units	1024	256	512	16	128	128

Table 8: GRU Network Results: Mean Absolute Error and Number of Output Units

We see in Figures 4-9 the 14 day forecast for each target variable as produced by the GRU network. As mentioned before, once the network was trained we used the last seven days of March 2022 and forecasted the next 14 days of April 2022. All target variables are normalized by dividing the the values by the maximum. The blue line corresponds to the last seven days of March 2022 and the orange line corresponds to the first 14 days of April 2022. We see that 6 is the only target variable that the network does not forecast (orange line) to decrease in value within the first couple days of April 2022. We also see that it is the only graph where the last day forecasted is greater than the last day of March 2022 (blue line). We see that the model forecasts the biggest drop in 4 within the first couple days of April 2022 from a normalized index value of 1.0 to slightly under 0.88. It then begins to level out at around 0.90 towards the last seven days.

We believe that the GRU network does not generalize the datasets well for most target variables as we see a repeating pattern for graphs 5, 7, 8, and 9. We believe that this is due to not having enough instances in our data and also that the model required a large amount of output units and thus overfitting to the training set. Overall we were able to train GRU networks that produced lower MAE scores than similar work done by L. Xu et al.

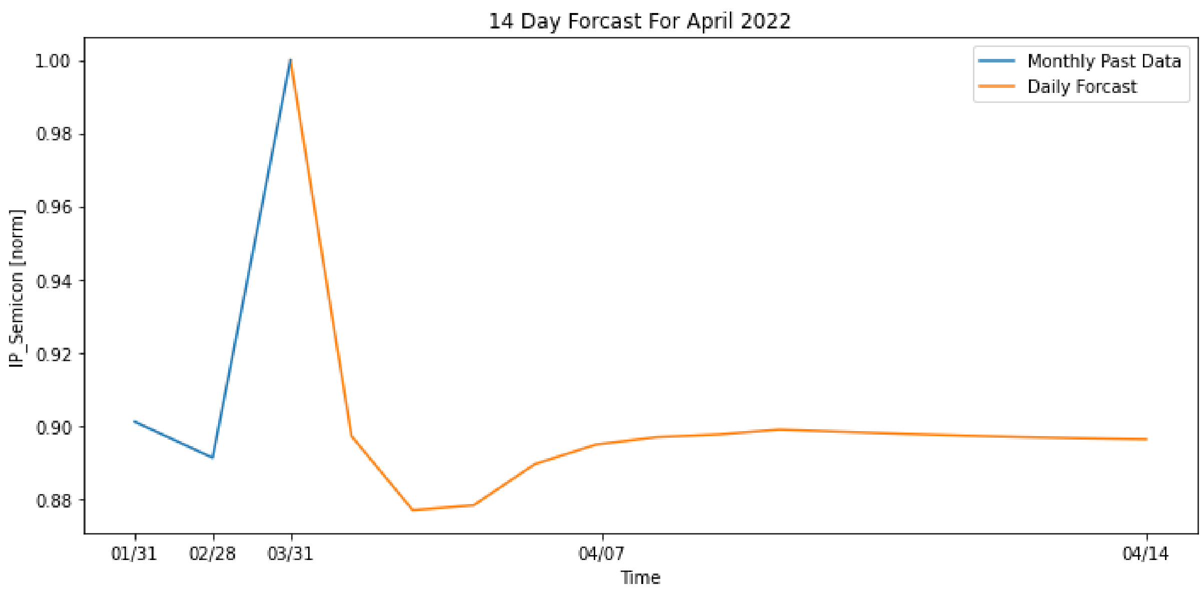

Figure 4: IP_Semicon 14 Day Forecast for April 2022 (KOR)

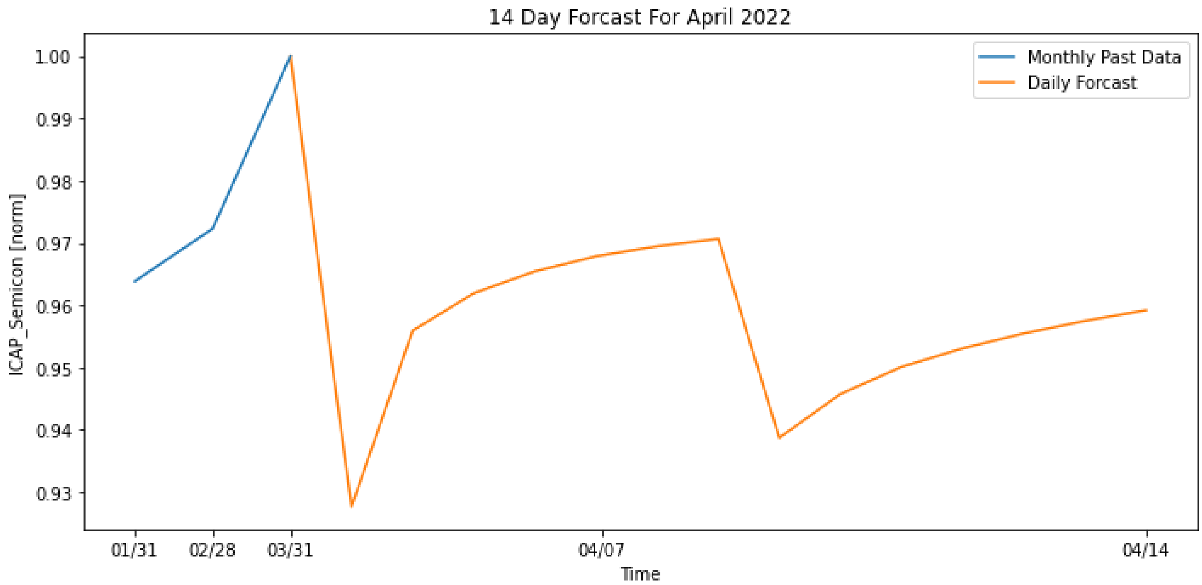

Figure 5: ICAP_Semicon 14 Day Forecast for April 2022 (KOR)

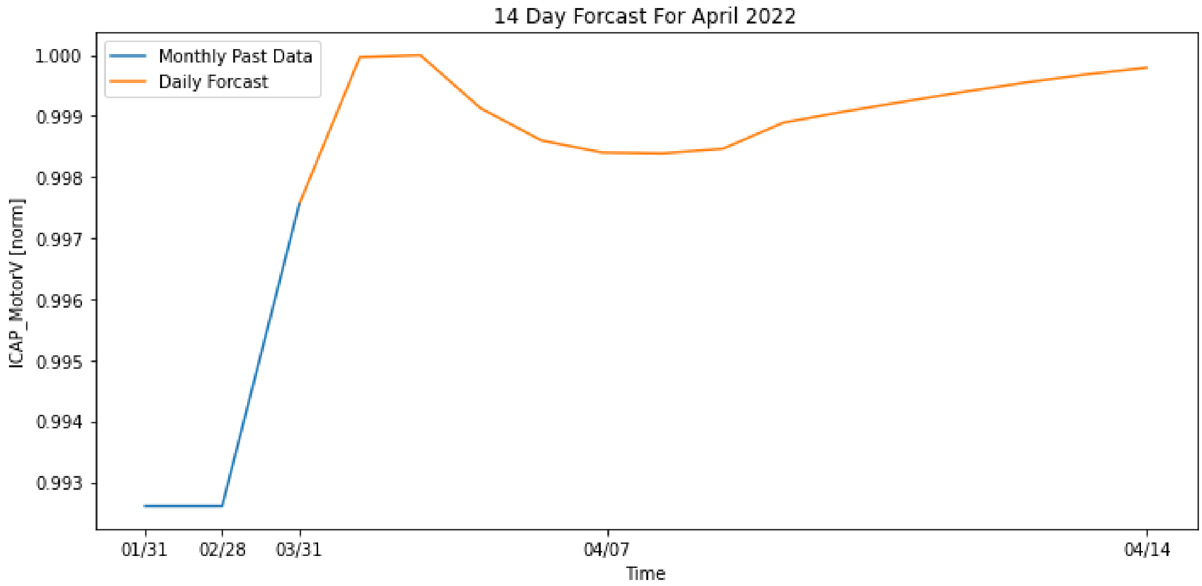

Figure 6: ICAP_MotorV 14 Day Forecast for April 2022 (KOR)

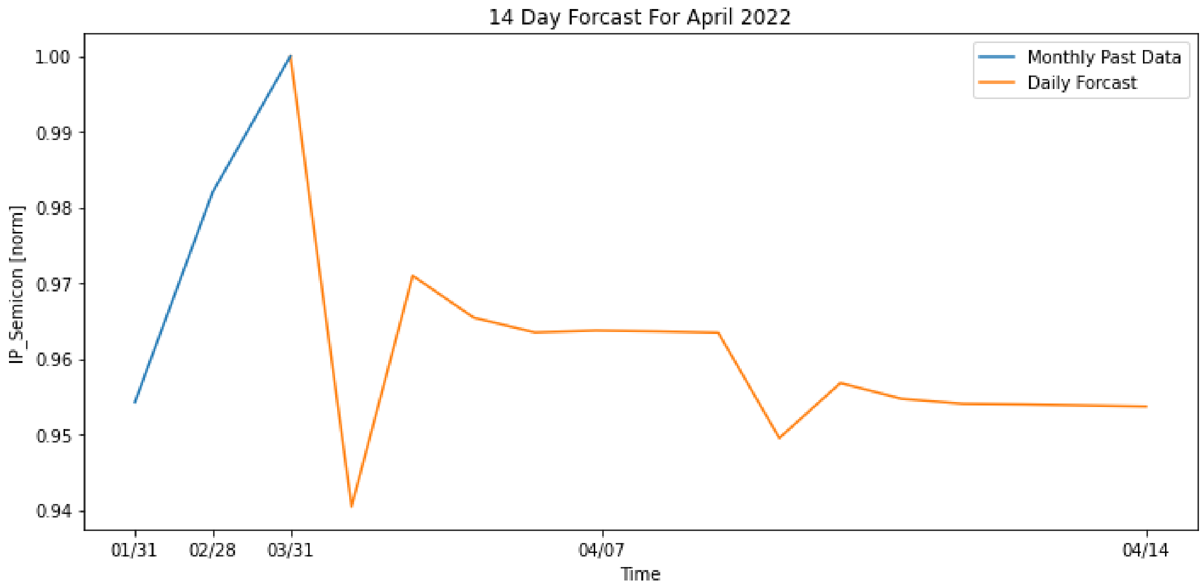

Figure 7: IP_Semicon 14 Day Forecast for April 2022 (USA)

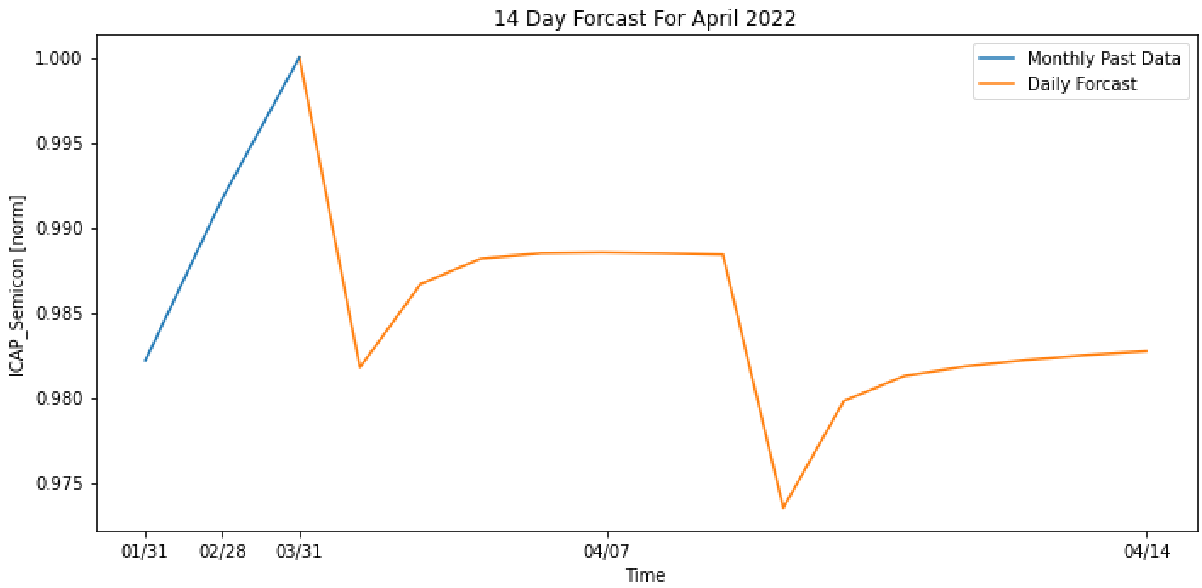

Figure 8: ICAP_Semicon 14 Day Forecast for April 2022 (USA)

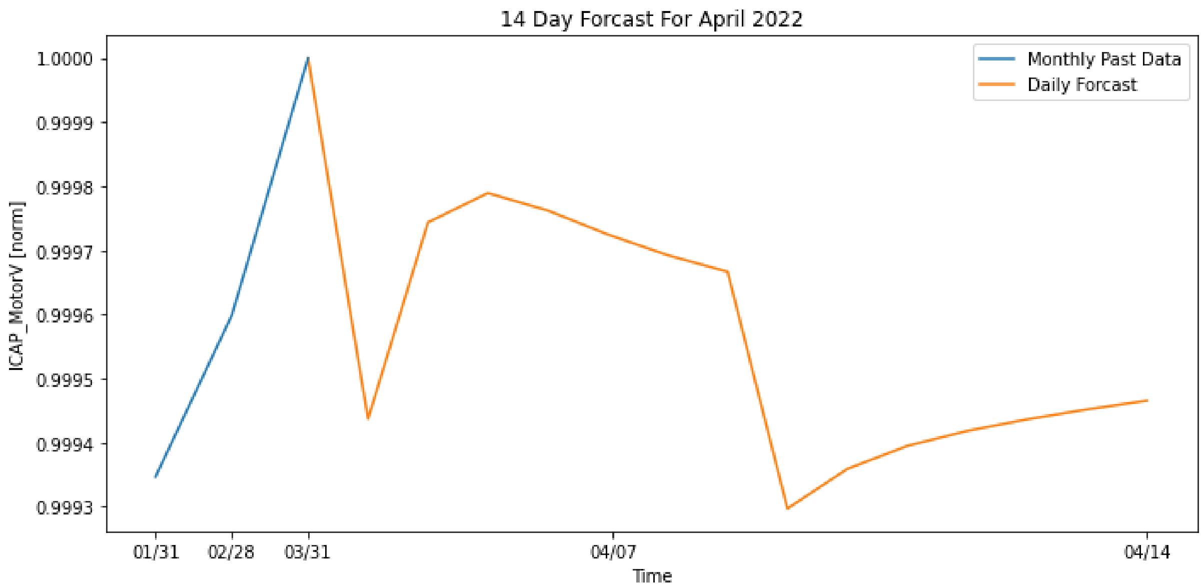

Figure 9: ICAP_MotorV 14 Day Forecast for April 2022 (USA)

6 Proposed Solution

Our analysis has shown that industry and COVID-related features can be used to predict semiconductor industrial production and capacity and motor vehicle manufacturing capacity. In addition, using the selected features in a GRU network, a forecasting model to predict future production and capacity has been shown to be effective but still requires further work. Moreover, taking the selected features and the forecasting model and combining them into a web dashboard and data-hub would allow for end users such as governments and private companies to be able to monitor production and capacity and prevent supply chain disruptions in the future. Our solution calls for the investment and development of SCDash (Semiconductor Diagnostic Accelerator and Supply-Chain Hub), a dashboard that automatically downloads and gathers trade, production, manufacturing, and COVID data and syncs the data into a central hub. Currently, trade, production, and manufacturing data are reported to different government agencies and the data is not easily accessible from a central database. Initially, SCDash would utilize existing data reporting tools from government agencies and extract the data into a hub, then combine COVID-related data, and update the information in a timely manner, to be able to provide projections and forecast potential disruptions to end users.

With the recent technology alliance between the United States and South Korea, and with South Korean semiconductor companies investing in manufacturing facilities in the United States, both governments investing and partnering with private companies to develop this dashboard would provide a great opportunity for diplomatic relations but also one solution to mitigating the disruptions in the global supply chain. Moreover, this analysis shows that motor vehicle manufacturing capacity is directly affected by semiconductor production and capacity which means that policies for investment in the production of semiconductors should continue to continue mitigating current supply chain issues. Therefore, monitoring the semiconductor production and capacity through a web dashboard would allow governments and private companies to plan, increase stockpile, make changes to inventory, and ultimately make better informed decisions. In addition to providing projections, SCDash will also include an alert and notification system that will alert the user of potential disruptions at a given time based on data. By creating a technology-based partnership for production forecasting, governments can use the dashboard to work with private companies and enact policies that would benefit production and capacity while maintaining public health and safety. In this way, supply chain disruptions can be better managed and if partnerships are successful, future iterations of the dashboard can be applied to other countries and other industries.

6.1 Prototype and User-Interface

SCDash is a web dashboard and provides a central hub to automatically synchronize trade, production, manufacturing, and COVID-19 data for each country. The dashboard will assist the government and industries with monitoring production outputs for semiconductor manufacturing during the continued COVID-19 pandemic. The main feature of the dashboard will be to capture data for industrial production, capacity, and inventory information so that users can make decisions while reviewing forecasted indicators for COVID-19 risk. The dashboard will collaborate with governments to ensure that the most up-to-date policies related to COVID-19 and public health mitigation efforts are included in the dashboard and also used as indicators to alert users. Moreover, when the risk level reaches a certain metric, the dashboard will send users a notification which will caution users of a potential disruptive time based on the data. Industry leaders can use this information in their supply chain management by revising their production plans, stockpile strategies, or alternative business partnerships to improve their operations and ensure that the disruptions are minimal to production.

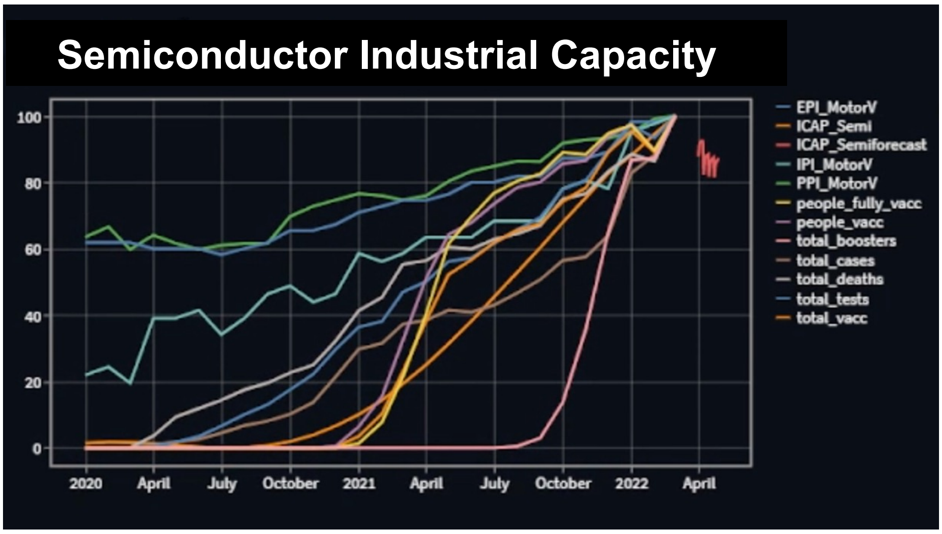

Figure 10: Semiconductor Capacity Projections with SCDash

As shown above, an initial mock-up of the SCDash will allow users to select their country and their industry and will display a graph forecasting the production and capacity for that industry. The raw data is also accessible and downloadable on the site so that users can verify the exact values of the industry and perform additional analytics with them for research and performance. In the next iteration of SCDash, with enough investment and support from both government and industry, the dashboard can be developed into a mobile application (see Appendix A.5). The mobile application will allow users to log in through a secure portal and show users their main home page which will inform them of their risk level, any potential disruptions, and level of impact, as well as production, capacity, and inventory. The goal of the application would be to allow government and private industries to form a partnership and for governments to enact policies with consideration of the supply chain and industry insights, while industries have a system to monitor and prepare for enacted policies and/or impacts from future COVID-19 variants, ultimately leading to a win-win balance between public health and safety and supply and demand.

6.2 Policy Recommendations

Based on the results of our feature selection, we recommend that both the United States and South Korea continue advocating and funding for vaccinations. Our results show that vaccine-related features are important For predicting semiconductor production and capacity and ultimately important for combating supply chain disruptions. For both countries, total vaccinations, people vaccinated, and people fully vaccinated were chosen for semiconductor production and industrial capacity. Although boosters were not chosen unanimously like vaccination features likely due to having less data for this particular feature, total boosters were selected for industrial capacity for the United States and production for South Korea, signifying that boosters are likely equally as important as vaccinations.

We also recommend policies that help promote the use of COVID tests, whether they be at home or done at a clinic, and make testing widely available and accessible. Testing was a salient feature for semiconductor production and capacity in South Korea and was also selected as an important feature for capacity for motor vehicles in both countries. Moreover, by having testing easier, more accessible, and widely available, both industries can be able to monitor their workforce and prepare to manage their staff to continue working with minimal disruptions. Moreover, industries can promote testing by enacting key policies that require testing for various situations such as when symptoms arise, a negative test to return to work after a given number of days off, and testing on a weekly or bi-weekly basis ensure overall workforce health. By being proactive and placing workforce health as a priority, the resiliency of both industries will continue to improve and thus allow for the supply chain to reach optimal levels of production and capacity.

Lastly, we also recommend that governments enact policies to support, invest, and provide additional funding to semiconductor manufacturing. Our results show that the trade, production, and capacity of semiconductors are important features for predicting motor vehicle industrial capacity, which shows that investment in the continued manufacturing of semiconductors is important to the global economy. Most industries rely on the use of semiconductors, from mobile phones, computers, laptops, and cameras, almost every modern technological device contains semiconductors. We cannot allow the industry to halt production, instead, governments should invest the necessary funds to build additional manufacturing plants, especially the United States and South Korea, who are the current global leaders in the production of semiconductors by market share.

7 Conclusion

This analysis examined the effect of COVID-19 on the global supply chain by analyzing the semiconductor and motor vehicle industries and comparing their trade and production in the United States and South Korea during the pandemic. The goal of the analysis was to determine if COVID-related features could predict trade, production, or price indices in the semiconductor and motor vehicle industries and which COVID-related features are most salient as predictors. The results show that for the United States, the production, capacity, and producer price index of semiconductors can be predicted by COVID-related features with vaccinations as the most salient predictor. Similarly, for South Korea, production and capacity for semiconductors can be predicted by COVID-related features with testing and vaccinations as the most salient predictors. Our results confirm that health containment policies such as mandatory masks and testing directly affected the production and capacity of semiconductors in South Korea. Surprisingly, government response indices such as the stringency index which includes lockdown style restrictions were not salient for predicting production or capacity for semiconductors in either country. For motor vehicles, testing was important for predicting industrial capacity, but COVID-related features were less salient for the industry, and instead, trade and production of semiconductors were more salient for predicting the industrial capacity for motor vehicles. Industrial production could not be predicted by our models, which shows that COVID-related features and semiconductor industry features may not have influenced the production of motor vehicles during the pandemic.

In addition, our analysis shows that a GRU network can be used to forecast semiconductor production, semiconductor capacity, and motor vehicle capacity with the selected COVID-related features. One solution to mitigating supply chain disruptions is to use the forecasting model and apply the projections to a user interface to monitor pandemic disruption risk, check inventory and stockpiling, and create a central hub for trade, production, and manufacturing data. Since supply chain and logistics data is not readily available, creating a technology-based partnership between governments and private industries, will allow for the monitoring and projection of production and capacity, and provide risk assessment caused by the ongoing pandemic. Furthermore, this analysis shows that policies to advocate for vaccinations, testing, and technology should continue to be able to build supply chain resilience and minimize future disruptions in both countries.

7.1 Limitations and Future Work

This analysis shows that semiconductor production and capacity can be predicted using motor vehicle trade and production features and COVID-related features. In addition, motor vehicle capacity can be predicted using semiconductor trade and production features and COVID-related features. The analysis is limited in several ways. First, the industrial production, capacity, and capacity utilization are measured as a ratio based on a base year, with the United States using 2017 as a base year, with the measurement as a percentage, and South Korea using 2015 as a base year, with the measurement as an index. This slight difference in how the production and capacity are measured may have affected the results. For this analysis, it was assumed that the difference would have little effect on the final results, nevertheless, in the future, the same metric should be used when performing cross-comparison studies between countries. Second, the data is gathered from different sources and although we attempted to match the data as closely as possible between the two countries, some differences in the reported data may affect the outcome of the results. In particular, data for motor vehicles is listed as transport equipment for South Korea and listed as a motor vehicle for the United States, although it is assumed that transport equipment only includes motor vehicles for South Korea. In the future, if the data collected is labeled the same across countries and industries, there can be more confident in the accuracy of the data. Moreover, a global classification of products and industries would be beneficial for research and analysis across industries and countries.

For this analysis, time series time steps were not incorporated into the regression models. Instead, each entry of the data is a monthly record, but time was not considered as a feature and models did not use time as a factor for feature selection. Instead, time series was only incorporated into the GRU network and applied to forecasting models. In addition, the forecasting models used deep learning networks which produce slightly different outputs with each iteration due to the nature of these networks randomizing and initializing weights. Also, deep learning networks may overfit the data, especially when there are few instances. Since our forecasting model is based on the subset of the dataset, there are fewer instances, and the model will experience overfitting even though we have added parameters in the model to handle overfitting. As such, future forecasting models can be enhanced to be more stable, robust, and tuned so that models output results more consistently. Moreover, although it has been shown that recurrent neural networks can forecast the spread of COVID-19 based on case count or infection rate, forecasting COVID-19 based on multiple factors is still novel. Therefore, the forecasting models presented in this analysis, although shown to be effective, due to the unpredictability of COVID-19, the models can only be used as an estimated projection, and differences in predicted versus actual outputs can be expected.

For future work, workforce data such as medical leave, sick days, unemployment rate, and percentage of active labor should be considered in analyzing production and manufacturing capacity. This data was not available for South Korea since no agency in South Korea monitors this data. The data does exist for the United States, and it would be interesting to see the role of human capital and its importance in the global supply chain during the pandemic. In addition, future work may want to focus specifically on policies and lockdowns without considering COVID-19 infections, testing, or vaccinations, to see to what extent government-enacted policies affect production and capacity. In this analysis, government response indicators were not as salient to the semiconductor and motor vehicle industries, but perhaps they could be salient in other industries. Moreover, future models can expand to research other industries that have been affected by the pandemic and other countries who enacted stricter lockdowns and closures who are integral in the manufacturing of key essential items, like semiconductors, that affect the global supply chain.

Works Cited

Ayris, Devante, et al. "Novel deep learning approach to model and predict the spread of COVID-19". *Intelligent Systems with Applications*, vol. 14, 2022, p. 200068.

Belhadi, Amine, et al. "Manufacturing and service supply chain resilience to the COVID-19 outbreak: Lessons learned from the automobile and airline industries". *Technological Forecasting and Social Change*, vol. 163, 2021, p. 120447.

Board of governors of the Federal Reserve System. *The Fed - Industrial Production and Capacity Utilization - G.17*, www.federalreserve.gov/releases/g17/Current/default.htm.

Bureau, U.S. Census. Census Bureau8217;s Small Business Pulse Survey Reveals Delays From Domestic, Foreign Suppliers — census.gov. [Accessed 14-Jun-2022].

Cai, Min and Jianwen Luo. "Influence of COVID-19 on manufacturing industry and corresponding countermeasures from supply chain perspective". *Journal of Shanghai Jiaotong University (Science)*, vol. 25, no. 4, 2020, pp. 409–416.

Covid-19 government response tracker. *Blavatnik School of Government*, www.bsg.ox.ac.uk/research/research-projects/covid-19-government-response-tracker.

CSSEGISandData. CSSEGISANDDATA/covid-19: Novel coronavirus (COVID-19) cases, provided by JHU CSSE. *GitHub*, github.com/CSSEGISandData/COVID-19.

Databases, tables amp; calculators by subject. *U.S. Bureau of Labor Statistics*, www.bls.gov/data/.

For Disease Control (CDC), Center. CDC Museum Covid-19 Timeline. *Centers for Disease Control and Prevention*, Jan. 2022. www.cdc.gov/museum/timeline/covid19.html.

Global, McKinsey. How COVID-19 is reshaping supply chains — mckinsey.com. [Accessed 14-Jun-2022].

Goel, Rajeev K, et al. "Supply chain performance and economic growth: The impact of COVID-19 disruptions". *Journal of Policy Modeling*, vol. 43, no. 2, 2021, pp. 298–316.

Guan, Dabo, et al. "Global supply-chain effects of COVID-19 control measures". *Nature human behaviour*, vol. 4, no. 6, 2020, pp. 577–587.

Guyon, Isabelle, et al. "Gene selection for cancer classification using support vector machines". *Machine learning*, vol. 46, no. 1, 2002, pp. 389–422.

Hayakawa, Kazunobu and Hiroshi Mukunoki. “Impacts of COVID-19 on global value chains”. *The Developing Economies*, vol. 59, no. 2, 2021, pp. 154–177.

Health, Think Global. Updated: Timeline of the coronavirus: Think global health. *Council on Foreign Relations*, www.thinkglobalhealth.org/article/updated-timeline-coronavirus.

Herald, The Korean. As it happened: South Korea’s deadly omicron experiment. [Accessed 14-Jun-2022].

House, The While. Fact sheet: Biden-Harris Administration Announces Supply Chain Disruptions Task Force to address short-term supply chain discontinuities. *The White House*, June 2021. www.whitehouse.gov/briefing-room/statements-releases/2021/06/08/fact-sheet-biden-harris-administration-announces-supply-chain-disruptions-task-force-to-address-short-term-supply-chain-discontinuities/.

Intelligence, Economist. Disruption, Digitisation, Resilience — citigroup.com. [Accessed 14-Jun-2022].

Kafchinski, Joseph E. US International Trade Data. *United States Census Bureau*, Apr. 2009. www.census.gov/foreign-trade/data/index.html.

Korean Statistical Information Services (KOSIS). *Kosis-Korean Statistical Information Service*, kosis.kr/index/index.do.

Phi, Michael. Illustrated guide to LSTM’s and GRU’s: A step by step explanation. *Medium*, June 2020. towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21.

Post, The Washington. South Korea’s coronavirus infection rate soars to one of the world’s highest with omicron-fueled surge. [Accessed 14-Jun-2022].

Ritchie, Hannah, et al. Coronavirus pandemic (COVID-19). *Our World in Data*, Mar. 2020. ourworldindata.org/coronavirus.

WHO. Who coronavirus (COVID-19) dashboard. *World Health Organization*, covid19.who.int/.

Xu, Lu, et al. “Forecasting COVID-19 new cases using deep learning methods”. *Computers in biology and medicine*, vol. 144, 2022, p. 105342.

Xu, Zhitao, et al. “Impacts of COVID-19 on global supply chains: facts and perspectives”. *IEEE Engineering Management Review*, vol. 48, no. 3, 2020, pp. 153–166.

A Appendix

A.1 Data Collection Flowchart

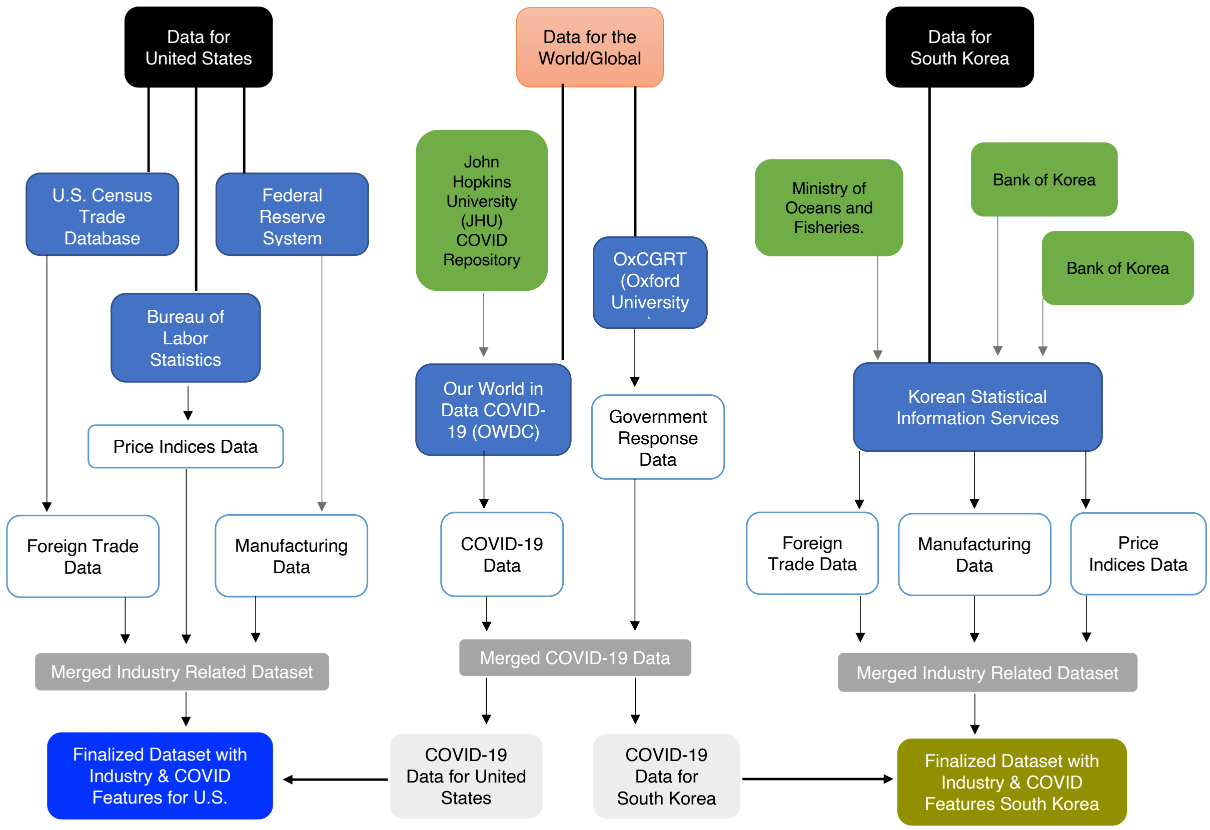

Figure 11: Flowchart of Data Sources

A.2 Correlation between Industry and COVID-19 Features

A.2.1 Correlation Heatmap for the United States

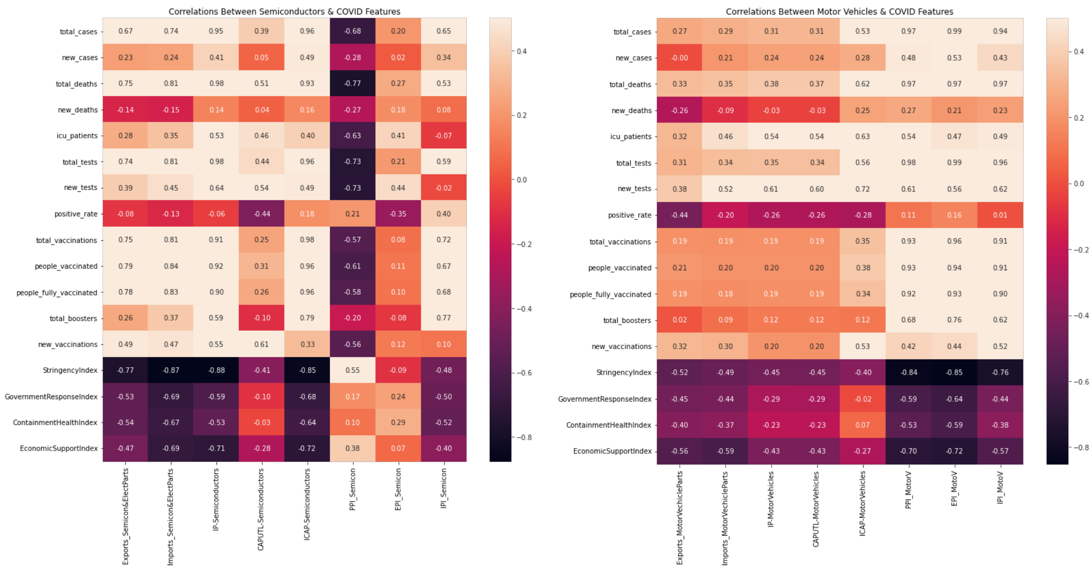

Figure 12: Semiconductor (left) and Motor Vehicles (right) with COVID-19 Features

A.2.2 Correlation Heatmap for South Korea

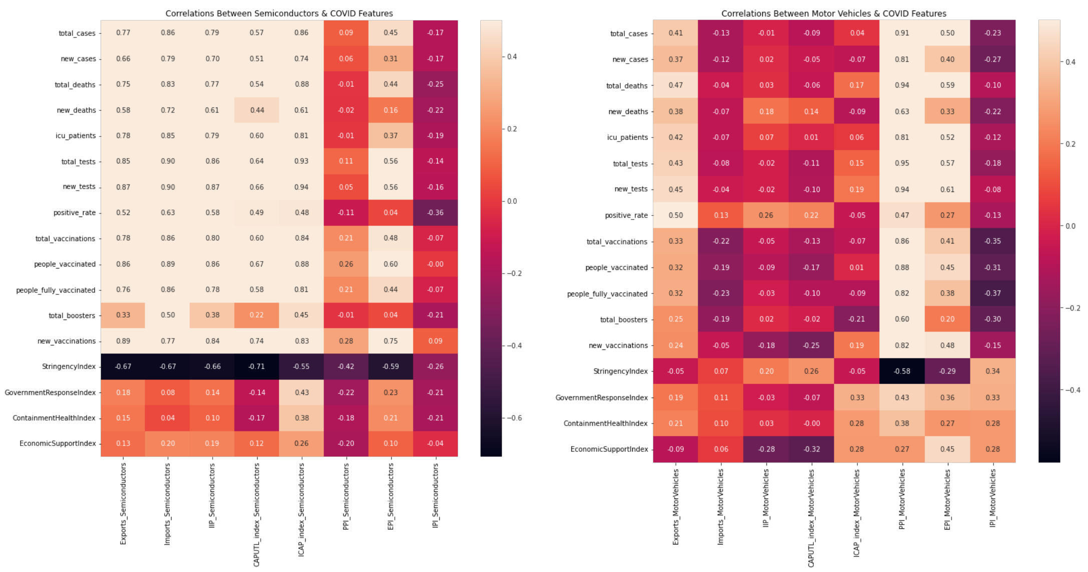

Figure 13: Semiconductor (left) and Motor Vehicles (right) with COVID-19 Features

A.3 Multicollinearity Analysis for COVID-Related Features

Multicollinearity analysis was performed to examine the COVID-related Features. Multicollinearity occurs when there are high correlations between independent variables. It was observed that multiple COVID-related features have high correlations with semiconductor and motor vehicle industry features. As such, the variance inflation factor (VIF) was examined to check for multicollinearity. Prior to checking the VIF scores, all data was normalized using standard deviation.

Results show that the VIF scores for both the United States and South Korea for the COVID-related features are very high. Multicollinearity exists between these features. Certain features such as new cases and new deaths for the U.S. have VIF scores below 5. In comparison, none of the VIF scores for South Korea are below 50. When visualizing the correlation coefficients in a heatmap for the United States, many yellows and greens appear which means that many of the features are highly positively correlated. Similarly, the heatmap for South Korea is primary yellow and gold, which shows that many of the features are highly positively correlated as well.

A.3.1 Variance Inflation Factor (VIF) for the United States

VIF for all COVID Features for United States:

	Features	VIF
0	total_cases	85.471819
1	new_cases	4.021206
2	total_deaths	68.443731
3	new_deaths	3.114127
4	icu_patients	8.974316
5	total_vaccinations	448416.953312
6	people_vaccinated	115045.042245
7	people_fully_vaccinated	68613.495641
8	total_boosters	8755.135507

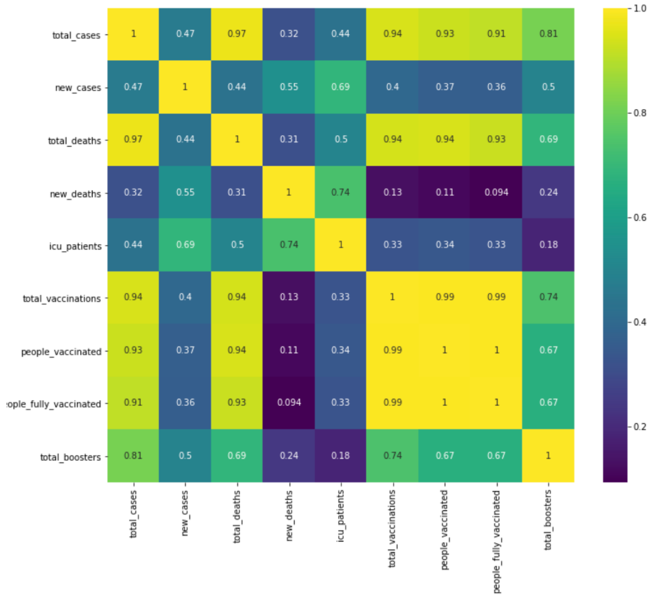

Figure 14: VIF Scores (left) and Correlation Heatmap (right) for COVID-19 Features

A.3.2 Variance Inflation Factor (VIF) for South Korea

VIF for all COVID Features for South Korea:

	Features	VIF
0	total_cases	266.889305
1	new_cases	74.705233
2	total_deaths	52.771827
3	new_deaths	460.499559
4	icu_patients	67.031595
5	total_vaccinations	258964.515772
6	people_vaccinated	43407.215930
7	people_fully_vaccinated	47483.981752
8	total_boosters	12022.504334

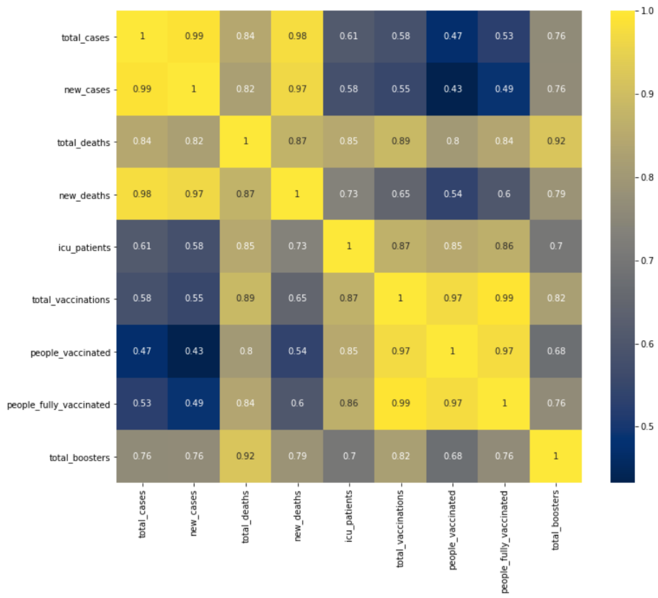

Figure 15: VIF Scores (left) and Correlation Heatmap (right) for COVID-19 Features

A.4 Additional Feature Selection Results

A.4.1 Results for CAPUTL and IPI for Semiconductors

Semiconductor Method / RMSE	USA		KOR	
	CAPUTL	IPI	CAPUTL	IPI
RF with Ridge	0.4764	0.6470	0.8717	1.1490
RF with Random Forest Regressor	0.6649	1.2743	0.7328	1.1133
Lasso	0.3113	0.1992	0.9496	0.5420
SVR	0.2376	0.2188	0.3759	0.2095

Table 9: RMSE Values for Semiconductor Feature Selection with COVID Subset

A.4.2 Results for CAPUTL and IPI for Motor Vehicles

Motor Vehicle Method / RMSE	USA		KOR	
	Exports	Imports	Exports	Imports
RF with Ridge	1.3946	2.3180	0.3384	1.6284
RF with Random Forest Regressor	1.0106	1.1437	0.5621	0.8504
RF with Lasso	1.0968	1.1510	1.2679	1.1058

Table 10: RMSE Values for Motor Vehicle Feature Selection with COVID Subset

A.4.3 Results for Semiconductors for the United States

Semiconductor Method / RMSE	USA – Full Dataset				
	IP	ICAP	CAPUTL	PPI	IPI
RF with Ridge	0.1235	0.6096	0.4592	0.1948	0.2022
RF with Random Forest Regressor	0.3099	0.4737	0.2846	0.1568	0.1621
Lasso	0.1487	0.3747	0.6298	0.1593	0.2662
SVR	0.1789	0.3361	0.5966	0.1607	0.3008

Table 11: RMSE Values for Semiconductor Feature Selection with Full Dataset

A.4.4 Results for Motor Vehicles for the United States

Motor Vehicle Method / RMSE	USA – Full Dataset				
	Exports	Imports	IP	ICAP	PPI
RF with Ridge	1.7188	1.7962	2.0498	0.2012	0.1196
RF with Random Forest Regressor	1.0998	1.1889	0.9163	0.1759	0.2711
RF with Lasso	1.1261	1.1838	1.2774	1.5065	1.4511

Table 12: RMSE Values for Motor Vehicles Feature Selection with Full Dataset

A.4.5 Results for Semiconductors for the South Korea

Semiconductor Method / RMSE	KOR – Full Dataset				
	IP	ICAP	CAPUTL	PPI	IPI
RF with Ridge	1.6716	1.2035	1.2297	1.2293	0.7121
RF with Random Forest Regressor	0.4809	0.7038	0.2325	1.3985	0.1619
Lasso	0.4239	0.3120	0.6965	0.8432	0.3213
SVR	0.4520	0.2964	0.6715	0.7920	0.3968

Table 13: RMSE Values for Semiconductor Feature Selection with Full Dataset

A.4.6 Results for Motor Vehicles for South Korea

Motor Vehicle Method / RMSE	KOR – Full Dataset				
	Exports	Imports	IP	ICAP	PPI
RF with Ridge	1.1090	0.6693	0.9701	2.3743	0.8101
RF with Random Forest Regressor	1.1727	0.5013	0.9801	0.9969	0.1900
RF with Lasso	1.1209	1.3736	1.0689	1.3742	1.5233

Table 14: RMSE Values for Motor Vehicles Feature Selection with Full Dataset

A.5 Additional User-Interface Prototype

A.5.1 SCDash Mobile Application Prototype

Figure 16: Prototype of SCDash User Homepage (left) and Projected Forecast and Risk (Right)