Anomalies Detection in Supply Chain Management

Alexander Pharaon[1] Moussa EL OUAFI[2] Yuang Chen[3] Hengyu Wang[4]

[1] The London School of Economics and Political Science
[2] École Normale Supérieure Paris Saclay
[3] University of Cambridge
[4] University of Oxford

Team 5

March 13, 2022

1 Topic Questions

A supply chain is a network established by a company and its suppliers with that oversees the production, flow, and distribution of products. We examine BigSupply Co.'s datasets, a multinational retailer that sells a wide range of products from clothing to electronics, to detect inefficiencies in its supply chain. To achieve this goal, we apply various statistical techniques to address the following questions:

1. What anomalies and inconsistencies are present in the data?
2. Is there a relationship between anomalous orders and those suspected to be fraudulent?

These are important questions to answer in order to improve company performance. The latter question in particular is crucial as, if historical accounting scandals have shown us anything, it is that corporate fraud can destroy a company.

2 Executive Summary

Although often overlooked by executive management and customers, supply chains are the arteries that enable company sales. While COVID-19 has created various temporary anomalies and bottlenecks in the supply chain, we sought to detect all anomalies since BigSupply Co. began documenting its sales. Anomalies can take different forms and arise due to various circumstances, both fraud-related and not; thus, we took four different approaches, a data science one and three machine learning models, to maximise the number of anomalies detected.

We began by exploring anomalies in the shipping times: expected, actual, and how late orders were. Orders were found to be late 0.57 days on average; however, this varied region-by-region. Intuitively, this has an impact on company performance as underestimated shipping time can increase the probability of a sale; however, there is a trade-off as misleading customers could result in fewer repeat orders. Furthermore, certain regions were found to have average delays that were statistically different from others. We recommend standardising delays across regions to a fixed number to remove the anomalies and prevent potential reputational damages that may result from being framed for discrimination.

Continuing our exploratory approach of the data, we noticed that the company lacks a coherent and systematic approach for shipping orders. Indeed, orders to countries far from the departments, such as Australia, are not done in batch and do not follow the order in which orders came in. We believe a more automated approach, similar to the one taken at Ocado's warehouses, would enable staff to find inventory quicker and thereby improve performance.

Furthermore, we detected anomalies in how the performance metrics changed across time. Firstly, there was a dramatic shift in the company's product line in late 2017 which was followed by a decrease in order quantity and a decline in performance. The change in products caused a change in the consumer demographics, shown by the proliferation of new accounts and the elimination of repeat orders by old consumers. While this could be considered anomalous, it seems to have occurred due to a structural change in the company and we recommend the company reverts its strategic direction to improve performance.

We developed three automated models for management to deploy to detect anomalies and potentially fraud. The first is a Decision Tree Classifier that assumes suspected frauds are frauds to learn important features in detecting and predicting fraud and late deliveries. This offered numerous insights into the supply chain bottlenecks that are causing late deliveries, such as the customer's city. Furthermore, it is a tool management can deploy to detect fraudulent orders and cancel them before they are shipped.

Our second model applies concepts from time series analysis to find if seasonality affects daily order profit. Furthermore, it challenges the assumptions of our first model by determining whether orders suspected to fraudulent are actually fraudulent. Finally, we provide some general results regarding outliers and anomalies in defining points where trends reverse.
Our final model combines recurrent neural networks and support vector machines to detect anomalous orders. Although the model has an impressive performance, management should not over-rely on it as it lacks explainability; however, customers may want explanations for why their orders were cancelled and deemed fraudulent. Nevertheless, the automated nature of the approach is effective in that it reduces labour costs and may improve operational efficiency.

3 Anomalies Found in the Exploratory Data Analysis

3.1 Anomalies in the Shipping Data

We began by taking the difference between the actual shipping duration and the scheduled time to find the delay customers experienced. We found this to be 0.57 days on average, which suggests the company may be misleading consumers in how long deliveries will take to increases sales. Indeed, the company could have calculated the trade-off between the number of first-time orders and repeat orders arising from misleading customers as worth it. This is supported by the fact that the average and median number of orders by a customer are 8.74 and 7.00, respectively. As such, the company knows that understating delivery times at check-out is worth it as customer retention, as measured by the number of repeat orders, is very high.

Interestingly, some regions, most notably Canada and Southern Europe, were anomalies in that they had an average delay that was statistically different from the other regions. To better understand why these regions were anomalies, we used the longitudes and latitudes to check whether delivery distance from the departments are a factor impacting shipping times.

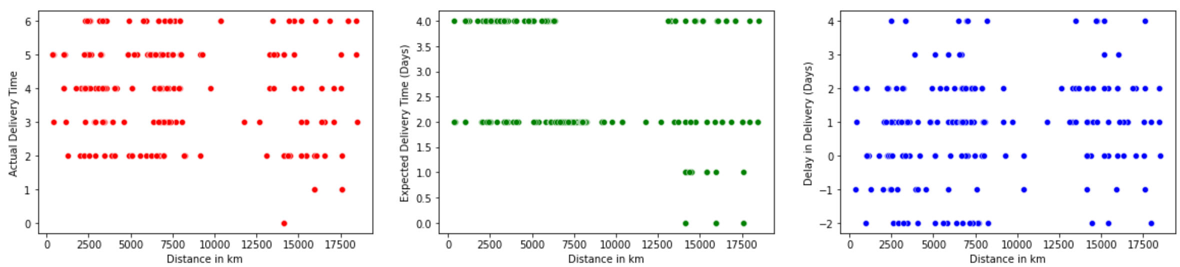

Figure 1: Relationship between the Distance to the Customer's Address and the Delivery Time

As can be inferred from Figure 1.3, the distance has no relationship with the shipping delay; therefore, the lack of explainability for certain countries strengthens our argument that they are anomalies with regards to shipping delays. More surprisingly however, Figure 1.1 shows that there is no relationship between the distance and how long shipping takes, with the distribution across delivery times approximately uniform. Furthermore, Figure 1.2 highlights another anomaly in the data: the orders from the furthest locations are sometimes predicted to be delivered on the same day whereas closer ones can take up to 4 days!

There are a few possible explanations for the findings summarized in Figure 1. Firstly, the company may have warehouses spread across the world and not just at the department locations. Secondly, it may use faster delivery methods, such as by plane, for orders that are further away. Thirdly, the anomalous orders may be fraudulent orders. This could be done by management to either reduce tax payable, supported by the strange losses reported on various orders, or to make their region appear more profitable such that they can achieve annual targets that their bonuses are linked to. Without more information on the nature of the company, it is difficult to conclude which possibility is most likely.

Given that distance clearly does not explain delays, we checked whether late delivery risk could be a contender. We obtained a correlation of 0.778 and a p-value of 0.000, which is less than any reasonable p-value threshold. Therefore, we reject the null hypothesis, which is that there is no correlation between an order being shipped late (ie. late delivery risk) and whether it arrives late or not. Although correlation does not necessarily imply causation, given our intuition of the problem, we can safely say that an order being shipped late is indeed a factor in causing late deliveries. This begs the question, why are orders being shipped late?

One possible answer to this question is that the company is waiting for further orders to come in such that they can ship everything at once. This is a common technique for taking advantage of economies of scale such that the average cost per delivery falls – this is common for companies that have their own transportation network. To see whether this was the case, we began by organising the dataset in relation to the time an order was placed at. We focused our analysis on Australia, as it is a country far away from the departments. Then, for each day, we check how many orders were not shipped on that day although a subsequent order was shipped on that day. If the company is handling its own shipping, you would expect this number to be low as orders would be sent together. We found that, in Australia on average, 35 orders are made per day but 16 of them are not shipped the same day even though a subsequent order was shipped to Australia on that day. This suggests the company pays for shipping on a per-unit basis and using an external courier, such as DHL. Furthermore, this analysis also shows that understaffed warehouses cannot be the cause of orders not being shipped on the same day, as company policy would most likely be to ship orders in a similar order to which they came in.

This restricts the possible reasons for orders being shipped late to either products being out of stock or that staff are unable to find products in the warehouse fast enough. We believe the former to be unlikely as their ecommerce site probably displays the quantity in-stock and prevents customers from buying out-of-stock items. Thus, the latter explanation is the most likely. We recommend the company invests in a better system for organizing its warehouses, possibly taking advantage of robots like Ocado does.

3.2 Anomalies in the Data Suspected to be Fraudulent

We looked for features that were indicative of fraud by comparing the distributions of such data suspected to be fraudulent and not. All distributions were statistically insignificant from each other, except for one. The distribution of customer IDs associated with orders was different for the data suspected to be fraudulent and not; however, as not all fraudulent orders may have been detected, this does not necessarily suggest that the population distributions are different. Nonetheless, this is worth exploring as it seems to have been an input relevant to the firm when deciding whether to label an order as suspicious.

To better investigate the distributional difference, we calculated the frequency of the various customer IDs – this is analogous to checking the number of orders made by a customer. On average, we found the number of orders to be 8.6 and 2.8 for the non-fraudulent and fraudulent datasets respectively, a statistically significant difference with a p-value of 0.000. This suggests that those committing fraud are constantly creating new accounts to decrease the risks of being caught.

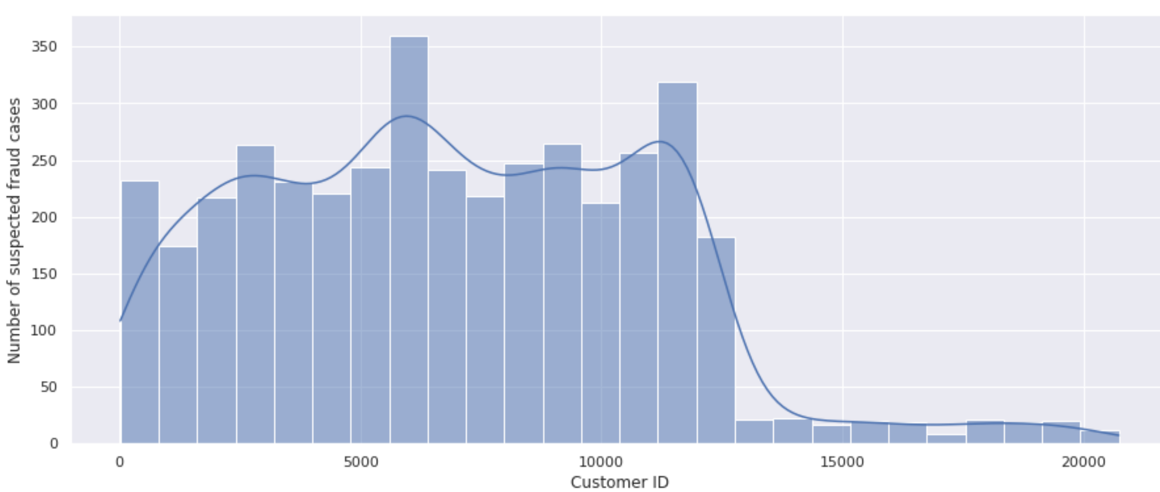

Figure 2: Distribution of Customer IDs by the Number of Suspected Fraud Cases

Another anomaly found is that older customers tended to engage in higher levels of fraud. As can be seen from the histogram in Figure 2, the levels of suspected fraud drop dramatically for the 1250$^{\text{th}}$

customer ID onwards. Since the customer ID is ascending with time, we can conclude that those committing fraud have gotten better at hiding it or that the levels of fraud dropped dramatically. Regardless, we recommend the company explore this further to ensure the continued accuracy of their fraud detection systems.

3.3 Anomalies in the Products being Sold

We found that there was a structural change in the company on the 2^{nd} of October 2017 at 12:46pm. Indeed, all products being sold previously (IDs less than 1200) stopped being sold; instead, a much smaller range of products (those with IDs greater than 1200) are solely sold instead. This can be observed in the scatterplot in Figure 3.

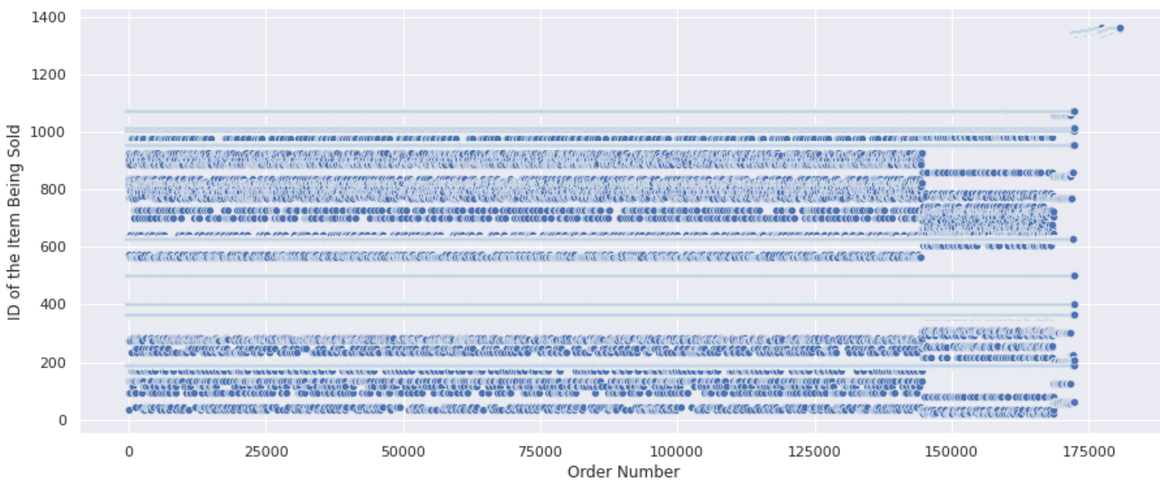

Figure 3: The Change in Products Sold Over Time

Upon investigating the new products being sold, we found that they did not all fall under the same category; for example, they started selling DVDs, toys, and first aid kits. Therefore, the reason for the company's sudden pivot is unclear and 2018 is an anomaly compared to the years prior. The evidence for classifying 2018 as an anomalous year is supported by the accompanying drop in the number of items customers bought per order after the product change, as shown in Figure 4.

Figure 4: Average Quantity Sold in an Order Across Time

3.4 Anomalies in the Regional Data

We looked at the numerical data provided by the orders dataset and saw how it differed region by region. We found some unexplainable anomalies in the data for certain regions.

First, using histograms, we looked at the frequency of numerical data, such as the distribution of discounts, across various countries. This resulted in lots of plots, so we performed various hypothesis tests to compare the distributions. To correct for the multiple comparisons, we divided our maximum p-value threshold of 0.05 by the number of pairwise comparisons. As before, the null hypothesis was that the distributions are equal. We found that lots of regions experienced distributions different from the rest of the world; however, these are not all necessarily anomalies as they can often be explained away by cultural and geographical differences. The features that were found to be statistically different in a region compared to the rest of the world are summarized in the table in Appendix A.

Secondly, we looked at how features evolved in a region over time. We found that the discount data was quite strange for most regions; indeed, more recently, orders have experienced discounts above \$300 – which is about \$200 higher than the levels reached before 2018. An extract of the plots produced, that for South Asia, is shown in Figure 5.

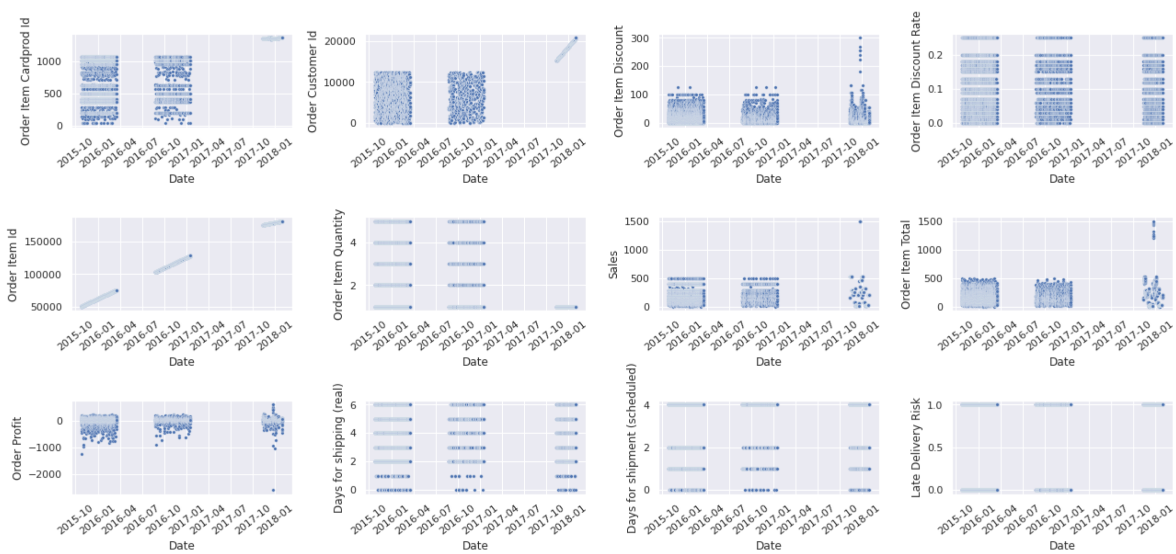

Figure 5: The Change in Feature Distributions over Time

Interestingly, we can see that there are horizontal gaps in the scatterplot due to prolonged periods of no sales in the region. This was something observed in several regions and, compared to the data of typical ecommerce sites like Amazon, is definitely an anomaly. Another anomaly can be seen by looking at the customer IDs in 2018 compared to those before. We see that repeat orders stop happening, or happen very infrequently, and that those that were purchasing pre-2018 stop purchasing from the site. This pattern was identified in the plots of every region; therefore, we can conclude that 2018 was an anomaly in relation to previous years. It seems there was a structural change in the company or that the levels of fraud may have increased.

4 Fraud Detection and Late Delivery Prediction using Machine Learning Models

4.1 Introduction

With digitization of information and rising of Internet of Things, more and more companies have to cope with huge amount of data. When processed properly, companies are able to notice hidden pattern

	Fraud Status	Late Delivery Status
Model Used	Decision Tree Classifier	Decision Tree Classifier
Accuracy	99.0832047711057%	99.19150209455415%
Confusion Matrix	$\begin{bmatrix} 35103 & 151 \\ 180 & 670 \end{bmatrix}$	$\begin{bmatrix} 16194 & 113 \\ 116 & 19681 \end{bmatrix}$
Recall	81.60779537149817%	99.42911993533394%
F1 Score	80.19150209455415%	99.4215857139249%
Cross Validation Error	0.97 (+/- 0.03)	0.98 (+/- 0.02)

Table 1: Performance of Decision Tree model on Fraud Detection and Late Delivery Prediction

of data to better improve their systems. While there are lots of options for data analysis, inspired by *Comparision-of-MLmodels-with-RNN* (Badvelu, 2020), we adopt machine learning methods and use Decision Tree Classifier as our main model with performance measure based on accuracy, recall and F1 score. Here we choose "suspected fraud" and "late delivery" as our training target.

4.2 Data Preparation

We first do data cleaning which involves eliminating missing values and getting rid of useless columns such as "product image". Since *orders* table connects information from products, categories and departments tables, and we want to see what influence all these variables have on our anomaly detection and prediction, we merge all these table together.

4.3 Data Modeling

Here we choose Decision Tree classification as our machine learning model. To measure the performance of our models. Two new columns are created in *orders* table with suspected fraud and late delivery to turn them into binary classification, which helps to measure performance of our models better. In order to measure performance more accurately all the columns with redundant values are dropped. For example, *late delivery risk* column because we can infer from products delivered late. And we convert all the non-numeric data into numeric ones so that all the data can be trained in the model. 80% of the data are for training and 20% are for testing. All the scores we use are multiplied with 100 for better understanding.

As we can see from Table 1, our model achieves a rather satisfying result. To make sure the model is acting correctly, a cross validation is implemented and the results are as follows.

As seen from the Table 1, the difference between scores of the model and scores of cross validation is negligible, our model is neither underfitted or overfitted.

4.4 Feature Importance

By using feature importance method from Scikit-learn, we have the following:

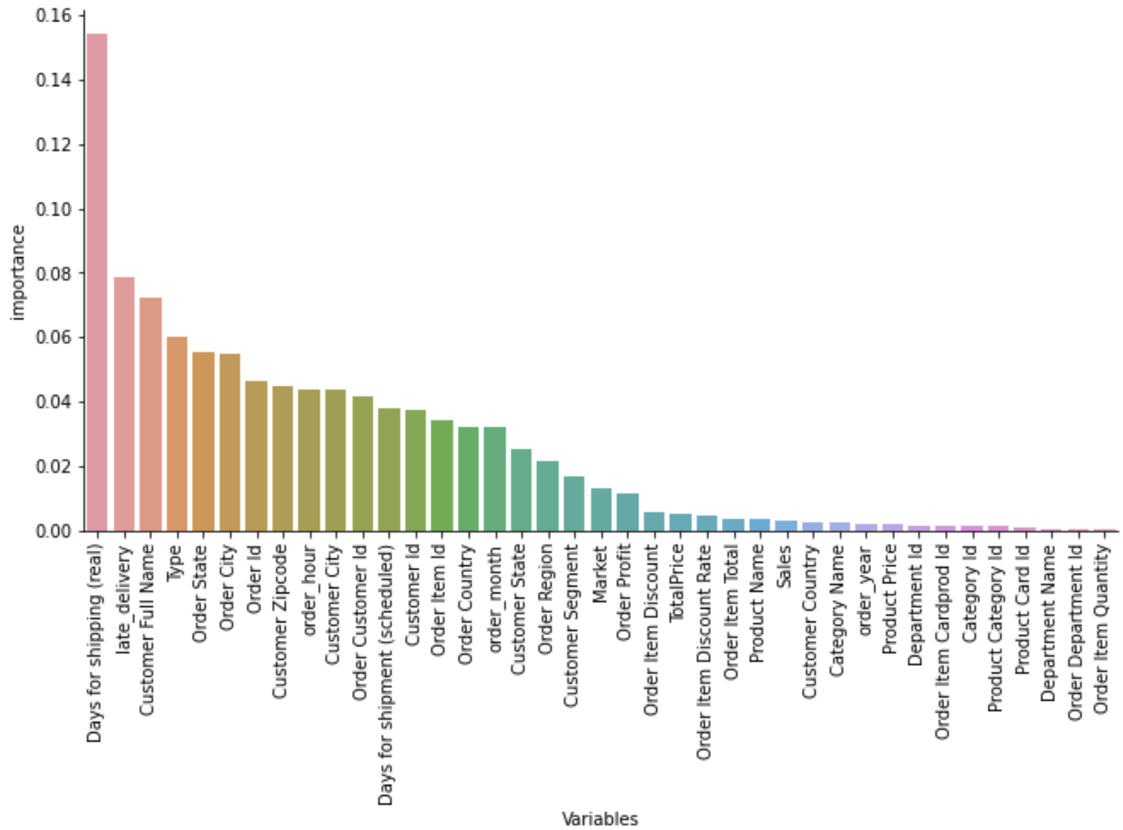

Figure 6: Feature Importance With Respect To Frauds

We can see from Figure 6 that "days for shipping (real)" takes an overwhelming importance over other variables, despite that this factor is not intuitively related to frauds. Other important factors such as "Customer Full name", "Order city" suggest company focus on customers conducting fraud.

Model Used	Decision Tree Classifier
Accuracy	86.94327498338134%
Recall	86.94327498338134%
Confusion Matrix	$\begin{bmatrix} 14023 & 2284 \\ 2430 & 17367 \end{bmatrix}$
F1 Score	86.9494820768107047%

Table 2: Performance After Removal of "Days for shipping (real)" and "Days for shipment (scheduled)"

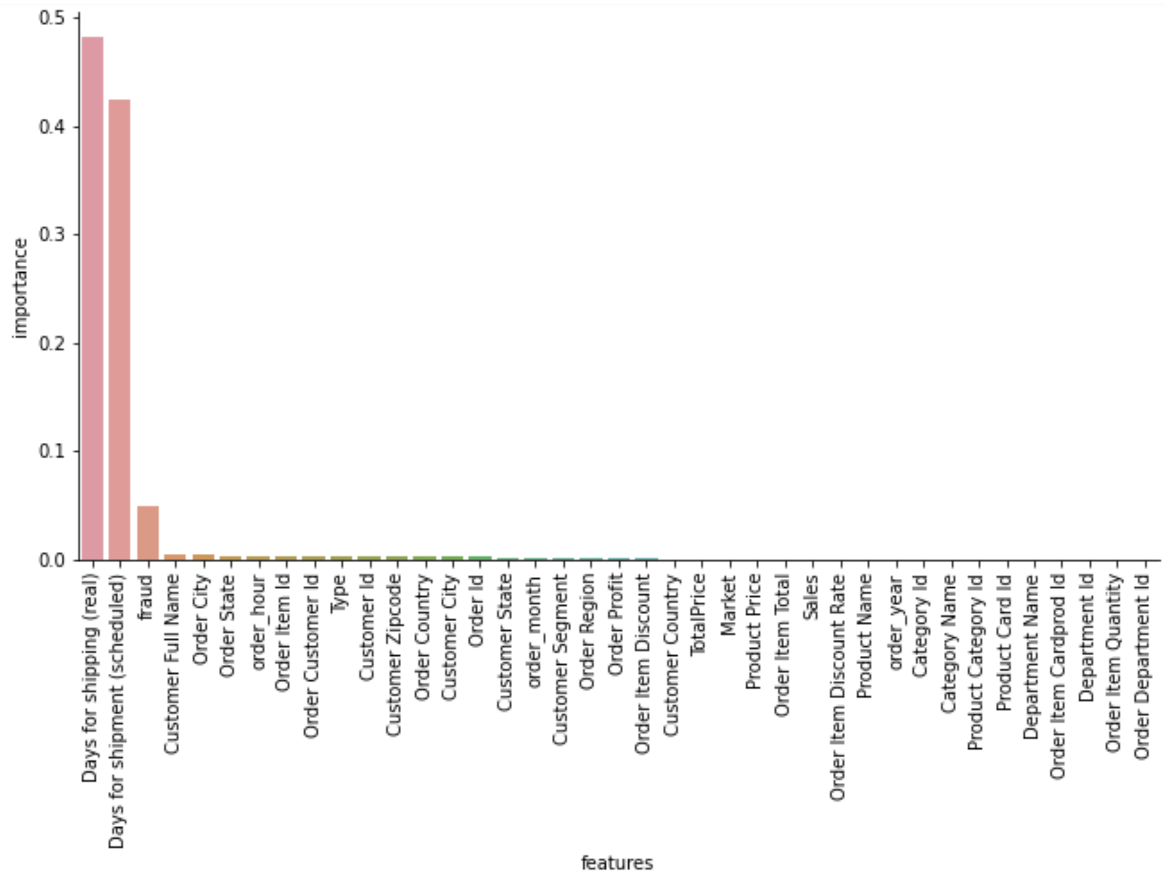

Figure 7: Feature Importance With Respect To Late Deliveries

From Figure 7, "Days for shipping (real)" and "Days for shipment (scheduled)" take up almost 90% of importance, which conforms to our intuition. It makes us wonder which factors will become important when we remove these two variables.

From Table 2, the model we trained still gives pretty good results of 87%.

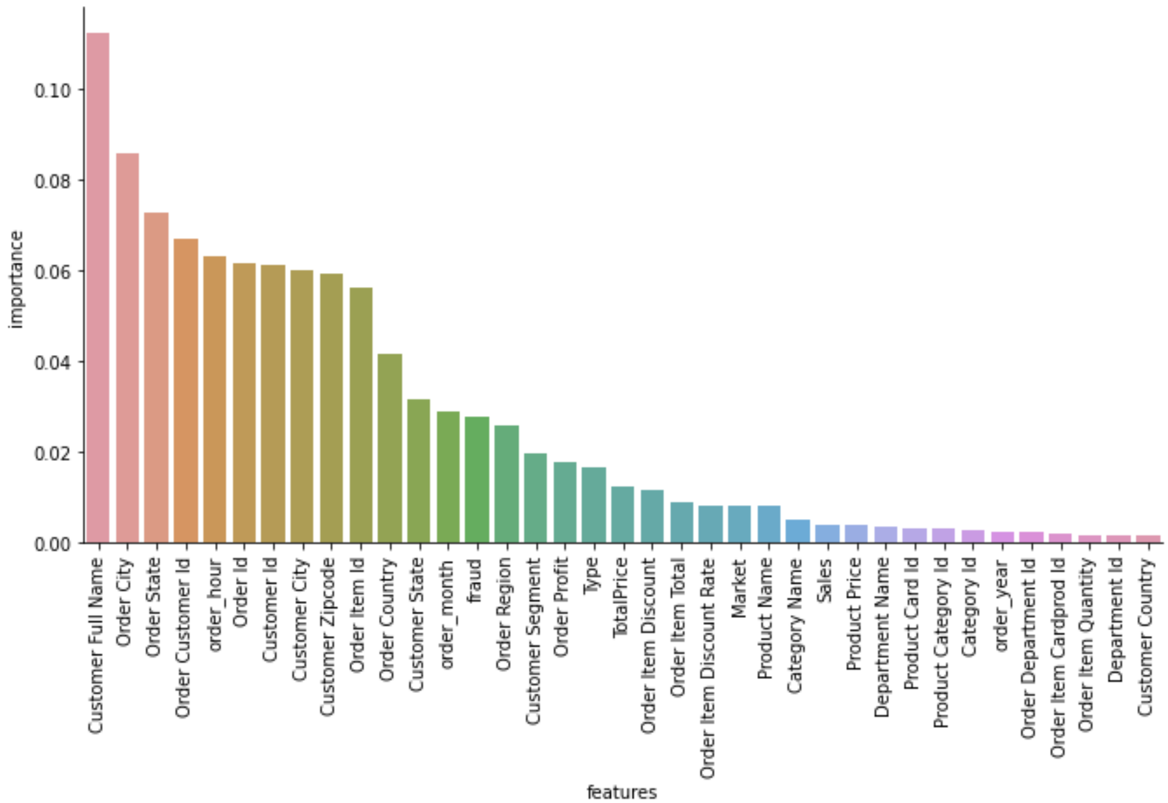

Figure 8: Feature Importance After Removal of "Days for shipping (real)" and "Days for shipment (scheduled)"

As seen in Figure 8, now factors like "Order City" and "Order State" shows more importance, which suggest company use different strategies for different order region to reduce late deliveries.

4.5 Conclusion

This sections uses Decision Tree Classifiers to implement Fraud Detection and Late Delivery Prediction with satisfying performance, which can help company's predict future frauds and late deliveries and alter their strategies to tackle these problems. It is worth noting that this model is based on assumption that "suspected frauds" are all or mostly actually frauds. It should only work when company are sure about which orders are real frauds.

5 Time Series Analysis of the data set

5.1 Introduction

In this section, we use models and concepts for time series analysis to make predictions, find out more about the seasonality patterns, Anomalies, Outliers and Change-Points in the data,. Due to computations limits and some missing data in the data set, we will focus our analysis on the daily **Order Profit** for all customers and in all markets.

We will also illustrate the use of the *ARMA* process and the *Ruptures* algorithm to analyse the orders that are suspected to be fraudulent.

5.2 Exploring the order profit

We make a three plots of the daily profits. first concerns the daily profit of the all orders including the ones with and without suspected frauds. The second concerns the daily profit of orders without suspected frauds. The last and the third plot concerns only the daily profit of the orders with suspected frauds.

Figure 9: Order Profit for all orders

We can notice that the non suspected fraud cases are still making huge loss in certain days which makes us assume that there are either problems with deliveries, cancelled order refunds or some hidden frauds the company couldn't detect. In certain days The suspected fraud cases are still making good profit, which make us assume that either the company didn't reach the expected profit or there is an inaccuracy in the data set's labels.

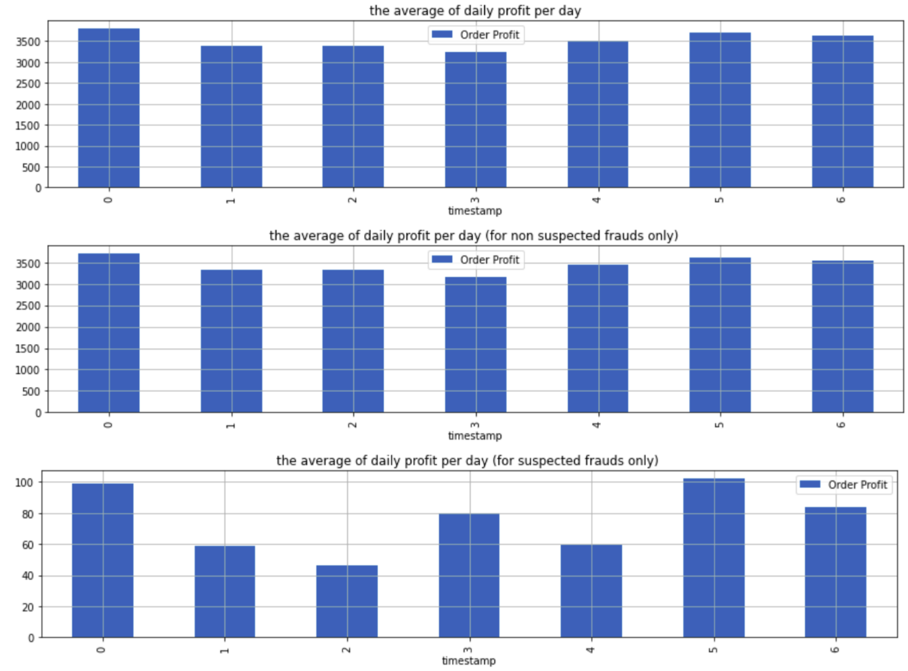

Figure 10: The average daily profit per day of the week

We can notice that the suspected frauds made generally lesser profit in the middle of the week, the least daily Order Profit is on Tuesday.

5.3 Seasonality Analysis

The purpose of this section is detect the seasonality of the signal which is the periodical component in the signal.

We tested two trends estimation for the signal, constant trend and polynomial trend.

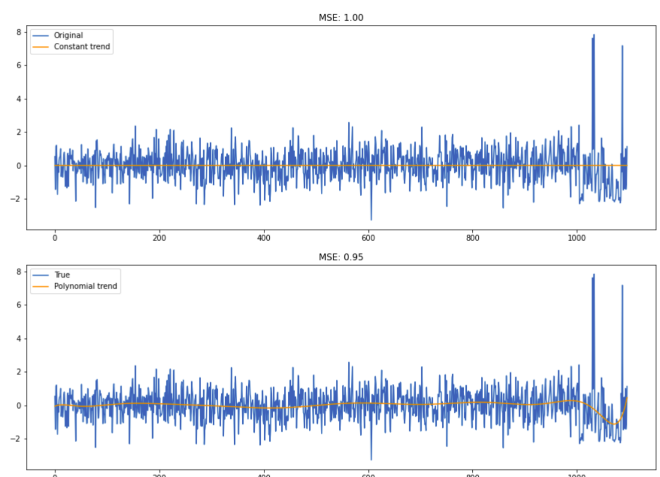

Figure 11: The Mean Squared Error and the trend

Let's estimate the main seasonality from the original signal with the periodogram of Power spectral density.

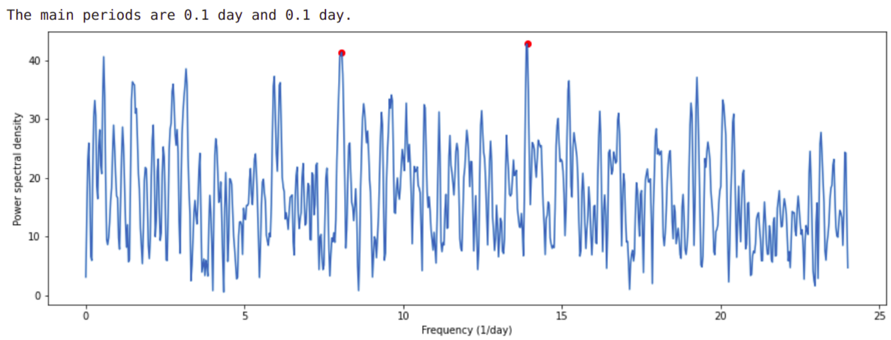

Figure 12: The periodogram of Power spectral density

We observe a periodogram that has two tightly separated spikes, and a signal that is aperiodic.

5.3.1 Harmonic regression

In an harmonic regression (with two harmonic components), the signal is modelled as follows:

$$y_t = \mu + A_1 \cos(2\pi f_1 t + \phi_1) + A_2 \cos(2\pi f_2 t + \phi_2) + \epsilon_t$$

where $\mu, A, \phi \in R$ must be estimated, the frequencies f_1 and f_2 are given, and ϵ_t is a white noise.

Let's extract the seasonal component with the Harmonic regression.

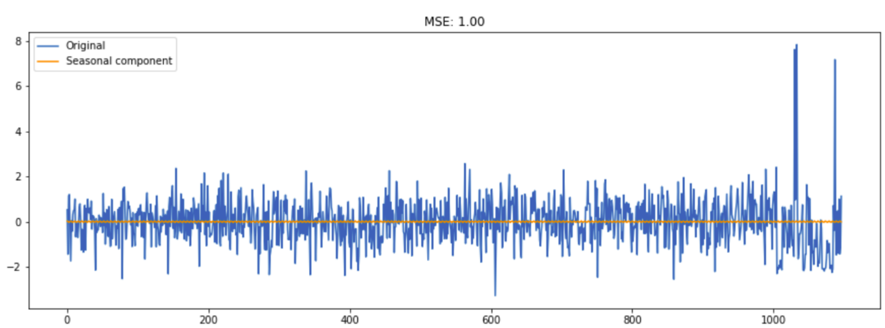

Figure 13: The periodogram of Power spectral density

The seasonal component of the daily orders(without suspected frauds) follows a constant trend.

5.3.2 Arima model

Here we used the Arima model and fitted it on the daily profit(**without suspected frauds**) with parameters (8,0,10). THen we ploted the residual signal of the daily profit (**with suspected frauds**) for comparison. The R-squared of the two residuals is 98%.

```
  ConvergenceWarning)
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                 1097
Model:                ARIMA(10, 0, 8)   Log Likelihood               -1535.988
Date:                Sun, 13 Mar 2022   AIC                           3111.977
Time:                        18:17:02   BIC                           3211.984
Sample:                             0   HQIC                          3149.815
                               - 1097
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0161      0.054     -0.299      0.765      -0.121       0.089
ar.L1         -0.1867      0.793     -0.235      0.814      -1.741       1.367
ar.L2         -0.3053      0.533     -0.572      0.567      -1.351       0.740
ar.L3          0.1321      0.388      0.341      0.733      -0.628       0.892
ar.L4          0.0589      0.312      0.189      0.850      -0.553       0.671
ar.L5         -0.1859      0.314     -0.592      0.554      -0.802       0.430
ar.L6          0.5494      0.355      1.547      0.122      -0.147       1.246
ar.L7          0.1746      0.539      0.324      0.746      -0.882       1.231
ar.L8          0.3816      0.323      1.180      0.238      -0.252       1.015
ar.L9         -0.0806      0.068     -1.180      0.238      -0.214       0.053
ar.L10         0.0014      0.085      0.017      0.987      -0.165       0.168
ma.L1          0.3100      0.793      0.391      0.696      -1.244       1.864
ma.L2          0.4203      0.533      0.789      0.430      -0.624       1.464
ma.L3          0.0090      0.353      0.025      0.980      -0.683       0.702
ma.L4         -0.0146      0.306     -0.048      0.962      -0.614       0.585
ma.L5          0.1997      0.307      0.650      0.516      -0.403       0.802
ma.L6         -0.5552      0.344     -1.613      0.107      -1.230       0.119
ma.L7         -0.2049      0.539     -0.380      0.704      -1.261       0.852
ma.L8         -0.3985      0.336     -1.187      0.235      -1.056       0.259
sigma2         0.9655      0.029     33.108      0.000       0.908       1.023
===================================================================================
Ljung-Box (L1) (Q):                   0.08   Jarque-Bera (JB):              3256.69
Prob(Q):                              0.78   Prob(JB):                         0.00
Heteroskedasticity (H):               2.04   Skew:                             1.08
Prob(H) (two-sided):                  0.00   Kurtosis:                        11.16
===================================================================================
```

Figure 14: Fitting The Arima model

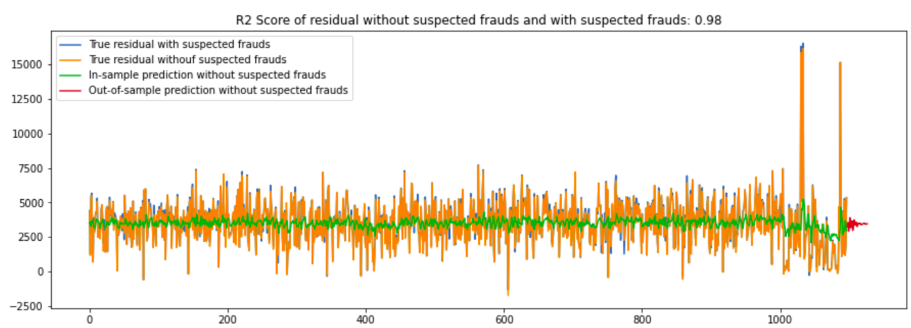

Figure 15: Prediction of the residual signals with Arima

Conclusion:

The residuals of the signal without suspected frauds don't differ much of the residuals of the signal with suspected frauds suspected frauds. We assume that the suspected frauds might not be entirely frauds, and perhaps the data contains more actual frauds which the company couldn't label as suspected frauds, for these reasons we can't confirm if the suspected frauds are actual frauds or not.

5.3.3 Stationarity checks

As we can tell from the previous subsection(4.3.2), the residual signal seem to be trend/level stationary.
To verify the stationarity of the signal, we will use theDickey-Fuller and the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests.

```
Results of Dickey-Fuller Test:
Test Statistic                -1.477067e+01
p-value                        2.325145e-27
#Lags Used                     2.000000e+00
Number of Observations Used    1.124000e+03
Critical Value (1%)           -3.436181e+00
Critical Value (5%)           -2.864115e+00
Critical Value (10%)          -2.568141e+00
dtype: float64
Results of KPSS Test:
Test Statistic            0.518116
p-value                   0.037586
Lags Used                13.000000
Critical Value (10%)      0.347000
Critical Value (5%)       0.463000
Critical Value (2.5%)     0.574000
Critical Value (1%)       0.739000
dtype: float64
```

Figure 16: ADFULLER and KPSS tests

The null hypothesis in the Dickey-Fuller test is that the signal has a unit root. The p-value is very less than the significance level of 0.05, therefore we can reject the null hypothesis. Threfore the residual signal doesn't have a unit root.

The null hypothesis of the KPSS test assumes that residual signal is trend/level stationary. Since the p-value is greater than the significance level of 0.05. Hence, we accept the null hypothesis.

5.4 Outliers Detection

Since we couldn't confirm with the ARIMA model if the suspected frauds are actual frauds or not, we want to detect outliers, which would help the company target the anomalies in the daily profit for their analysis.

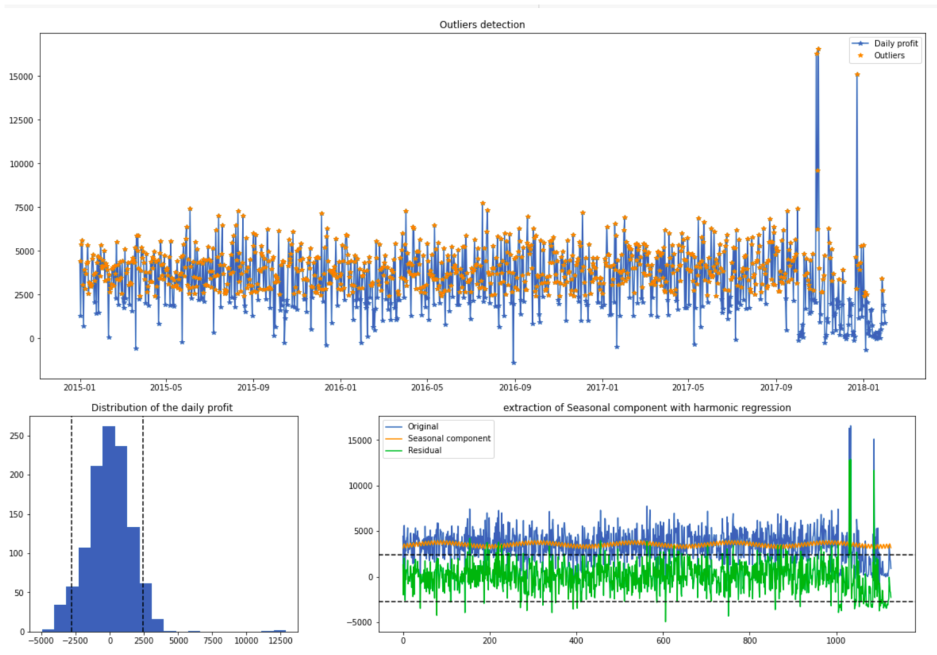

Figure 17: outliers detection with quantile thresholds = (0.05, 0.95)

5.5 Change-Point detection

To further analyse the anomalies of the signal, we chose to use the **Ruptures** algorithm developed to detect change points in time series. This algorithm is very useful for the segmentation of non-stationary signals.

Due to computation limits, we chose to only analyse the first 1000 samples of the signale and compare the change points detected by the algorithm with the 12 suspected frauds orders in these samples.

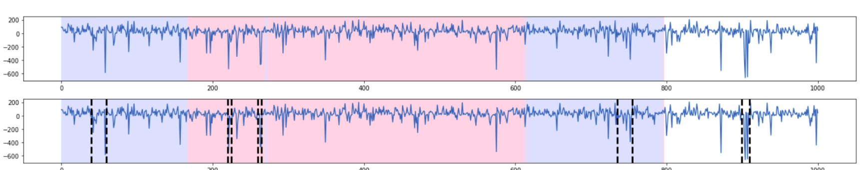

Figure 18: Change point detection with number of breakpoints =12.

The result of our analysis shows that none of suspected frauds is listed as a change point. Why?

We assume that there are many reasons for our findings: - using the daily profit is not good, because if we have more products that are making huge profit but lesser orders that were fraudulent, the daily profit won't change significantly. We believe that this is the main problem with our method, but due to computation limitation we couldn't proceed with analysing the order of specific products and specific regions. - these suspected frauds might not be frauds, which might be accurate with the fact that in some days the dialy profits of only suspected frauds were positive and huge.

5.6 Conclusion of Time Series Analysis

In this section we tried to use time series analysis to the daily profit and check if the orders with frauds are actual frauds or not. We didn't get a clear answer to this problem since the data set seems to probably contain more frauds than the orders with suspected frauds. We also suspect that the orders with suspected frauds are not actual frauds, so the company should verufy if the data is well labed (we suspect it's not the case). We provided two methods to detect outliers and detecting change points. The signal does have a lot of outliets and chane-points that need further analysis to relate them to fraud or to anomalies in the management of deliveries.

6 Anomaly Detection and Forecasting with Deep Learning

6.1 Background

Two central challenges in supply chain management are how to detect anomalies and forecast demands. Managers expect fraud orders to be detected for further root cause analytics. Besides, the balance of supply and demand is crucial in retail inventory planning.

The volume of supply chain transactions is overwhelming. Lower human bias and high efficiency can be achieved by leveraging machine learning-based methods. To deal with these specific questions, half of the Best-in-Class logistics companies are investing in AI capabilities. In the study, we carried out a time-series LSTM Autoencoder scheme to automate those processes by following methods proposed by [Nguyen, H.D., et al.].

6.2 The Necessary Algorithms

Let's briefly review some key concepts in machine learning for our proposed method.

Long short-term memory (LSTM) is a special architecture in recurrent neural networks (RNN) suitable for sequential supply chain data. Each cell contains a forget gate, an input gate and an output gate. It is the weight matrix that contains dependencies between the variables. To prevent gradient vanishing/explosion issues, we chose Rectified Linear Unit (ReLU) function as the activation.

Autoencoder is an unsupervised network for dimension reduction. The encoder compresses the input vector x $\in R^n toz \in R^d, whered < nandziscalledthelatentvector, whilethedecoderdecompresseszintothe\hat{x}.OurproposedLS$

6.3 Proposed Methods

6.3.1 Data engineering

We split the raw 180,519 data into a training set (80%) and test set(20%). Let $x^{(i)}$ denote a time series average profit of order i. Each observed data was re-scaled by using the formula

$x^{(i)}_{scaled} = \frac{x^{(i)} - x^{(i)}_{min}}{x^{(i)}_{max} - x^{(i)}_{min}}$

6.3.2 Time series forecasting using LSTM

The LSTM networks were trained based on a sequence of orders $x^{(1)}, x^{(2)}, ..., x^{(N)}$, where N is the number of training data. We assigned a sliding window size m=10, which is the number of data fed into the LSTM simultaneously. More rigorous assignment of window size can be regarded as a further research topic. Too small m implies loss of information about dependencies between neighboring data, while too large m will increase computation exponentially. In our study, setting m=10 seems reasonable because the previous time-series data are summarized by the memory cell in LSTM. The value of average profit at order m+1, i.e. $x^{(\hat{m+1})}$ can be predicted based on the sequence $x^{(1)}, x^{(2)}, ..., x^{(m)}$. The process will be terminated when the window slide to the final training data. We defined the loss function as

$L = \sum_{i=m+1}^{N} \|\hat{x}^{(i)} - x^{(i)}\|.$

The weight parameters within LSTM networks were trained by minimizing L. After training, managers can use the model for prediction. For example, the profit at order N+1 can be forecasted by feeding the observed data $x^{(N-m+1)}, x^{(N-m+2)}, ..., x^{(N)}$ into the LSTM nets.

In our method, the learning rate, the number of cells, and the dropout are hyperparameters, which were optimized by performing a grid search.

6.3.3 Anomaly detection using LSTM Autoencoder and SVM

[Nguyen, H.D., et al.] proposed that using LSTM alone is not efficient for anomaly detection in some situations as the dependence of the predicted average profit on the predicted values of other variables (i.e: order city, product ID, department ID)is ignored. The Autoencoder LSTM method for anomaly detection is proposed as follows.

The same notations are adopted as the previous section. The input sequence $X^{(i)} = x^{(i)}, x^{(i+1)}, ..., x^{(i-m+1)}$, where m is the sliding window size. After feeding into the trained LSTM Autoencoder, it was recreated as $\hat{X^{(i)}} = \hat{x^{(i)}}, \hat{x^{(i+1)}}, ..., \hat{x^{(i-m+1)}}$, i = m+1,. . . N. We calculated the prediction error vectors for each sequence $\hat{e^{(i)}} = \hat{X^{(i)}} - X^{(i)}$. Then, we performed anomaly detection on those error vectors by using single-class Support Vector Machines. One possible future research direction is to investigate which statistical models are appropriate for describing the error vectors.) The advantage of using SVM is that it does not rely on any probability distribution assumptions of the error vectors.

6.4 Experiments and Results

6.4.1 Average profits forecasting

We implemented our proposed methods on the supply chain management data of Big Supply Co. The objective was to forecast the average profit of each order in the testing set by using the data in the training set. To do so, the training data were fed to train parameters with the LSTM Autoencoder model and the trained model can be utilized to make predictions. The hyperparameters, including the learning rate, dropout and number of cells, were optimized by a grid search method to achieve a lower mean squared error (MSE). We used a sliding window of size 10, i.e., we take the data of 10 consecutive orders to predict the profit of the next order. The learned hyperparameters are as follows: #cells = 8, learning rate = 0.0001, dropout = 0.1, #epoch = 50. The process was iterative and thus time-consuming. After that, the model was retrained by using the optimal hyperparameters and then can be used to forecast average profits. We depicted a graph showing the comparison between the true average profits and the predicted ones in Figure 19. While Figure 20 is a sliced piece of Figure 19 for visibility, it shows that the forecasted profits are quite close to the ground truth.

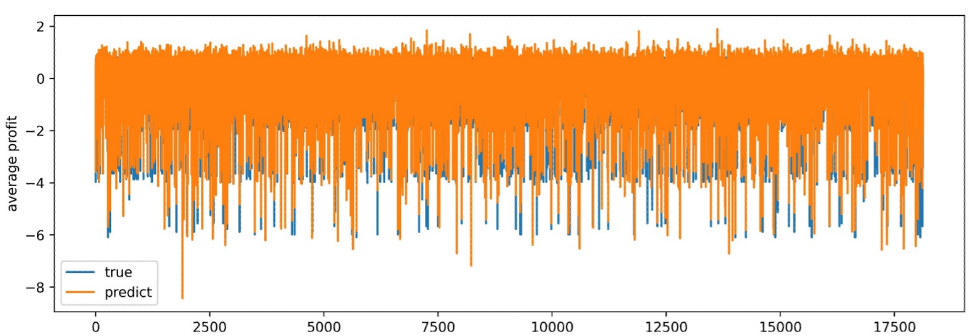

Figure 19: A comparison between real average profits and predicted profits using LSTM.

Figure 21 indicates that the Mean Squared Errors on both the training set and test set are decaying remarkably, which implies that the proposed algorithm converges.

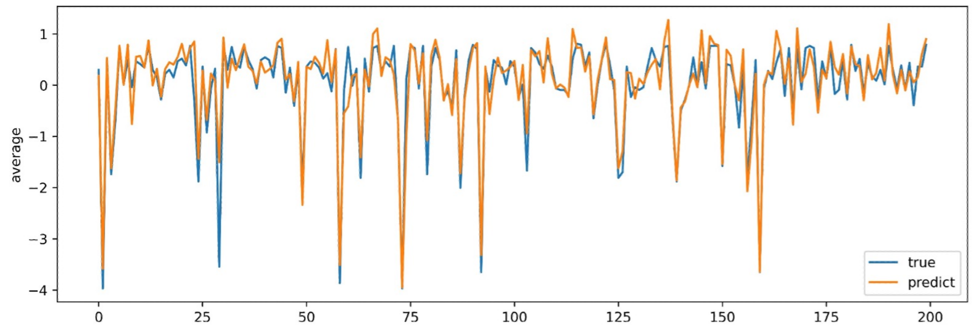

Figure 20: A fragment of Figure 1 for the first 200 orders.

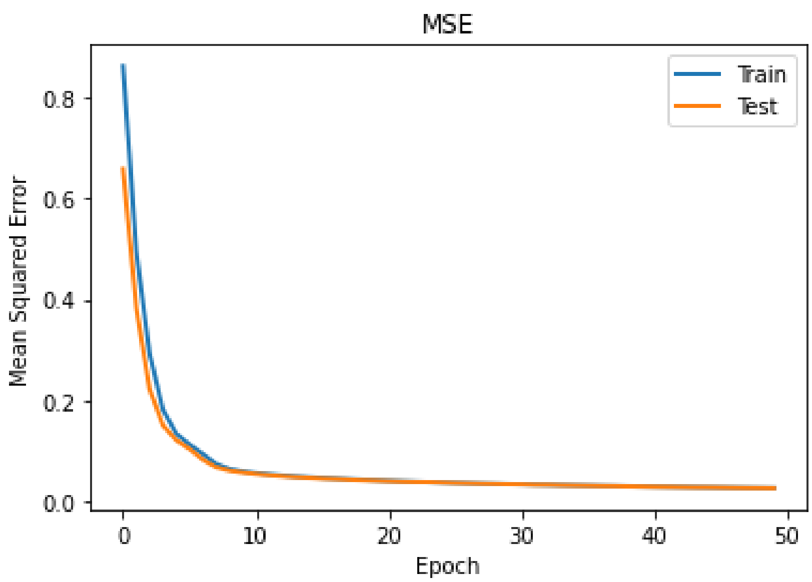

Figure 21: Plot of training and test error from the proposed model.

6.4.2 Anomaly detection.

The LSTM Autoencoder and SVM were trained on the same data set as in the previous section. Then, anomaly detection was done on the test set. To visualize the difference between normal data and anomalies, we calculated the error vector between the learned representation and the real data for each order. Then, we used Principal Component Analysis (PCA) to extract the two most significant eigenvalues. As shown in Figure 22, there is a clear boundary between the normal data (yellow) and the anomalies (purple). Therefore, the LSTM Autoencoder effectively extracted meaning features from the input data, and the SVM quite accurately classified anomalies versus normal orders.

In Figure 23, we highlighted all the classified anomalies with red crosses. Pointing out those anomaly data automatically with AI can save much time for supply chain managers. They can quickly find them and then try to investigate what factors caused these frauds with their specialist knowledge. Figure 24 is a slice of Figure 23 for clearer visualization.

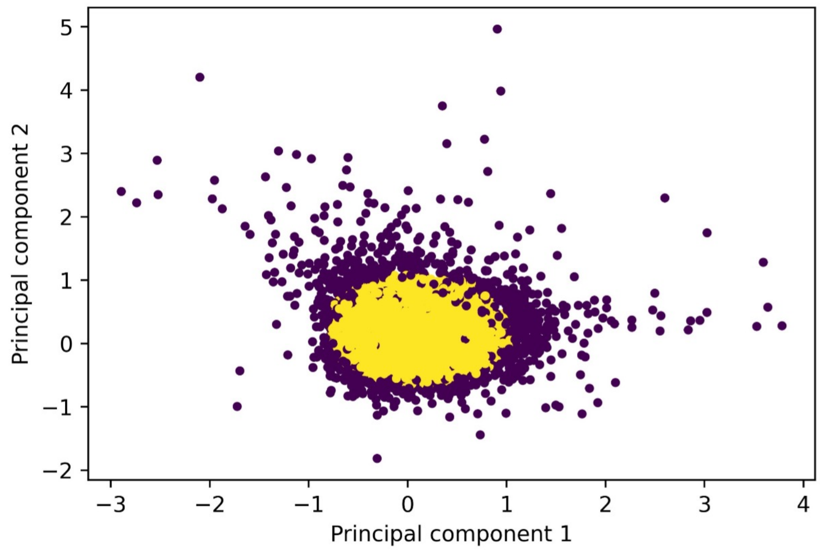

Figure 22: Comparisons between representation learned from LSTM Autoencoder and the orginal data via PCA.

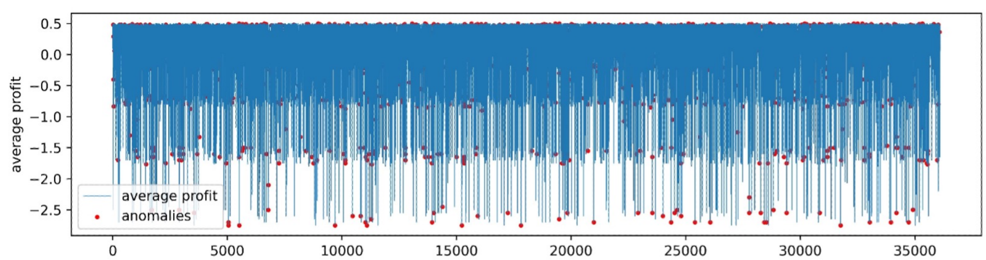

Figure 23: The anomaly detection for real data based on LSTM Autoencoder networks and SVM.

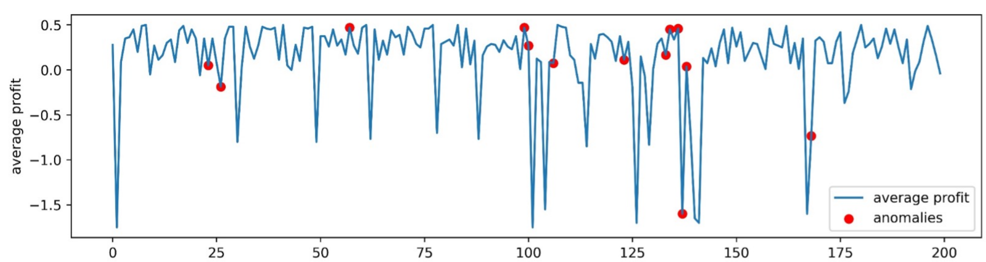

Figure 24: A fragment of Figure 5 for the first 200 orders.

We evaluated the performance of our method by the following measures:

· Accuracy $= \frac{TP+TN}{TP+FP+TN+FN} = 0.958302163854487$

· Precision $= \frac{TP}{TP+FP} = 1.0$

· Recall $= \frac{TP}{TP+FN} = 0.958302163854487$

· F-score $= 2 \times \frac{Precision \times Recall}{Precision + Recall} = 0.48935357451083034$

Where TP (True Positive) stands for the number of anomalies correctly diagnosed as anomalies, TN (True Negative) stands for the number of normal orders classified as normal, FP (False Positive) Stands for the number of normal orders incorrectly classified as anomalies, and FN (False Negative) stands for the number of frauds incorrectly diagnosed as normal. By definition, Precision measures how accurate the prediction is, while Recall evaluates how complete the anomalies are diagnosed. To trade-off between the two conflicting objectives, we defined F-score.

6.5 Conclusion

In the section, we tried to address the two most critical questions in supply chain management, including forecasting and anomaly detection, by applying LSTM Autoencoder networks and SVM algorithms. The experiments showed that our proposed approach worked quite well in terms of low errors and high accuracy. In the future, we expect to exploit more input features in the supply chain data and to perform root cause analysis, which would improve the interpretability as opposed to to the black-box neural networks-based method.

References

[1] Badvelu, J. (2020). COMPARATIVE-STUDY-OF-DIFFERENT-MACHINE-LEARNING-MODELS-FOR-SALES-PREDICTION-AND-FRAUD-DETECTION. [online] GitHub. Available at: https://github.com/JaswanthBadvelu/Comparision-of-MLmodels-with-RNN [Accessed 13 Mar. 2022].

[2] GitHub. (n.d.). deepcharles - Overview. [online] Available at: https://github.com/deepcharles [Accessed 13 Mar. 2022].

[3] Nguyen, H.D., et al. "Forecasting and Anomaly Detection Approaches Using LSTM and LSTM Autoencoder Techniques with the Applications in Supply Chain Management." International Journal of Information Management, vol. 57, Apr. 2021, p. 102282, 10.1016/j.ijinfomgt.2020.102282.

[4] [Truong2020] Truong, C., Oudre, L., Vayatis, N. (2020). Selective review of offline change point detection methods. Signal Processing, 167. [abstract] [doi] [pdf]

[5] www.laurentoudre.fr. (n.d.). Laurent Oudre - Home Page. [online] Available at: http://www.laurentoudre.fr/ast.html [Accessed 13 Mar. 2022].

7 Appendix

7.1 Appendix A

The table below summarizes all distributions that were found to be statistically different from all other regions.

Region	Feature	Statistic	Threshold	P-value
Central America	Order Item Cardprod Id	0.055	0.009	0
Central America	Order Customer Id	0.056	0.009	0
Central America	Order Item Id	0.42	0.009	0
Central America	Order Item Quantity	0.028	0.009	0
Central America	Sales	0.016	0.009	0
Central America	Order Item Total	0.015	0.009	0
South America	Order Item Cardprod Id	0.05	0.012	0
South America	Order Customer Id	0.05	0.012	0
South America	Order Item Id	0.399	0.012	0
South America	Order Item Quantity	0.024	0.012	0
South America	Sales	0.022	0.012	0
South America	Order Item Total	0.019	0.012	0
Caribbean	Order Item Cardprod Id	0.048	0.015	0
Caribbean	Order Customer Id	0.05	0.015	0
Caribbean	Order Item Id	0.387	0.015	0
Caribbean	Order Item Quantity	0.031	0.015	0
Northern Europe	Order Item Cardprod Id	0.026	0.014	0
Northern Europe	Order Customer Id	0.033	0.014	0
Northern Europe	Order Item Discount	0.03	0.014	0
Northern Europe	Order Item Id	0.315	0.014	0
Northern Europe	Sales	0.037	0.014	0
Northern Europe	Order Item Total	0.037	0.014	0
Northern Europe	Order Profit	0.023	0.014	0
Western Europe	Order Item Cardprod Id	0.028	0.009	0
Western Europe	Order Customer Id	0.037	0.009	0
Western Europe	Order Item Discount	0.018	0.009	0
Western Europe	Order Item Id	0.363	0.009	0
Western Europe	Order Item Quantity	0.021	0.009	0
Western Europe	Sales	0.033	0.009	0
Western Europe	Order Item Total	0.032	0.009	0
Western Europe	Order Profit	0.02	0.009	0
Southern Europe	Order Item Cardprod Id	0.025	0.014	0
Southern Europe	Order Customer Id	0.033	0.014	0
Southern Europe	Order Item Id	0.314	0.014	0
Southern Europe	Sales	0.027	0.014	0
Southern Europe	Order Item Total	0.026	0.014	0
Eastern Asia	Order Item Cardprod Id	0.14	0.016	0
Eastern Asia	Order Customer Id	0.156	0.016	0
Eastern Asia	Order Item Discount	0.026	0.016	0
Eastern Asia	Order Item Id	0.406	0.016	0
Eastern Asia	Order Item Quantity	0.063	0.016	0
Eastern Asia	Sales	0.053	0.016	0
Eastern Asia	Order Item Total	0.05	0.016	0
Eastern Asia	Order Profit	0.029	0.016	0
Southeast Asia	Order Item Cardprod Id	0.145	0.014	0
Southeast Asia	Order Customer Id	0.161	0.014	0
Southeast Asia	Order Item Discount	0.034	0.014	0
Southeast Asia	Order Item Id	0.419	0.014	0
Southeast Asia	Order Item Quantity	0.06	0.014	0
Southeast Asia	Sales	0.053	0.014	0
Southeast Asia	Order Item Total	0.051	0.014	0
Southeast Asia	Order Profit	0.034	0.014	0
Oceania	Order Item Cardprod Id	0.101	0.014	0
Oceania	Order Customer Id	0.117	0.014	0
Oceania	Order Item Id	0.451	0.014	0
Oceania	Order Item Quantity	0.05	0.014	0

Region	Feature	Statistic	Threshold	P-value
Oceania	Sales	0.041	0.014	0
Oceania	Order Item Total	0.04	0.014	0
Oceania	Order Profit	0.028	0.014	0
South Asia	Order Item Cardprod Id	0.105	0.016	0
South Asia	Order Customer Id	0.119	0.016	0
South Asia	Order Item Id	0.294	0.016	0
South Asia	Order Item Quantity	0.047	0.016	0
South Asia	Sales	0.034	0.016	0
South Asia	Order Item Total	0.033	0.016	0
South of USA	Order Item Cardprod Id	0.047	0.022	0
South of USA	Order Customer Id	0.048	0.022	0
South of USA	Order Item Id	0.443	0.022	0
West of USA	Order Item Cardprod Id	0.048	0.016	0
West of USA	Order Customer Id	0.052	0.016	0
West of USA	Order Item Id	0.453	0.016	0
West of USA	Order Item Quantity	0.029	0.016	0
US Center	Order Item Cardprod Id	0.051	0.018	0
US Center	Order Customer Id	0.055	0.018	0
US Center	Order Item Id	0.448	0.018	0
US Center	Order Item Quantity	0.029	0.018	0
East of USA	Order Item Cardprod Id	0.048	0.017	0
East of USA	Order Customer Id	0.049	0.017	0
East of USA	Order Item Id	0.451	0.017	0
Central Africa	Order Customer Id	0.058	0.033	0
Central Africa	Order Item Id	0.576	0.033	0
North Africa	Order Item Cardprod Id	0.047	0.024	0
North Africa	Order Customer Id	0.061	0.024	0
North Africa	Order Item Id	0.581	0.024	0
Eastern Europe	Order Item Cardprod Id	0.051	0.022	0
Eastern Europe	Order Customer Id	0.06	0.022	0
Eastern Europe	Order Item Id	0.584	0.022	0
Eastern Europe	Order Item Quantity	0.041	0.022	0
West Asia	Order Item Cardprod Id	0.048	0.018	0
West Asia	Order Customer Id	0.048	0.018	0
West Asia	Order Item Id	0.591	0.018	0
West Asia	Order Item Quantity	0.031	0.018	0
West Asia	Order Item Total	0.032	0.018	0
West Africa	Order Item Cardprod Id	0.047	0.023	0
West Africa	Order Customer Id	0.065	0.023	0
West Africa	Order Item Id	0.583	0.023	0
Southern Africa	Order Item Id	0.575	0.04	0
Central Asia	Order Item Id	0.574	0.058	0
Canada	Order Customer Id	0.095	0.044	0
Canada	Order Item Id	0.575	0.044	0
Canada	Days for shipping (real)	0.07	0.044	0
East Africa	Order Customer Id	0.078	0.032	0
East Africa	Order Item Id	0.578	0.032	0