

Spare Parts Demand Forecasting

Gonzalo Atienza Lama, *Master's Degree Thesis*, Master in Connected Industry, Universidad Pontificia de Comillas - ICAI.

Supervised by Tilak Singh, Mitsubishi Fuso Truck and Bus Corporation (MFTBC)

Resumen—Este proyecto tiene como objetivo averiguar la situación actual de MFTBC* respecto a la predicción de la demanda de piezas de repuesto. Actualmente la empresa tiene ciertas herramientas implementadas para ayudarles a realizar dicha predicción. A lo largo de este proyecto se ha realizado un análisis de los resultados obtenidos mediante los métodos ya implementados y se han implementado nuevos modelos con el objetivo de mejorar el rendimiento del modelo en los grupos de piezas en los que el rendimiento actual no es bueno.

Gracias al análisis realizado en el modelo de predicción actual se descubrió que el rendimiento de dicho modelo en piezas que tienen una demanda y un precio bajo es bastante malo, por lo que se decidió que los nuevos modelos entrenados deberían centrarse en mejorar la predicción en dichos grupos. Tras probar diferentes modelos y realizar diferentes experimentos se observó que los modelos ‘gradient boosting trees’ dan muy buenos resultados, llegando a ser mejores que los métodos de predicción actuales en ciertos casos.

Abstract—This project aims to understand the current situation of MFTBC* regarding the spare part’s demand forecasting topic. Currently, the company has some tools already implemented to help them make this forecast. Throughout this project, an analysis of the results yielded by the already implemented methods has been carried out as well as the implementation of new methods with the aim of improving the performance of the model in the groups of parts in which the current method’s performance is not good.

Thanks to the analysis carried out for the current forecasting model, it was discovered that the performance of the model in parts that have low demand and low price is quite bad, so it was decided that the new models should focus on improving the performance in these groups. After testing different models and carrying out different experiments, it was observed that the ‘gradient boosting trees’ models give very good results, outperforming the current forecasting models in certain cases.

I. INTRODUCTION

During the last two decades companies have changed the way they operate, taking advantage of the new technologies that are appearing in the industry. One specific task that is gaining a lot of importance recently is demand forecasting. Thanks to the advancements in the machine learning or deep learning technologies, there are a huge number of different algorithms that can be used for this task. However, although there are many different models that can be used, not all them have a good performance, so, usually, companies need to try different models until they find the optimal algorithm. This is a crucial activity for companies operating in certain industries.

The demand forecasting problem is a very important task for companies that usually manage huge amount of parts. For instance, spare part demand forecasting is an extremely important activity for companies operating in the automotive or the aeronautical industries, not only because the amount of spare parts they usually manage is extremely high, but also because the demand profile of some of the spare parts is not continuous but intermittent, meaning there are several time periods where there is no demand at all. Due to this fact it is very important to have a reliable forecast algorithm in order to have a good supply chain management and avoid high inventory costs.

The fact of having spare parts with strange demand profiles makes it more difficult to find a forecast algorithm that yields good results. For this reason, companies usually separate spare parts according their demand characteristics [2][3]. More specifically, two parameters representing the demand regularity in time (ADI) and in quantity (CV²) [2][3] are computed for each spare part, and depending on the value of these parameters, the spare part is assigned to one of the four categories shown in Fig. 1. This categories represent demand profiles that are easy to forecast (Smooth category), demand profiles that usually require more complex forecast algorithms (Intermittent and Erratic categories), and demand profiles that are unforecastable (Lumpy category) [2][3]. Additionally, there is a fifth category, called ‘Non-moving’ that represent spare parts with absolutely no demand in the last 12 months.

$$ADI = \frac{\text{Total NAmount of Time Periods}}{\text{Amount of Time Period with Demand Greater than 0}} \quad (1)$$

$$CV^2 = \left(\frac{\text{Standat Deviation}}{\text{Mean}} \right)^2 \quad (2)$$

This document aims to show the results obtained when analyzing the current forecasting methods that are already implemented in MFBTC, analyzing, not only the forecast error but also other aspects of the forecast like the bias, or how the price or demand of the spare parts can affect the performance of the model. Additionally, this document also explains the new methods that were implemented and tested, giving a brief theoretical explanation of them and showing the results of analyzing the forecast error and the bias of these models. Finally, after analyzing all the new models implemented, the model with the most promising results will be selected and compared to the current methods to see if any improvement was achieved.

• Mitsubishi Fuso Bus and Truck Corporation (MFBTC) is an integral part of Daimler Truck AG, the commercial vehicles business of the Daimler Group [1].

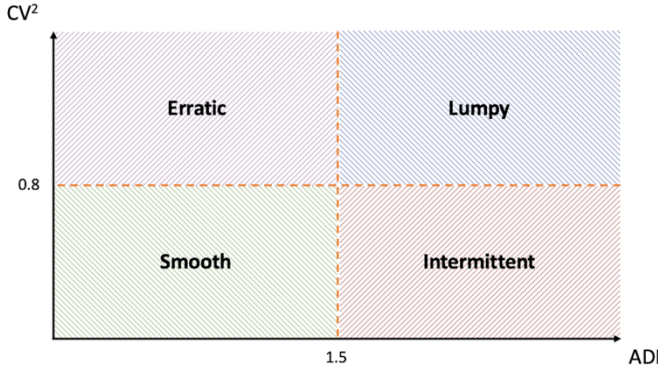


Fig. 1. Categories used for the initial classification of the spare parts.

II. CURRENT METHODS ANALYSIS

A. Exploratory Analysis of the data

Before showing any results an exploratory analysis of the data was performed. When performing this analysis on the dataset, there were several things that were discovered and that need to be explained, because some important decisions that affected the analysis were taken due to these discoveries.

Firstly, it was noticed that the spare parts included in the dataset could belong either to the Dubai plant (plant 5756), or to the Atsugi plant (plant 5758). Also, inside each plant, spare parts could belong to the domestic market (FD) or the export market (FE). It is important to know this because many of the results in this document group spare parts by plant and market.

Secondly, it was discovered that the company has two different forecasting methods, called APO and Big Data. Analyzing the forecasts of these methods it was noticed that the Big Data method is only available for plant 5758 and that it was implemented in August of 2017. Finally, it was also discovered that the company recently decided to stop using the Big Data method, reason why the current method analysis was only performed for the APO forecast. However, a comparison between APO and Big Data is still performed to see if stopping using the Big Data method is a good decision.

Additionally, the dataset used for this project contained information from March 2016 to February 2019, however, this dataset included some variables that were computed using only the last 12 months of data available, so it was decided that the APO analysis was going to be performed using only the last 12 months of data available (March 2018 to February 2019).

Finally, it was seen that, when separating the spare parts by their demand profile there were only 10% of the spare parts belonging to the smooth category (Table I), however, it was also seen that almost 90% of the demand came from this category, showing that these spare parts are extremely important. This led to the decision of focusing this project just on the smooth spare parts, trying to see the performance of the current methods in this segment of spare parts and trying to improve this performance on the groups of spare parts where the current method is not giving good results.

To sum up, all the project was focused on the smooth spare parts and the analysis performed only included the last 12 months of data. Additionally, the comparison between APO and Big Data was done only with spare parts from plant 5758.

TABLE I
RESULTS OF CLASSIFYING THE SPARE PARTS DEPENDING ON THEIR DEMAND PROFILE AND THE PERCENTAGE OF THE TOTAL DEMAND COMING FROM EACH CATEGORY.

Category	Amount of spare parts	Percentage of demand (%)
Non-moving	137,363	0.00
Smooth	25,221	89.58
Intermittent	44,720	5.45
Erratic	3,331	3.17
Lumpy	41,900	1.8

B. APO and Big Data comparison results

This comparison was carried out with the objective of discovering if one of these methods is clearly better than the other and to see if stopping using the Big Data was a good idea. The error metric used in this comparison and in the rest of the analysis was the mean average percentage error (MAPE), which is very useful in this project since there are many spare parts with different demand levels and, since the MAPE is a scale-free error, it is easier to compare the performance of the model in all spare parts.

In order to compare the Big Data method and the APO method, the MAPE of each of these methods was calculated for each individual spare part, and then they were compared in order to see which method was more accurate for each spare part. The results of this comparison can be seen in Fig. 2.

These results show there are almost 12,000 spare parts for which the APO method is better than Big Data and that there are approximately 3,000 spare parts for which both methods are equally good. This means there are 15,000 spare parts for which APO has an equal or better performance than Big Data. Knowing that this analysis has a total of 25,221 spare parts, these 15,000 spare parts represent almost 60% of the total.

Looking at these results it can be concluded that the APO method is better than the Big Data method and it seems that stopping using the Big Data method was a good decision.

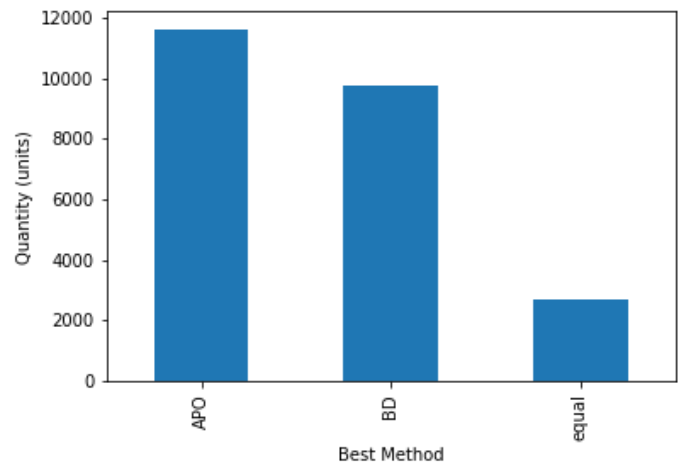


Fig. 2. Amount of spare parts for which each forecasting method is better (has a lower MAPE).

C. APO Analysis results

This analysis was performed on different aspect of the forecast, focusing mostly on the bias, the forecast error (MAPE) and the effect of the price and demand on the performance of the model.

The first aspect of the data that was analyzed was the price and demand of the spare parts, making two new classification processes, one for grouping the spare parts depending on their demand and another one grouping them by price. In both these classification processes there were 9 different groups, representing spare parts with very low price/demand (groups 1 to 3), spare parts with an average value for the price/demand (groups 4 to 6), and spare parts with high price/demand (groups 7 to 9). All the demand and price groups, along with the condition that the spare parts need to meet in order to be included in them can be seen in Table II and Table III, respectively.

TABLE II

CATEGORIES FOR THE DEMAND CLASSIFICATION ALONG THE CONDITION TO BE INCLUDED IN THEM.

1. - Low Low	0 <Demand <= 10
2. - Low Medium	10 <Demand <= 25
3. - Low High	25 <Demand <= 40
4. - Medium Low	40 <Demand <= 50
5. - Medium Medium	50 <Demand <= 70
6. - Medium High	70 <Demand <= 120
7. - High Low	120 <Demand <= 500
8. - High Medium	500 <Demand <= 1000
9. - High High	1000 <Demand

These two classification processes helped separate spare parts with low demand from spare parts with high demand and analyze the performance of the APO model in these groups separately. The same was done but separating spare parts with low price from spare parts with high price. The results obtained after performing these classification was very interesting, as most of the spare parts belonged to low price and low demand groups. The demand classification results can be seen in Table IV and the price classification results in Table V.

TABLE III

CATEGORIES FOR THE PRICE CLASSIFICATION ALONG THE CONDITION TO BE INCLUDED IN THEM (PRICE IN YENS).

1. - Low Low	0 <Price <= 500
2. - Low Medium	500 <Price <= 1000
3. - Low High	1000 <Price <= 2000
4. - Medium Low	2000 <Price <= 4000
5. - Medium Medium	4000 <Price <= 7000
6. - Medium High	7000 <Price <= 10000
7. - High Low	10000 <Price <= 20000
8. - High Medium	20000 <Price <= 50000
9. - High High	50000 <Price

Looking at the results, it is easy to realize that almost 70% of the spare parts are characterized by having a low price, belonging to price groups 1, 2 and 3, where the price of the spare parts is lower than 2,000 yens. The demand classification results are similar, having more than 80% of the spare parts with a demand lower than 40 units per month (demand groups 1, 2 and 3).

TABLE III

AMOUNT OF SPARE PARTS IN EACH PRICE CATEGORY (MIDDLE COLUMN) AND THE PERCENTAGE OVER THE TOTAL AMOUNT OF SPARE PARTS (RIGHT COLUMN). THE CLASSIFICATION WAS PERFORMED USING THE AVERAGE PRICE OF THE LAST 12 MONTHS OF DATA AVAILABLE.

Price_class	Price_cnt	Price_cnt_percentage
1 - Low Low	10383	41.31
2 - Low Medium	3558	14.16
3 - Low High	3175	12.63
4 - Medium Low	2768	11.01
5 - Medium Medium	1770	7.04
6 - Medium High	1002	3.99
7 - High Low	1265	5.03
8 - High Medium	631	2.51
9 - High High	583	2.32

Once the classification was done, the average MAPE of each of the price and demand groups was computed. The results obtained were very interesting as it showed that the APO method has more or less the same performance in all price groups, being the MAPE of all groups between 48% and 53%. The only two price groups in which APO seem to have a worse performance than in the rest of price groups are price group 8 and price group 9, where the MAPE obtained are 51.31 and 53.43, respectively.

TABLE IV

AMOUNT OF SPARE PARTS IN EACH DEMAND CATEGORY (MIDDLE COLUMN) AND THE PERCENTAGE OVER THE TOTAL AMOUNT OF SPARE PARTS (RIGHT COLUMN). THE CLASSIFICATION WAS PERFORMED USING THE AVERAGE DEMAND OF THE LAST 12 MONTHS OF DATA AVAILABLE.

Demand_class	Demand_cnt	Demand_cnt_percentage
1 - Low Low	15757	62.69
2 - Low Medium	4046	16.10
3 - Low High	1497	5.96
4 - Medium Low	591	2.35
5 - Medium Medium	692	2.75
6 - Medium High	835	3.32
7 - High Low	1283	5.10
8 - High Medium	252	1.00
9 - High High	182	0.72

On the other hand, when computing the MAPE for the demand groups it was noticed that there is huge difference in the performance of the APO forecast in high demand groups and in low demand groups. Looking at Table V, where the MAPE for the demand groups can be seen, it is easy to see that the MAPE of the high demand groups is very good, having a MAPE lower than 30% for demand group 8 and lower than 25% for demand group 9. However, the MAPE obtained for low demand groups is always higher than 40%, being higher than 50% for low demand groups. This analysis shows that low demand groups should be a priority when looking for new methods.

TABLE V
MAPE FOR EACH OF THE DEMAND GROUPS USING THE 12 LAST MONTHS OF DATA TO COMPUTE IT.

Demand_class	MAPE_Err
1 - Low Low	53.55
2 - Low Medium	45.46
3 - Low High	41.11
4 - Medium Low	38.99
5 - Medium Medium	38.35
6 - Medium High	37.17
7 - High Low	32.18
8 - High Medium	29.11
9 - High High	23.78

Additionally, when analyzing the low demand groups, it was found that, inside each demand group, the worst MAPE was obtained for very high price groups (price group 9) and the second worst MAPE was achieved for very low price groups (price group 1). Taking into consideration that there is a huge amount of spare parts in low price groups, it was decided that low demand and low price groups should be the groups for which the new methods had to give a better performance than the APO method.

The next aspect of the data that was analyzed was the forecasting error. For this, the MAPE was computed for each combination of plant and market (Table VI). Additionally, the weighted MAPE by orderlines (Table VII) was also computed in order to see what is the average error of each orderline.

TABLE VI
MAPE FOR EACH COMBINATION OF PLANT AND MARKET COMPUTING IT WITH THE LAS 12 MONTHS OF DATA AVAILABLE.

Plant_ID	Market_ID	MAPE_Err
5756	FE	66.21
5758	FD	46.51
	FE	60.93

Name: MAPE_Err, dtype: float64

The results obtained were a little alarming at first since all of the MAPE are very high, however if analyzed more closely, plant 5758 and market FD has the lowest MAPE, which is something good since most of spare parts belong to this plant and market. Also, looking at the weighted MAPE it can be seen that the error decreases almost 10%, showing that the average error of each orderline is less than 40% for plant 5758 market FD and it is higher than 50% for the other two combinations of plant and market.

TABLE VII
WEIGHTED MAPE BY ORDERLINES FOR EACH COMBINATION OF PLANT AND MARKET COMPUTING IT WITH THE LAS 12 MONTHS OF DATA AVAILABLE.

Plant_ID	Market_ID	WeightedMAPE
5756	FE	55.19
5758	FD	37.81
	FE	50.22

Name: WeightedMAPE, dtype: float64

It is also important to remember that most of the spare parts belong to the low price and low demand groups, where the MAPE is very high, meaning that these low price and low demand groups have a huge impact in the overall MAPE of the model. Additionally, since most of the spare parts belong to plant 5758 and market FD, the overall MAPE is very close to the MAPE obtained in that combination of plant and market, in fact, the overall weighted MAPE for APO is 39.67%, which is much closer to the weighted MAPE of plant 5758 market FD than to the weighted MAPE of the other combinations of plant and market.

Another aspect of the data that was analyzed was the manual forecast. This does not give any new insights on the performance of the APO method, however the results obtained in this analysis are very shocking and worth seeing. The spare parts included in this part of the analysis are spare parts for which the APO forecasted is omitted and, instead, a manually computed forecast is used.

The results of this showed that the manual forecast is performed on a 9.4% of the spare parts and that it happens, on average, between 9 and 10 times for each spare part in a 12 month period. When analyzing the MAPE before and after the manual forecast (Table VIII and Table IX, respectively), it was seen that the MAPE increases more than a 5% in plant 5758 and market FD, showing that the manual forecast is not working correctly. In fact, it was discovered that the manual forecast occurred 22,847 times during the last 12 months, and that it was more accurate than the APO forecast only 960 times, which is 4.2% of the times.

TABLE VIII
MAPE FOR EACH COMBINATION OF PLANT AND MARKET BEFORE MANUAL FORECAST (ONLY USING MANUALLY FORECASTED SPARE PARTS FOR COMPUTING THE MAPE).

Zfuso_Plant_Location	Zfuso_Market	MAPE_Err
5756	FE	60.12
5758	FD	68.32
	FE	63.61

Name: MAPE_Err, dtype: float64

The last aspect which is analyzed is the bias of the model. This is a very important aspect because, although the theory says that the forecast models should be unbiased, in this case it is preferable to have a slightly positive bias for the low price groups. This is mainly because the inventory cost of these spare parts is not very high, and through having a positive bias in the model it is ensured to always have some extra spare parts in the inventory in case they are needed. However, this only applies for low price spare parts. The model should still be unbiased for high price spare parts.

TABLE IX
MAPE FOR EACH COMBINATION OF PLANT AND MARKET AFTER MANUAL FORECAST (ONLY USING MANUALLY FORECASTED SPARE PARTS FOR COMPUTING THE MAPE).

Zfuso_Plant_Location	Zfuso_Market	MAPE_FinalForecast
5756	FE	60.12
5758	FD	73.65
	FE	54.29

Name: MAPE_FinalForecast, dtype: float64

For analyzing the bias, the average percentage error of each spare part was computed and then this value was analyzed so that is the average percentage error was negative, it meant that the spare part had a positive bias in the forecasts and if the percentage error was positive it meant that there was a negative bias in the forecast of that spare part. However, if the average percentage error was between -0.05 and 0.05, that spare part was said to have no apparent bias. The results of this analysis is shown in Table X.

TABLE X

BIAS RESULTS FOR EACH COMBINATION OF PLANT AND MARKET CALCULATING THE AVERAGE ERROR OF THE LAST 12 MONTHS OF DATA.

Zfuso_Plant_Location	Zfuso_Market	Material_Bias	
5756	FE	Negative bias	451
		Positive bias	365
		No apparent bias	87
5758	FD	Positive bias	9717
		Negative bias	8446
		No apparent bias	3568
	FE	Negative bias	1460
		Positive bias	748
		No apparent bias	293

Name: Material_Bias, dtype: int64

The results showed that there were more spare parts with a positive bias in plant 5758 and market FD, however the number of spare parts with a negative bias is still very high. The opposite happened in the other two combinations of plant and market, where there were more spare parts with a negative bias. This same analysis was performed but only for the low price groups, obtaining the results shown in Table XI.

TABLE XI

BIAS RESULTS FOR EACH OF THE LOW PRICE GROUPS CALCULATING THE AVERAGE ERROR OF THE LAST 12 MONTHS OF DATA.

Price_class	Material_Bias	
1 - Low Low	Positive bias	4926
	Negative bias	3839
	No apparent bias	1618
2 - Low Medium	Positive bias	1512
	Negative bias	1428
	No apparent bias	618
3 - Low High	Negative bias	1341
	Positive bias	1322
	No apparent bias	512

Name: Material_Bias, dtype: int64

The results of the bias analysis on low price groups showed that there are more spare parts with a positive bias in price groups 1 and 2, which is something beneficial as explained before. However, the number of spare parts with negative bias is still very high in all low price groups, so this is something that needs to be improved in future methods as well.

E. APO Analysis Conclusions

During this analysis there were several things that were discovered and that affected the decisions taken for the rest of the project. The first thing discovered was that the performance of the APO method in low demand groups is not very good, so it was decided to focus the new methods on improving the performance in this specific groups of spare parts, especially the low demand and low price groups.

Secondly, it was shown that the manual forecast is not giving good results, as most of the times the APO forecast is more accurate than the manual forecast. This analysis can help the company realize that something need to be done with respect to the manual forecast.

Thirdly, it was discovered that the APO performs much better on the combination of plant 5758 and market FD, which is something good as most of the spare parts belong to this combination of plant and market. On the other hand, the performance on the other two combinations of plant and market are very bad, however, it was decided that this is not a priority yet since improving the performance on low demand groups can also bring an improvement in these combination of plant and market.

Finally, it was shown that, although it seems there is an overall positive bias in the model and in the low price groups, there is still a huge amount of spare parts with a negative bias in their forecasts. This is something to consider in future models, although it is also true that this problem is not very difficult to solve.

III. NEW METHODS

As stated at the beginning, this project was only focused on the smooth spare parts. The smooth spare parts are characterized by a regular demand both in quantity and in time ^{[2][3]}, this means that the demand should not be very difficult to forecast using conventional methods. In fact, many of the paper that discuss this topic suggest trying machine learning algorithms for time-series forecasting, like ARIMA, ARMA, etc. ^[4] in this situation. Additionally, other more complex models can also be used for this problem, like LSTM ^[4] or gradient boosting trees ^{[4][6]}. Finally there are also some researchers that show how combining two of these models and creating a hybrid model ^[5] can give very good results in some cases.

For this project it was decided that 3 new method would be tried, carrying out different experiments with them. The three new methods that were implemented and tested were the gradient boosting trees method (XGB method from now on), an LSTM model and a hybrid model combining the XGB and the LSMT models.

The XGB method basically tries to treat the problem like a linear regression problem. The algorithm behind this model is known as 'additive regression trees' and it consist on creating and adding new models to the existing ones in order to improve the error. When adding new models, it uses the gradient descent to minimize the loss, reason why it is called gradient boosting tree method. This method has been implemented using the XGBoost library for python and using the GridSearchCV function from the sklearn python package to tune the parameters of the model.

The LSTM is a deep learning model, more specifically it is a type of recurrent neural network (RNN), which are known for their capability of holding information from the past in a very efficient way. The model developed in the project is a deep learning model, which consist on 6 different layers: A LSTM layer followed by a Dropout layer, then another LSTM layer followed by another Dropout layer, and, finally, a third LSTM layer followed by a Dense layer. This model was implemented using the keras library for python and using tensorflow backend.

Finally, the hybrid model is a combination of the previous two models. These model are combined by following 4 simple steps [5]. The first step is to train an initial XGB model and select the most important features for this model. The second step is to train an univariate LSTM model for the most important features. The third step is to include the outputs of the LSTM model as new features and, lastly, the fourth step is to retrain the XGB model with the new dataset.

During this part of the project there were a total of 8 different models trained, 5 of which were XGB models, 2 were LSTM models and 1 was the hybrid model. All these different models are explained below.

- XGB experiment 1: model trained using all the spare parts in the dataset.
- XGB experiment 2: models trained using only spare parts from plant 5758.
- XGB experiment 3: model trained using only spare parts from plant 5758 and market FD.
- XGB experiment 4: model trained using only spare parts from plant 5758 and combining the demand of the two markets in each plant.
- XGB experiment 5: model trained using only spare parts belonging to low demand and low price groups.
- LSTM experiment 1: model trained using all the spare parts in the dataset.
- LSTM experiment 2: model trained using all the spare parts in the dataset and normalizing the value of all the features.
- Hybrid model: model trained will all the spare parts in the dataset.

As it can be seen, the XGB method was chosen to be the one in which most of the time was focused. One of the reasons behind this decision is that the company is currently trying a new SAP tool, called SAP IBP, which provides different algorithms for forecasting the demand of the spare parts, one of which is an XGB algorithm. For this reason it was thought that it would be interesting to develop an XGB model from scratch and use the results obtained to decide what new features need to be created in the SAP tool to obtain the most optimal performance. Another reason behind this decision was that the XGB models are really fast to train, making it more convenient for trying many different experiments in order to find a model that outperforms the APO method.

For training these models a new dataset was created, including features like the price, the average demand, the average number of orderlines, etc. Also, for all the experiments the dataset was split between training, cross validation and test in the same way, having 6 time periods for testing, 1 time period for cross validation and 17 time periods for training. All the results shown (bias, MAPE, etc.) have been computed with the forecast for the 6 time periods used for testing the models.

A. Experiment Results

The results obtained from these different experiments were very good in some case and very bad in others. The results from the most promising experiments are shown later as well as commenting the performance of all the different models.

The XGB models gave very good results in general, with the only exception of experiment 2, where the MAPE obtained for each demand group was considerably worse than the one obtained for APO. The rest of the experiments gave very similar results, being the MAPE for some demand groups better than APO and others worse, however, since each experiment had a different amount of spare parts for computing the MAPE, it was hard to tell if they performed better or worse without making a proper comparison between the two models. However, after the analysis performed on the XGB experiments it was seen that the most promising experiments were experiment 3 and experiment 5.

The MAPE values obtained for each of the demand groups in experiment 3 can be seen in Table XII. It is easy to see that, although the MAPE for demand group 1 is a little high, the MAPE for demand groups 2 and 3 is really good, being 40% or less. Also, the results of this experiment for high demand groups is also extremely good, being less than 20% for demand group 9.

TABLE XII
MAPE OF EACH OF THE DEMAND GROUPS FOR XGB EXPERIMENT 3
COMPUTING IT WITH THE 6 FORECASTS OBTAINED DURING THE TESTING
PROCESS.

Demand_class	
1 - Low Low	57.150082
2 - Low Medium	40.040952
3 - Low High	35.461655
4 - Medium Low	32.189858
5 - Medium Medium	33.034974
6 - Medium High	29.881004
7 - High Low	25.990937
8 - High Medium	23.682740
9 - High High	18.355741

Name: MAPE_XGB, dtype: float64

The Bias analysis of this experiment also showed there is an overall positive bias in the model, having more than 11,000 spare parts with a positive bias. On the other hand, there are almost 7,000 spare parts with a negative bias. These bias results seem to be better than the bias results from APO, however, there is still a huge amount of spare parts with a negative bias and this problem should be fixed before implementing the final model.

Finally, the overall weighted MAPE for experiment 3 is 38.46%, showing the good performance of this experiment. However, as said before, this experiment has less time periods than the APO analysis, so, even though the weighted MAPE of this experiment is lower than the weighted MAPE for APO, a proper comparison needs to be performed before concluding that this experiment is better than APO.

Regarding XGB experiment 5, the MAPE of each of the demand groups for this experiment can be seen in Table XIII. It can be seen that in this case the MAPE for demand groups 2 and 3 is a little worse than in experiment 3, however, the MAPE for demand group 1 is better than before, which is good because this is the group with the largest amount of spare parts in it. In fact, the MAPE for demand group 1 is the best one obtained so far.

TABLE XII
MAPE OF EACH OF THE DEMAND GROUPS FOR XGB EXPERIMENT 5
COMPUTING IT WITH THE 6 FORECASTS OBTAINED DURING THE TESTING
PROCESS.

Demand_class		
1 - Low Low	53.919718	
2 - Low Medium	44.325150	
3 - Low High	38.933919	
Name: MAPE_XGB,	dtype: float64	

The bias analysis of this experiment showed that the amount of spare parts with a positive and with a negative bias is very similar, although it seems there are more spare parts with a negative bias. This is the only negative aspect of this model, since the MAPE obtained are very good and it seems this a very promising experiment.

In this case the overall weighted MAPE of this experiment is 39.95%, which is a little higher than before, but, it is important to remember that this MAPE is not calculated with the same amount of spare parts as in experiment 3, so the results can change when comparing the results with the spare parts they both have in common.

Regarding the LSTM and the hybrid models, the results obtained were very discouraging, especially for the LSTM models. The two experiments tried with the deep learning model yielded an overall MAPE of over 90% in both cases, showing that, either the model was not defined correctly or that it does not work well for this project. The hybrid model's results were also a little disappointing, although it makes sense to obtain bias results from this model since it is using the LSTM model.

B. XGB Experiment 3 vs XGB Experiment 5

As stated before, XGB experiment 3 and XGB experiment 5 are the most promising experiments among all the different models that have been tried. As a reminder, experiment 3 included all the spare parts from plant 5758 and market FD and experiment 5 included all the spare parts belonging to low price and low demand groups. This means that this comparison includes only spare parts belonging to low price and low demand groups and from plant 5758 and market FD.

For this comparison, the MAPE of each experiment was computed again, but this time only with the spare parts they have in common. The results obtained can be seen in Table XIV and Table XV.

TABLE XIV
MAPE OF EACH OF THE DEMAND GROUPS FOR XGB EXPERIMENT 5
COMPUTING IT WITH THE 6 FORECASTS OBTAINED DURING THE TESTING
PROCESS.

Model	Weighted MAPE
XGB experiment 3	40.62
XGB experiment 5	38.45

Looking at the overall MAPE it can be seen that experiment 5 seems to have a better performance than experiment 3, having a lower MAPE for low price and low demand spare parts from plant 5758 and market FD. Additionally, after comparing both method's MAPE for each spare part and choosing the model with the lowest MAPE it was seen that, for 65% of the spare parts, experiment's 5 performance is equal or better than experiment's 3 performance.

TABLE XV
MAPE OF EACH OF THE DEMAND GROUPS FOR XGB EXPERIMENT 5
COMPUTING IT WITH THE 6 FORECASTS OBTAINED DURING THE TESTING
PROCESS.

Zfuso_Plant_Location	Zfuso_Market	Best_Method	Best_Method_cnt	
5758	FD	Experiment 3	4101	33.83
		Experiment 5	5331	43.98
		equal	2690	22.19
Name: Best_Method_cnt_percentage, dtype: float64				

The only aspect in which experiment 3 seems to be better than experiment 5 is in the bias, since experiment 5 has more spare parts with a negative bias, however, since this problem is easy to fix and can be fixed in the future, experiment 5 was chosen as the best model and it was the one used for the comparison against the current methods.

C. XGB Experiment 5 vs APO

The comparison between XGB experiment 5 and APO included the spare parts belonging to a low price and a low demand group. For this comparison, only 6 time periods were used for computing the MAPE since XGB experiment 5 only has a forecast for the last 6 time periods of data available.

The results obtained for the overall MAPE of each model can be seen in Table XVI. It is visible that the MAPE obtained for experiment 5 is approximately 1.5% better than the MAPE for APO, showing that this experiment performs better on low price and low demand groups.

TABLE XVI
MAPE FOR XGB EXPERIMENT 5 AND APO COMPUTING IT WITH THE 6 LAST
TIME PERIODS AND ONLY WITH SPARE PARTS FROM LOW PRICE AND LOW
DEMAND GROUPS.

Model	Weighted MAPE
XGB experiment 5	40.33
APO	41.77

Additionally, the MAPE of each method for each combination of plant and market can be seen in Table XII. This results show how the experiment 5 not only performs better in general, but also has a lower MAPE in every possible combination of plant and market.

TABLE XVII
MAPE FOR XGB EXPERIMENT 5 AND APO COMPUTING IT WITH THE 6 LAST
TIME PERIODS AND ONLY WITH SPARE PARTS FROM LOW PRICE AND LOW
DEMAND GROUPS GROUPING BY PLANT AND MARKET.

Plant	Market	Weighted MAPE	
		XGB experiment 5	APO
5756	FE	51.73	54.52
5758	FD	38.81	40.17
5758	FE	50.14	51.90

It can be seen that the XGB experiment 5 is a better model for low demand and low price spare parts. It is true that the improvement achieved can seem to be very little, since it is only a 1.5%, however, considering the amount of spare parts included in the project, which is more than 13,000 spare parts, this means that a lot of money can be saved using this new model.

Thanks to this comparison it has been discovered that the current methods implemented in the company can still be improved. Also, obtaining good results in this XGB model is very good, as this model can be also used for assessing the performance of the new SAP tool that will be implemented.

IV. CONCLUSIONS AND NEXT STEPS

During this document, the results of the current forecasting method have been presented, showing in which groups of spare parts it has a bad performance that should be improved and showing vary valuable insights about other aspects like the manual forecast, the bias or the forecast error. Additionally, the results of the new methods have also been analyzed, showing how an improvement has been achieved in the low price and low demand spare parts.

This project has shown how difficult it is to find a good forecasting model for all the spare parts, being necessary to separate the spare parts into different groups and training a different model for each group. In this project spare parts have been grouped by plant, market, demand and price, obtaining the best results when grouping by demand and price. This results showed that a better method than the APO can still be achieved by using the XGB method.

Nevertheless, the XGB method trained during this project is a simple model, since it was developed in a short period of time. This means that the results obtained from the new methods can still be improved.

One way of improving this results can be to update the data and include information from 2020, this will make the dataset bigger, having more time periods to train the model and more time period for testing it. Another way can be trying to fit a model with less spare parts, focusing only in one low demand group instead of the three that were used in experiment 5.

Other ways of improving the performance of the XGB model can be trying new ways of grouping the spare parts, like performing a dynamic time wrapping (DTW) analysis, which will group spare parts with similar demand profile shapes. Finally, the LSTM model can be fixed and trained with different datasets until some good results are achieved, and then try the hybrid model again.

REFERENCES

- [1] MFTBC, "About | Mitsubishi Fuso Bus and Truck Corporation", 2020 [Online], Available: <https://www.mitsubishi-fuso.com/corporate-profile/> [Accessed: August 14, 2020].
- [2] Pacheco Soto, Ivan. (2015). Classification and forecasting for Inventory Management of Spare Parts.
- [3] Frepple. (). Demand classification: Why forecastability matters. Available: <https://frepple.com/blog/demand-classification/> [Accessed: April 27, 2020].
- [4] N. Tyagi, "Introduction to time series analysis: Time-series forecasting machine learning methods & models", 2020. Available: <https://medium.com/analytics-steps/introduction-to-time-series-analysis-time-series-forecasting-machine-learning-methods-models-ecaa76a7b0e3> [Accessed: 12 May, 2020]
- [5] S.-I. L. Hugh Chen Scott Lundberg, "Hybrid gradient boosting trees and neural net-works for forecasting operating room data", Research paper, 2017.
- [6] R. Agarwal, "Using gradient boosting for time series prediction tasks", 2019. Available: <https://towardsdatascience.com/using-gradient-boosting-for-time-series-prediction-tasks-600fac66a5fc> [Accessed: June 5, 2020].