УДК 519.6

ROBUST AND SCALABLE PROCUREMENT FORECAST IN LOGISTICS

Ulyanov Sergey V. [1], Filipyev Andrey V.[2]

[1]*Grand PhD in Physical and Mathematical Sciences, professor;*
Dubna State University;
19 Universitetskaya Str., Dubna, Moscow region, 141980, Russia;
Chief Researcher;
Joint Institute for Nuclear Research;
6 Joliot-Curie Str., Dubna, Moscow region, 141980, Russia;
e-mail: ulyanovsv@mail.ru.

[2]*PhD student;*
Dubna State University;
19 Universitetskaya Str., Dubna, Moscow region, 141980, Russia;
e-mail: avfilipev@gmail.com.

This article aims to reveal that machine learning algorithms can be applied in a variety of commercial companies in order to improve developing intelligent systems. The major task which would be discussed in the developing robust forecast model and lauching it on scalable data platform. Besides algorithms and software, the problems of changing processes would be considered and possible solutions suggested.

Keywords: Machine Learning, Intelligence Logistic Systems, Scalable Data Platform Architecure, Decision Making Systems, Artificial Intelligence.

For citation:
Ulyanov S. V., Filipyev A. V. Robust and scalable procurement forecast in logistics. *System analysis in science and education*, 2022;(3):1-7. Available from: http: //sanse.ru/download/472.

НАДЕЖНЫЙ И МАСШТАБИРУЕМЫЙ ПРОГНОЗ ЗАКУПОК В ЛОГИСТИКЕ

Ульянов Сергей Викторович[1], Филипьев Андрей Владимирович[2]

[1]*Доктор физико-математических наук, профессор;*
Государственный университет «Дубна»;
141980, Московская обл., г. Дубна, ул. Университетская, 19;
Главный научный сотрудник;
Объединенный институт ядерных исследований,
141980, Московская обл., г. Дубна, ул. Жолио-Кюри, д. 6;
e-mail: ulyanovsv46_46@mail.ru.

[2]*Аспирант;*
Государственный университет «Дубна»;
141980, Московская обл., г. Дубна, ул. Университетская, 19;
e-mail: avfilipev@gmail.com.

Статья призвана показать, что алгоритмы машинного обучения могут применяться в различных коммерческих компаниях для улучшения разработки интеллектуальных систем. Основная задача – разработка надежной модели прогноза и ее запуске на масштабируемой платформе данных. Помимо алгоритмов и программного обеспечения, будут рассмотрены проблемы изменения процессов и предложены возможные решения.

Ключевые слова: машинное обучение, интеллектуальные логистические системы, архитектура масштабируемой платформы данных, системы принятия решений, искусственный интеллект.

Для цитирования:

Ульянов С. В., Филипьев А. В. Надежный и масштабируемый прогноз закупок в логистике // Системный анализ в науке и образовании: сетевое научное издание. 2022. №3. С. 1-7. На англ. языке. URL:http://sanse.ru/download/472.

1. Introduction

Forecasting demand for fast-growing businesses is a key factor for successful development in highly competitive markets. There are lots of reasons to understand future sales beginning from satisfying customers desire of purchasing certain products affects loyalty up to optimizing stocks in stores. Every type of market has its specificities of business processes and it depends on various parameters like the sales of food products with limited expiring periods or electrical goods with the expensive cost of supplying and storage. The market has evolved into a “pull” environment with customers more demanding and discriminating, dictating to the supplier what products they desire and when they need them delivered [1].

This article aims to reveal the possible baseline approach for organizing demand forecasting for a wide chain of stores to standardize procurement processes and provide a unified tool for the store managers for ordering necessary product ingredients for coming periods. The main purpose was to predict the consumption of ingredients for every unit of sales because the stores’ chain has different parameters of demand, supply period and transport for each unit of sales. The responsibility for ordering ingredients rests on the store managers and each of them has their own approach for forecasting demand on ordinary days, holidays, local events, etc. Thus, our goal was not only providing the prediction with higher accuracy than manual managers forecast but also scaling robust results and developing infrastructure for future machine learning model improvements based on the additional parameters of promotions, marketing activities, geolocation, purchasing power.

The reproducible and stable forecast metric results do not guarantee the success of the project. There are multiple algorithms for generating time series prediction but the first goal was to give business value for the company in such spheres as decreasing loss because of lack of opportunity of sailing products because of ingredients absence. To solve this problem, we had to develop a forecasting model and check the possibility of integration provided prediction into the operating processes of stores and the scaling of this algorithm for every pair of store and ingredient. This means that provided predictions have to be reliable for forecasting thousands of time series which should be applicable in lots of stores.

Dodo Pizza operates in a very competitive market. Only after a few years of development, it became a leader in the local market and opened its stores in more than thirteen other countries. The issue of forecasting ingredients consumption solves not only the problem of reducing the overstock and out of stock in the point of sales, but also could provide a united prediction of necessary products for the logistic company and food producers for a lingering period. It helps to optimize food production and scale beforehand.

2. Method

The providing of a robust forecast of ingredients does not guarantee the absolute success of a project. The first steps were about manual providing forecast results through sending predicted numbers by machine learning model to stores managers and tests revealed that it is very important to standardize business processes inside stores before developing a module of automated procurement.

2.1. Process standardization

Manual tests have shown that the company has to change the processes of predicting necessary ingredients and ordering them through the whole way from producers to the sale stores.

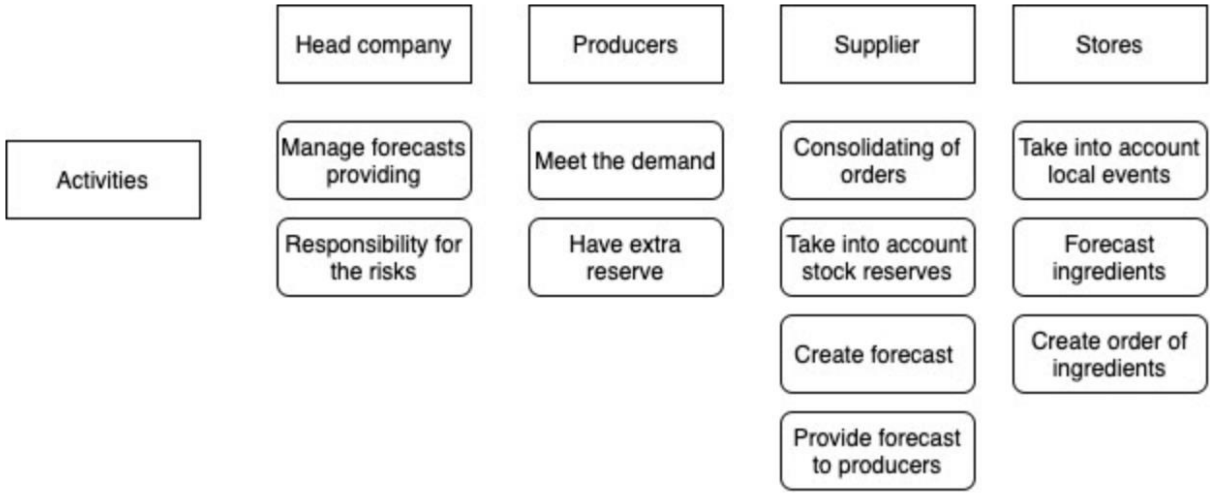

Fig. 1. Main responsibilities of supply chain units

Figure 1 illustrates the main units of the supply chain and their responsibilities. As we can see there is no one responsible unit for predicting and it affects the whole process. The main problem is that every store has its approach to plan future orders based on its understanding of the market. Supplier is responsible for consolidating all store forecasts and create their forecast with added reserves for unforeseen circumstances. Producers have to understand that in the case of unexpected sales increasing they have to have an extra reserve for ingredients, it can happen when the head company launches new products with strong promoting.

Improving the planning process via increasing accuracy of predicting for our company was possible and easier if the first step would be creating one point of responsibility inside the head office. This requests investment in developing a new approach of forecasting based on machine learning algorithms and specific IT infrastructure for creating and deploying forecasts every day for every sale store.

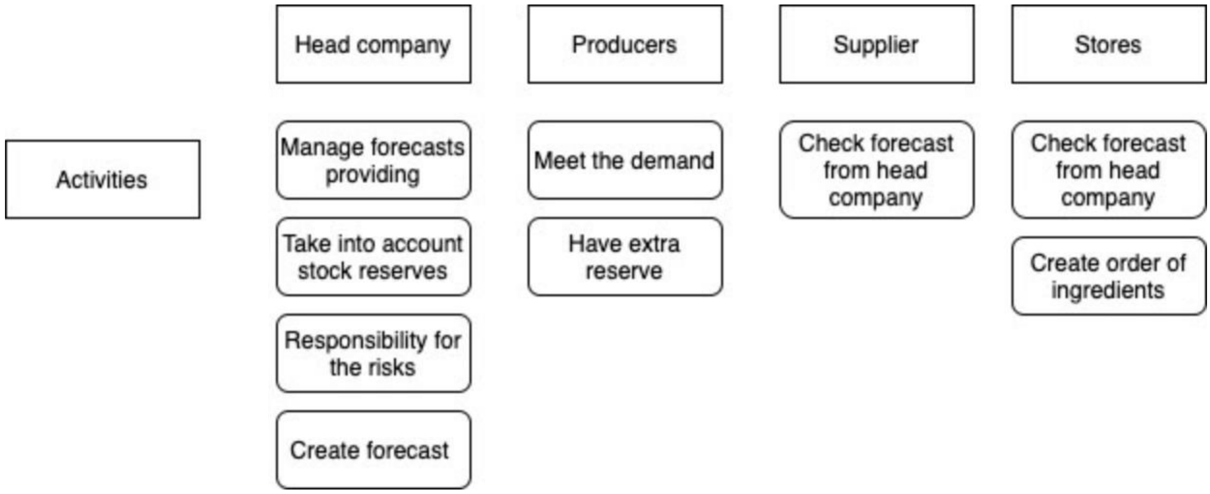

Fig. 2. Changed scheme of responsibilities

As we can see in Figure 2 the main idea is to move responsibility to the head company. It helps to concentrate on improving the accuracy of forecast in one place and have an understandable process of including all items which affect orders inside the prediction processes like regional and local actions, marketing, launching new products, creating a reserve and necessary time for supplying ingredients to different regions.

2.2. Forecasting approaches

The purpose of time series forecasting is to predict some future actions in order to make decisions based on understanding that some value will be equal to some number with certain accuracy. Today there are various approaches for solving the predicting problem and in this article, we compare the ARIMA algorithm and Facebook Prophet Library which is based on the Holt-Winter method.

2.2.3. ARIMA

This approach is widely used in solving time series forecasting including building models for supply chains[2]. Our assumption is based on the idea that if demand data is stationary it can be represented by autoregressive integrated moving average model.

The ARIMA is generally denoted as ARIMA(p,d,q), where parameters are non-negative numbers:

- *p* is autoregressive terms;
- *d* is differences;
- *q* is moving averages.

This kind of model assumes that that Y_t is a linear function of the preceding values and is given by equation (1):

$$Y_t = \alpha_t Y_{t-1} + \varepsilon_t \qquad (1)$$

We can describe this formula as each observation consists of a random component and a linear combination of the previous observations, and α_1 in this equation is the self-regression coefficient.

The behavior of the time series may be affected by the cumulative effect of some processes. For example, stock status is constantly modified by consumption and supply, but the average level of stocks is essentially dependent on the cumulative effect of the instantaneous changes over the period between inventories. Although short-term stock values may fluctuate with large contingencies around this average value, the level of the series over the long term will remain unchanged. A time series determined by the cumulative effect of activity belongs to the class of integrated processes. Even if the behavior of a series is erratic, the differences from one observation to the next can be relatively low or even oscillate around a constant value for a process observed at different time intervals. This stationarity of the series of differences for an integrated process is a crucial characteristic viewed from the statistical analysis side of the time series. Integrated processes are the archetype of nonstationary series. A differentiation of order 1 assumes that the difference between two successive values of Y is constant. An integrated process is defined by equation (2)[3]:

$$Y_t = Y_{t-1} + \varepsilon_t \qquad (2)$$

where the random disturbance $\varepsilon_t$is white noise.

The moving averaging process is a linear combination of the current white noise with one or more previous perturbations. The moving average order indicates the number of previous periods embedded in the current state. Thus, a moving average is defined by equation (3):

$$Y_t = \varepsilon_t - \theta\varepsilon_{t-1} \qquad (3)$$

Box and Jenkins[4] method was founded to develop a practical approach in order to perform ARIMA models. The Box–Jenkins principle consists of three iterative steps of model identification, parameter estimation, and diagnostic checking steps[5]. The main rule to identify the model is that if a time series is obtained from an ARIMA process, it should have some theoretical autocorrelation properties. By matching the theoretical and empirical autocorrelation patterns, we make it possible to identify one or several potential models for the given time series. Box and Jenkins principle proposed to use the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the sample data as the basic tools to identify the order of the ARIMA model.

2.2.3 Facebook prophet

Facebook Prophet is an open-source library for generating time series that can be used for forecasting in a various tasks. It gives fast setting up fitting model and robust result that independent on a lot of external facts. Importantly, it is also designed to have intuitive parameters that can be adjusted without knowing the details of the underlying model[6]. It gives data analysts an oppurtinity for effective model tuning and fast building solutions that can answer their hypothesis questions.

The core is based on decomposable time series model with three main components: trend, seasonality and holidays. Following formula decribes the combination of these parameters:

$$f(x) = g(t) + s(t) + h(t) + \epsilon_t \qquad (4)$$

Here g(t) is the trend function which models non-periodic changes in the value of the time series, s(t) represents periodic changes (e.g., weekly and yearly seasonality), and h(t) represents the effects of holidays which occur on potentially irregular schedules over one or more days. The error term t represents any idiosyncratic changes which are not accommodated by the model.

This specification is similar to a generalized additive model (GAM) that was described by Hastie & Tibshirani[7], a class of regression models with potentially non-linear smoothers applied to the regressors. Here time as a regressor is used but possibly several linear and non-linear functions of time as components. Modeling seasonality as an additive component is the same approach taken by exponential smoothing[8]. Multiplicative seasonality, where the seasonal effect is a factor that multiplies g(t), can be accomplished through a log transform.

The generalized additive model decomposes easily and accommodates new components as necessary, for instance when a new source of seasonality is identified.

Developers of this approach, in effect, framing the forecasting problem as a curve-fitting exercise, which is inherently different from time series models that explicitly account for the temporal dependence structure in the data. This formulation provides a number of practical advantages in comparing with ARIMA:

1. *Flexibility* - the model accommodates seasonality with multiple periods and lets the analyst make different assumptions about trends;
2. Unlike with ARIMA models, the measurements do not need to be regularly spaced, and analyst does not need to interpolate missing values by removing outliers;
3. The forecasting model has easily interpretable parameters that can be changed by the analyst to impose assumptions on the forecast. Moreover, analysts typically do have experience with regression and are easily able to extend the model to include new components.

2.3 Developing scalable forecast service

The main idea of developing scalable service was in building data platform which can be used in various machine learning projects inside our company. We had to provide data analyst tools that could speed up creating solutions without wasting time on gathering and preprocessing data. Also, these set of tools has to be independent on a data size because the story of our company keeps information about customers, sales and connected data for about 10 years.

2.3.3 Data Platform Architecture

The main information system that automates all business processes inside every our store is based on cloud technologies. It has lots of advantages to use managed services and one of them is the how fast we can launch new solutions and support them.

In order to run our procurement forecast model we decided to start developing own data platform and build trastable data source that can collect data from various databased of our distributed information system. The conception of using Delta Lakes[9] as a main storage gave us some important advantages that we discuss later.

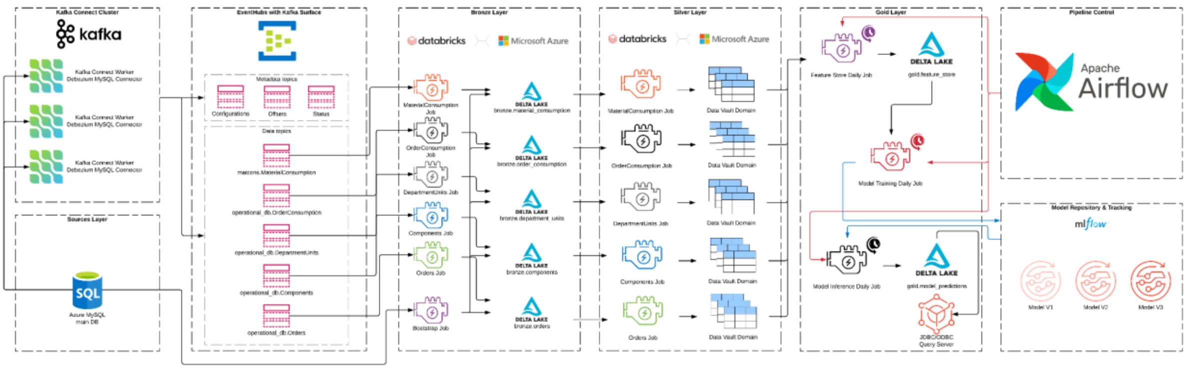

Fig. 3. Data Platform Architecture

This technical architecture contains the idea of gathering data and keeping it on three different layers:

- *Bronze Layer* - keeps data as it is in the data source, also called *Raw Data*;
- *Silver Layer* - the first step after preprocessing *Raw Data* and changing it to the Data Vault[10] analytical architecture if it is neccessary for the final business task;
- *Gold Layer* - the place where machine learning models can provide their inferences into the production systems.

This Data Platform artchitecure let us provide infrastructure with fast accessable and trustable information.

Over every layer we used Apache Spark which can work with data in a distributed and robust way. This set of tools is wide useable in the subject of developing high load distributed machine learning solutions[11]. We connected data platform with Apache Spark and our forecast model and built robust and scalable procurement solution.

3. Results

Our developed model that was lauchned on data platfotm and changed business procceses let us scaled this approach to 165 stores and provides procurement forecasts for moreover 20 000 time series everyday.

Table 1. Forecast result statistics

Check date	**Unique time series**	**MAE**	**MAPE**	**SMAPE**
2021-09-25	**8776**	6,30	49,85%	21,43%
2021-11-01	8180	6,43	55,25%	22,68%
2021-12-14	21089	6,29	51,24%	21,87%
2021-12-28	21735	6,65	54,53%	22,66%
2022-01-11	21960	5,52	47,25%	20,52%
Mean		6,24	51,62%	21,83%

We can coclude that our approach can provide stable solution and it reasonable to develop this model further.

4. Conclusion

It is reasonable to say that we achieved success in the first stage of developing our procurement forecsat system. However, it is neccessary to pay attention to the fact that not only the forecast model gives us an advantages of using machine learning algorithms in logistins. There some of things which are strongly connected inside this approach.

Firstly, if the developed model has to help a great amount of people inside the company there is the high accuracy that you will have to change the processes inside the company. One of the insights of McKinsey researches proves that embedding analytics is as much about change management as it is about data science. It suggests that not only upgrading software is critically necessary, but a strong change management strategy is needed [12].

Finally, if your company works on high compatitive markets it is worth to find a way of developing data platform and processes around in order to support further developing of data based projects. Every machine learning projects costs a fortune, but in our hands make this developing faster, cheaper and build strong data strategy for the company.

References

1. Wisner J. D., Tan K. C, Leong G. K. Principles of supply chain management: a balanced approach. Mason: Thomson South-Western, 2011.
2. Kenneth G. An ARIMA Supply Chain Model. Management Science. 2005. Vol.51, No 2. Pp.305-310. https://pubsonline.informs.org/doi/pdf/10.1287/mnsc.1040.0308.
3. Forecasting of demand using ARIMA model. International Journal of Engineering Business Management. Vol. 10. Pp. 1–9. https://journals.sagepub.com/doi/10.1177/1847979018808673.
4. Box George E. P, Jenkins G. Time series analysis, forecasting and control. San Francisco: Holden-Day, 1970.
5. Hanke J. E., Reitsch A. G. Business forecasting. 5th ed. Englewood Cliffs. 1995.
6. Taylor S. J., Letham B. Forecasting at scale. 2017. PeerJ Preprints 5:e3190v2. https://doi.org/10.7287/peerj.preprints.3190v2.
7. Hastie T., Tibshirani R. Generalized additive models: some applications. Journal of the American Statistical Association. 1987. Vol. 82, No 398. Pp. 371–386.
8. Gardner E. S. Exponential smoothing: the state of the art. Journal of Forecasting. 1985. Vol. 4, Pp. 1–28.
9. Armbrust M., et al. Delta lake: high-performance ACID table storage over cloud object stores. Proceedings of the VLDB Endowment Vol. 13, No 12. 2020. Pp. 3411-3424. DOI: 10.14778/3415478.3415560
10. Priebe T., Neumaier S., Markus S. Finding Your Way Through the Jungle of Big Data Architectures. 2021 IEEE International Conference on Big Data (Big Data). IEEE, 2021. DOI: 10.1109/BigData52589.2021.9671862.
11. Khan Mohammad Arafat Ali, et al. Nowcasting the Financial Time Series with Streaming Data Analytics under Apache Spark.2022. arXiv preprint arXiv:2202.11820.
12. Jit Kee Chin, Mikael Hagstroem, Ari Libarikian, and Khaled Rifai. Advanced analytics: Nine insights from the C-suite. McKinsey & Company. 2018. https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/ advanced-analytics-nine-insights-from-the-c-suite.