ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)

Volume no 11 Issue 1 January 2024

Opportunities of Using Artificial Intelligence Algorithms to Improve the Sourcing Process

[1] Lackes, Richard, [2] Sengewald, Julian

[1] [2] Dept. of Business Informatics, TU Dortmund, Dortmund, Germany
Email ID: [1] richard.lackes@tu-dortmund.de, [2] julian.sengewald@tu-dortmund.de

Abstract*— The article examines applications of AI algorithms in the field of procurement and sourcing. Based on a common project with a German enterprise the use of machine learning methods to determine the price of purchased parts and components will be presented and discussed. The project implementation is described in detail using the CRISP procedure model. The price estimates of the AI model can be used in particular for new products. Although there is not any quotation for a new product we can estimate how much should it cost at the procurement market by our machine learning model. When price negotiations with suppliers begin, the responsible purchaser has received helpful information on the expected or target price and can therefore better evaluate any offers from the potential supplier. An innovative two-stage technique with Clustering and Artificial Neural Nets showed the best results. The basis for this approach is a large data collection of previously supplied components with their specifications and their procurement prices. Additionally to the price forecast one get the most important price and delivery cost driver so it gives some hints for reducing the delivery price or to control the negotiation process.

Keywords: Artificial Intelligence and sourcing, product price, procurement cost data, price forecast, neural networks, machine learning algorithms.

I. INTRODUCTION

One important function in industrial enterprises is sourcing and procurement. It has the task to supply the enterprise with the resources it needs to provide its production processes. Its objective is to ensure timely and cost-effective supply. The resources to be procured are primarily materials, preliminary products, auxiliary products, but also machines, tools and services. The resource "staff" is usually the responsibility of the HR department and therefore not part of procurement department. The procurement processes can be differentiated into strategic and operational procurement. While the operational area includes the actual implementation of procurement with order monitoring and goods delivery, and thus the day-to-day business, strategic procurement provides the necessary framework conditions. This includes the long-term sourcing strategy and the determination of requirements, as well as the search for and selection of suppliers and the negotiation of conditions and prices with suppliers.

In this strategic area in particular, new approaches to planning support with the help of Artificial Intelligence (AI) algorithms can open up potential for improvement. Over the years, ERP systems (e.g. SAP S/4 HANA) have collected a large amount of purchasing and procurement-relevant data (e.g. supplier offers, prices, delivery conditions, procurement market developments etc.) that can be meaningfully evaluated using machine learning methods. A PwC study from 2022 shows the progress of digitalization and thus also the possibility of using AI processes for data analysis in procurement (PricewaterhouseCoopers 2022). According to the study, the digitization rate will be 72% by 2025. In addition, the 800 companies participating in the study intend to spend an average of EUR 1.28 million annually on the digitalization of procurement processes. In addition to cost reduction and digital transformation, the strategic priorities of the procurement department are risk management, supplier procurement and supplier selection (PricewaterhouseCoopers 2022).

Some of the potentials of using AI methods are outlined below. Using the example of the provision of adequate price data by AI methods, the selection of suppliers and the assessment of supplier quotations will be significantly improved. Based on a practical project with a company in the electrical industry, the price assessment for new products using machine learning methods will be described and discussed in detail.

II. LITERATURE REVIEW AND AI POTENTIALS

AI potentials in strategic procurement and sourcing were primarily examined with regard to rationalization, demand forecasting and supplier evaluation and selection.

Cost reduction potential in procurement and sourcing can be identified by analyzing historical order and consumption data. For example, (Tako and Robinson 2012) used AI methods to uncover inefficiencies in the stock management. The bullwhip effect is the result of such inefficiencies. Using historical data, it could be identified and avoided in the future with suitable behavioral rules - generated by AI models. The bullwhip effect in particular also highlights deficits in risk and information management that could be avoided through better digitalization. In the area of digital transformation, software agents can be used to open up new purchasing sources or systematically evaluate tendering and auction platforms. Electronic product catalogs and purchasing platforms open up further potential for rationalization.

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)

Volume no 11 Issue 1 January 2024

Dedicated forecasting models such as those described in (van Steenbergen and Mes, 2020) can be used to determine the quantitative purchasing volumes. They use various machine learning methods and combine K-means, random forest and quantile regression forest methods. Product characteristics of existing and new products as well as historical demand data of existing products are used to make predictions about overall demand (van Steenbergen and Mes 2020). The K-means algorithm is used to group the demand patterns into profiles. Random forest algorithms are then used to forecast the profiles, total demand and distributions for new products. This supports inventory management and reduces the risks associated with introducing new products. AI models for forecasting sporadic demand can be found in (Nikolopoulos et al. 2016).

The selection of suppliers is one of the most important tasks of a company's procurement department (Mohammed et al. 2018). The classification of suppliers can be implemented using powerful criteria through a k-means or self-organizing map method as well as neural networks (Segura and Maroto 2017; Tavana et al. 2016).

In supplier search and selection, AI models can also be used to provide data that improve both the negotiation process itself and the (potential) further search for suppliers. In particular, the negotiation process can be significantly improved by providing condition-relevant data such as price information and delivery times. This means that the AI models should provide the buyer with price data in advance, which, according to the AI models, are "fair" or expected product prices. Equipped with this important information, he can then better enter into price negotiations with a supplier. Furthermore, supplier offers can be better assessed. In the case of price quotations for a product that is to be newly procured, the usual procedure lacks a reference value that can be used to assess whether the quotation is good or bad. The procedure to demand several quotations (e.g. 3) doesn't solve the problem because it implies to choose only the relatively best quotation without ensuring that all the available quotations are not possibly too expensive.

However, the AI could - even for completely new products - determine "typical" prices to be expected by scouring previous deliveries and quotations for similar products (Ćwikła et al. 2020) and evaluating the price data there by an AI model. These price data corresponds with the cost data that the enterprise will be faced when the delivery will start. Previous research focused mostly on (artificial) neural networks or linear regression (Bode 2000; Bodendorf et al. 2021; Cavalieri et al. 2004; Sonmez 2011) while some authors have started benchmarking these methods against gradient boosting trees or support vector regression (Loyer et al. 2016).

Another approach was to estimate delivery price resp. cost data based on its similarity to other products (Mousavi et al. 2015; Ben-Arieh 2000). Our findings are based on a similar approach but is distinguished by the machine learning model and our two-stage architecture to get the price forecast.

III. AI BUSINESS PROJECT

A. Business and Data Understanding

We investigated a case study with a procurement dataset from a German company that manufactures home and electrical products that we accessed during an industry project. Our study's focus is specifically on (electrical) resistor delivery price estimation. For this purpose, we got access to the procurement database of the ERP-system. In figure 1 the data model for our project is depicted.

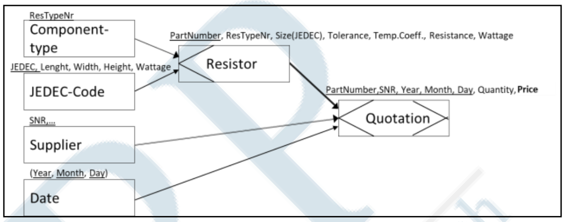

Figure 1. Sourcing Data Base (excerpt of ERP)

Suited SQL statements selected the records for resistor quotations with their attributes. Summarized we got a data table with about 180.000 rows. Each row describes the quotation of a part number at a special date with its price data and attributes respectively its features that are shown in Table 1.

Because we wanted to perform our price estimate before the procurement market analysis, we did not integrate the supplier data and only used the technical characteristics of the resistors. Resistors are standardized products that can be characterized by their technical specifications like (nominal) resistance, wattage, etc. The ResType characterizes different materials. The attribute Size (JEDEC) references only to a special code in the JEDEC table that contains data like length, width, height, and wattage of electronic components. JEDEC is an abbreviation for Joint Electronic Tube Engineering Council and is the global leader in developing open standards for the microelectronics industry. For the later used mining table, we substituted the size resp. JEDEC-code by its technical content features. Higher electrical powers also cause the resistor material to heat up and heat causes the material's electrical resistance to decrease. The temperature coefficient (TempCoeff) is used to quantify this reduction. Resistors can deviate from the nominal resistance value as a result of variations made during the manufacturing process. The attribute tolerance of a resistor is used to measure these variances. The tolerance and temperature coefficient can both be thought of as quality indicators.

B. Data Preparation and Feature Engineering

The records within this data set belong to the procurement activities of the domestic appliance manufacturing company and reflect the real usage of the procurement software database. To determine the features of our mining table, we first had to cleanse the database. For that we verify if there are only records in the database that accurately reflect the real-world entities and identify entries that are not consistent, with the latter definition: Empty records, zero prices and duplicates were removed, as a result of our verification procedure.

IFERP

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)

Volume no 11 Issue 1 January 2024

Table 1. Overview of Features (m = metric, c = categorical)

Name	Val	Description
Part.Number	ID	Unique ID.
Date	ID	The date on which the quotation was recorded, used as an additional ID.
Price	m	Price quotation at a date and current prices.
Resistance	m	Nominal resistance in $k\Omega$.
Wattage	m	How much electrical power a resistor tolerates.
Size(JEDEC)	c	JEDEC-code with a reference to the data in the JEDEC table (length, width, height, wattage).
ResType	c	Material and construction type of resistors (e.g. Carbon-Composit, Cermet, wire-wound).
Temp.Coeff	c	Indicator of deviation from nominal resistance when exposed to temperature changes.
Tolerance	c	Deviation from nominal resistance under real circumstances and load measured as a percentage.

The label of a data set is the attribute that should be learned by machine learning algorithms and is of major importance for the business utility of the entire machine learning solution. Because we aim to estimate the price of a component (identified by the part number), we choose the attribute price, as indicated on the quotation of a supplier, as our label. Several problems associated with this had to be solved. At first, you can maintain that there were about 180.000 data records representing different component quotations and only about 2000 different part numbers. On average about 90 different price data for one part number. Which one should be selected?

Given that the quotations were obtained between 2014 and 2018, the second challenge was determining

how to make the price data independent of inflation and technical progress. Therefore, one needs to analyze the historical development of the price over time and modify the respective price data of a quotation to make them comparable. All price data was adjusted to the base level of 2019. For the transformation, we used a producer price index for electronic and optical products issued by Eurostat to account for price developments specific to that sector. Therefore, this price index reflects the cost that is typically relevant for procurement departments. That way, we can see the "true" level of prices. After that price adaptation, we aggregate our dataset (in SQL grouped by part number and without the date attribute) and decide to use the mean price of each number as our label. For the estimation of procurement price risk, the standard deviation was also recorded.

Regarding feature engineering, we choose to add features that affect the price of a resistor (table II). By the JEDEC data, we included the dimensions width, length and height and derived the corresponding volume. In addition, after discussing with engineers, an important observation was that the surface area on a circuit board is more limited than the volume.

Therefore, we include the new feature area, the occupied space on a circuit board. To reflect the performance of a resistor relative to its dimensions, we calculate the ratio of resistance and wattage compared to the surface area of a resistor (Resistance_area, Wattage_area). Similarly, we proceeded with volume and its derived attributes (Resistance_volume, Wattage_volume). Area efficiency and volume compactness indicator were also constructed. The mining table for the machine learning algorithms contains the identifier part number, also five attributes from Table I (excluding Size(JEDEC), Date, Price) and eight additional features from Table II and the new label Price_adj.

2456 -1304

ISSN (Online)

International Journal of Science, Engineering and Management (IJSEM)

Volume no 11 Issue 1 January 2024

Table 2. Overview of Derived Features (m=metric, C=Categorical)

Name	Val	Description
Price_adj	m	Price adjusted for inflation in 2019 prices according to the ECB's producer price index of electronic components published.
Area	m	$Width \times Length$
Area efficiency	m	$\dfrac{area}{(\frac{1}{2}(Width + Length))^2}$
Compactness	m	$\dfrac{volume}{(\frac{1}{3}(Width + Length + Height))^3}$
Resistance_area	m	Performance indicator that relates the resistance to the area of a resistor.
Resistance_volume	m	Performance indicator that relates resistance to the volume of a resistor.
Volume	m	$Width \times Length \times Height$
Wattage_area	m	Performance indicator that relates wattage to the area of the resistor.
Wattage_volume	m	Performance indicator that relates wattage to the volume of the resistor.

C. Data Modeling

To get a first impression concerning the relationship between the chosen features and the price data a correlogram can be used. High correlations are indicated by a more intense coloring.

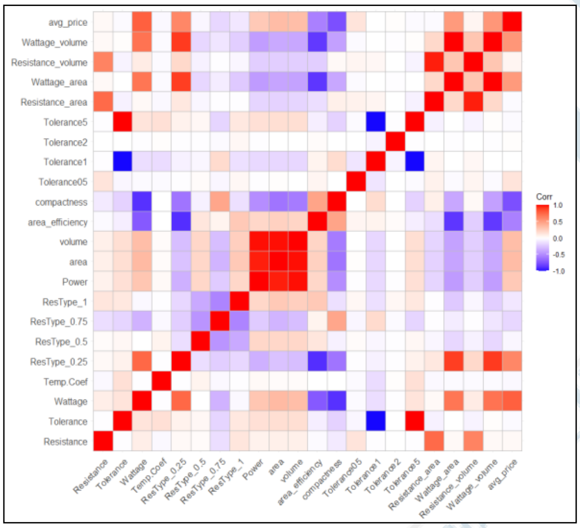

Figure 2: Correlogram of features and label

For example, wattage, ResType=0.25, width and compactness appear to be the most important predictor variables. The first two have a price-increasing effect, the last a price-reducing effect.

We decide to use a neural network (NN) model to solve our price/cost estimation problem as an instance of the class of nonlinear regression problems. For prediction tasks, NN is a popular and often quite performant method. In our case, we employed a feedforward NN with 30 input neurons and one output neuron. Category features were replaced by suitably coded dummy variables (one-hot encoding). Furthermore, we normalize the input data to handle the very different data domains of the features.

As a normalization method, we tried both min-max-normalization and z-transformation. The z-transformation calculates each value by referencing the mean value and standard deviation. It indicates the multiplier to the standard deviation that corresponds to the difference between the initial value and the mean value.

To get the right parameter setting, we execute hyperparameter-tuning by experimenting with different network topologies. Various numbers of hidden layers and different sizes of the hidden layers were simulated. According to our analysis, none of these topologies is superior to the X-X-1 structure (X=30). To evaluate the predictive model, a 70:30 split was performed, resulting in a training data set and a test data set. The performance indicator is the standard deviation of the average price.

D. Model Performance and Results

After predictive validation, the mean scaled average error was 16.9% (z-normalization) and 17.6% (min-max-normalization). In comparison with a "naive" estimation (the overall average price) that results in an average error of approximately 44% the NN model has a considerably better performance. In the following, we concentrate on the better results of the z-normalization procedure.

Nevertheless, we wanted to improve this result. The basic idea was that an NN for all components might perform less well than a NN that differentiates the components according to their specifications in subgroups. Then, for each subgroup of parts, a dedicated NN can be modeled. To implement this

ISSN (Online) 2456 -1304

International Journal of Science, Engineering and Management (IJSEM)

Volume no 11 Issue 1 January 2024

idea, we used a two-stage clustering before prediction.

In the first step, the k-means algorithm is used to define subgroups. In our case, k = 3 was the best choice. Remarkably, the largest group has on average the lowest price level (2.24 $), the second largest group has the medium-price level (4.08 $) and the smallest group has the highest price level (6.53 $). The standard deviation is quite similar in all groups. In the second step, a distinct NN was trained for each cluster. Using these approaches, we obtain NN1, NN2, and NN3 using the same network structure, but different weights, to assess the price of their constituent parts.

During the test and deployment process, a new component whose price should be predicted must first be allocated to the right subgroup. This can be done by calculating the distance to the cluster centroids based on their technical characteristics and assigning the component to its nearest cluster. According to this decision, the cluster-specific NN can be used to make a price prediction.

Using this two-stage procedure we obtained the performance data shown in Figure 3. The red line indicates the performance in the one-stage case. Surprisingly, not all results can be improved by subgrouping. For components in the class of mid and low-range prices, the prediction has deteriorated slightly (around 3 to 4 percentage points).

On the other hand, it is possible to predict high-priced components with great accuracy. Only a 4% deviation to the right value is an excellent result. Because these parts are the most expensive ones, they determine the final product costs to a large extent. And therefore, it is decisive to get a good estimation of their price data. The lower-priced parts have a minor influence on total procurement costs and therefore although their percentage deviation has worsened this has only minor implications. Measured in absolute values, even larger percentage deviations for the low-priced item yield only small dollar amounts.

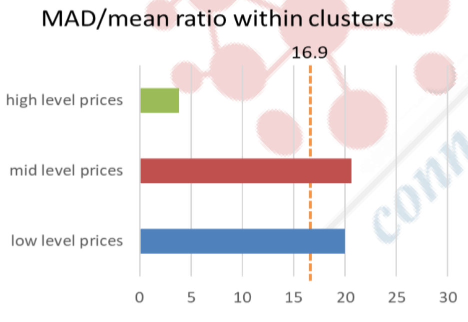

Figure. 3: K-means clustering with k=3 and z-Norm.

To get more information about the price drivers of electronic components one can investigate which features are important. By that, the design decision concerning alternative components can be considerably facilitated. Therefore, we analyzed the feature importance separately. Several features have an importance value greater than 0.15. Especially the wattage, the temperature coefficient, and the wattage-area are important for the prediction.

Moving on to the two-stage case we can detect a different view. In all clusters, only a few features have an importance value above 0.15. In cluster one with the high-price components, only two features are above 0.15. The temperature coefficient, representing the needed quality, has by far the greatest influence, followed by Wattage_area with just over 0.15. The last feature represents the power per area unit.

In the second cluster with the medium-price components we got a similar situation whereby here the temperature coefficient as quality indicator completely dominates (the only feature above 0.15). With a value of 0.225, it is more than twice as high as the second most important parameter.

The last cluster with the lowest price components gives a completely different picture. Only the feature Resistance_volume gets an importance value above 0.15. It characterizes the resistance value (measured in ohm) per volume unit. Quality indicators are of minor importance.

IV. DISCUSSION

A. Managerial Implications

We discover that the neural network's predictive performance is quite strong, particularly when utilizing the z-transformation. The average prediction error has a standard deviation between 15 and 20%, which corresponds to the typical price premium. An intriguing finding of the investigation is that embracing variability by forming subgroups can boost performance. This discovery is interesting because most of the previous research studied an alternative machine learning approach, but we employ neural networks that are specialized for specific types of resistors. Therefore, we demonstrate that two-stage models are more accurate predictors than one-stage models. Specifically, the two-stage approach provides a solid foundation for an accurate price estimate of new goods that can be used during the negatiations with a potential supplier or to judge a quotation.

Regarding the assessment of features, it is possible to determine if some specifications are really necessary (i.e. some quality requirements) because they can be an important price driver. As demonstrated, the temperature coefficient, as a quality characteristic, has the greatest impact on pricing.

Concerning the search for suited suppliers one can use the estimated price data for a new component as a kind of benchmark to compare it with the quotation of found supplier. If it will be judged as to high the search process can be continued. In the other case one can stop the searching process and choose the best supplier.

B. Future Research

This study lays the foundation for future research into the potential of machine learning to support the price negatiations for the sourcing department. From a technical standpoint, future research on applied machine learning should examine not only the price level but also the price deviation. Predicting the price deviation could be accomplished by the same method as predicting the average price, with the added advantage of obtaining a range of probable price data, which is a valuable indicator of procurement price risk. In addition, additional research might be conducted on the benefits of utilizing additional market data (i.e., procurement market characteristics, supplier data) to improve the estimates' precision. From a decision science and business intelligence perspective, it is promising that machine learning techniques can be used in supply chain operations to create awareness regarding uncertainties in the product price.

REFERENCES

[1] Ben-Arieh, D., Cost estimation system for machined parts, International Journal of Production Research, vol. 38, no. 17, pp. 4481–4494, 2000, doi: 10.1080/00207540050205244.

[2] Bode, J., Neural networks for cost estimation: Simulations and pilot application, International Journal of Production Research, vol. 38, no. 6, pp. 1231–1254, 2000, doi: 10.1080/002075400188825.

[3] Bodendorf, F., Merbele, S. and Franke, J., Deep learning based cost estimation of circuit boards: a case study in the automotive industry, International Journal of Production Research, pp. 1–22, 2021, doi: 10.1080/00207543.2021.1998698.

[4] Bodendorf, F., Lutz, M., Michelberger, S. and Franke, J., An empirical investigation into intelligent cost analysis in purchasing, Supply Chain Management: An International Journal, 2021, doi: 10.1108/SCM- 11-2020-0563.

[5] Brinke, E.T., Lutters, E., Streppel, T. and Kals, H. "Variant-based cost estimation based on Information Management," International Journal of Production Research, vol. 38, no. 17, pp. 4467–4479, 2000, doi: 10.1080/00207540050205235.

[6] Cavalieri, S., Maccarrone, P. and Pinto, R., Parametric vs. neural network models for the estimation of production costs: A case study in the automotive industry, International Journal of Production Economics, vol. 91, no. 2, pp. 165–177, 2004.

[7] Ćwikła, G., Grabowik, C., Bańczyk, K. and Wiecha, Ł., Assessment of Similarity of Elements as a Basis for Production Costs Estimation, in 14th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2019), Cham, 2020, pp. 386–395.

[8] Loyer, J.-L., Henriques, E., Fontul, M. and Wiseall, S., Comparison of Machine Learning methods applied to the estimation of manufacturing cost of jet engine components, International Journal of Production Economics, vol. 178, pp. 109–119, 2016.

[9] Mohammed, A.; Setchi, R.; Filip, M.; Harris, I.; Li, X.: An integrated methodology for a sustainable two-stage supplier selection and order allocation problem. In: Journal of Cleaner Production 192, 2018, S. 99–114. DOI: 10.1016/j.jclepro.2018.04.131.

[10] Mousavi, S. M., Vahdani, B. and Abdollahzade, M., An intelligent model for cost prediction in new product development projects, IFS, vol. 29, no. 5, pp. 2047–2057, 2015, doi: 10.3233/IFS-151682.

[11] Nikolopoulos, Konstantinos I.; Babai, M. Zied; Bozos, Konstantinos: Forecasting supply chain sporadic demand with nearest neighbor approaches. In: International Journal of Production Economics 177, 2016, S. 139–148. DOI: 10.1016/j.ijpe.2016.04.013.

[12] PricewaterhouseCoopers: Digital Procurement Survey 2022. Online verfügbar unter https://www.pwc.de/de/strategie-organisation-prozesse-systeme/operations/digital-procurement-survey.html, zuletzt aktualisiert am 09.04.2022, zuletzt geprüft am 09.04.2022.

[13] Segura, Marina; Maroto, Concepción: A multiple criteria supplier segmentation using outranking and value function methods. In: Expert Systems with Applications 69, 2017, S. 87–100. DOI: 10.1016/j.eswa.2016.10.031.

[14] Sonmez, R., Range estimation of construction costs using neural networks with bootstrap prediction intervals, Expert Systems with Applications, vol. 38, no. 8, pp. 9913–9917, 2011, doi: 10.1016/j.eswa.2011.02.042.

[15] Tako, Antuela A.; Robinson, Stewart: The application of discrete event simulation and system dynamics in the logistics and supply chain context. In: Decision Support Systems 52 (4), 2012, S. 802–815. DOI: 10.1016/j.dss.2011.11.015.

[16] Tavana, Madjid; Fallahpour, Alireza; Di Caprio, Debora; Santos-Arteaga, Francisco J.: A hybrid intelligent fuzzy predictive model with simulation for supplier evaluation and selection. In: Expert Systems with Applications 61, 2016, S. 129–144. DOI: 10.1016/j.eswa.2016.05.027.

[17] van Steenbergen, R. M.; Mes, M.R.K.: Forecasting demand profiles of new products. In: Decision Support Systems 139, 2020, S. 113401. DOI: 10.1016/j.dss.2020.11