Towards Predicting Supplier Resilience: A Tree-Based Model Approach

Maximilian Enthoven
University of St. Gallen
maximilian.enthoven@unisg.ch

Ivo Blohm
University of St. Gallen
ivo.blohm@unisg.ch

Erik Hofmann
University of St. Gallen
erik.hofmann@unisg.ch

Philipp Gordetzki
University of St. Gallen
philipp.gordetzki@unisg.ch

Abstract

With looming uncertainties and disruptions in today's global supply chains, such as lockdown measures to contain COVID-19, supply chain resilience has gained considerable attention recently. While decision-makers in procurement have emphasized the importance of traditional risk assessment, its shortcomings can be complemented by resilience. However, while most resilience studies are too qualitative in nature and abstract to inform supplier decisions, many quantitative resilience studies frequently rely on complex and impractical operations research models fed with simulated supplier data. Thus there is the need for an integrative, intermediate way for the practical and automated prediction of resilience with real-world data. We therefore propose a random forest-based supervised learning method to predict supplier resilience, outperforming the current human benchmark evaluation by 139 percent. The model is trained on both internal ERP data and publicly available secondary data to help assess suppliers in a pre-screening step, before deciding which supplier to select for a specific product. The results of this study are to be integrated into a software tool developed for measuring and tracking the total cost of supply chain resilience from the perspective of purchasing decisions.

1. Introduction

Today's global supply chains are subject to increasingly turbulent environments [1]. Their inherently complex and interdependent nature makes them highly vulnerable to unexpected disruptions. For example, sudden events with high economic impact such as Hurricane Katrina in 2005 [2], the Japanese tsunami in 2011 [3], and most recently the COVID-19 crisis in 2020 and 2021 [4] caused widespread production halts and revenue losses.

In order to mitigate the extent of disturbance induced by such events, numerous companies have implemented approaches for observing and managing supply and production risks, such as comprehensive risk assessments and insurance plans [5]. However, these approaches have limitations. For example, the statistical assumptions in risk assessments quickly reach their limits when considering low-probability events [6]. Moreover, risk management faces challenges on how to deal with ontological uncertainties (”unknown unknowns”) and increasing interdependencies in a globalized world [7]. Finally, risk management is often ”based on the notion of stability as the 'normal' state of affairs” [8].

Resilience aims to complement risk management by focusing not only on mitigating risks, but also on building capabilities through targeted investments [9]. Furthermore, it aids in addressing those disruptive events, whose probability models are prone to error and emphasizes the recovery after an unpredictable disruption has occurred [10]. As resilience has become a central topic for procurement and supply management, decision-makers are increasingly looking for ways to integrate it into their supplier decisions [11]. Research currently offers two directions for assessing supplier resilience: On the one hand, most qualitative studies require deep internal knowledge about the suppliers [12] and lack performance metrics [13]. These often include extensive surveys and interviews susceptible to high levels of subjectivity [3]. On the other hand, quantitative studies frequently rely on complex and impractical operations models fed with simulated supplier data [14, 15, 16], and thus have difficulties projecting reality. These studies consider disruption events and the behavior of suppliers during simulated disruption events (”reactive resilience”). Thus, there is a gap between qualitative and quantitative research for quick and accurate supplier resilience assessment, in which resilience is considered as part of daily business (”proactive resilience”).

As ISO 9001 requires annual supplier audits, most companies currently evaluate their suppliers with high manual effort in semi-annual or annual

URI: https://hdl.handle.net/10125/79540
978-0-9981331-5-7

periods retrospectively [17]. Meanwhile, procurement professionals have an abundance of data at their fingertips: in-house ERP systems containing supplier information, historical orders, and delivery data as well as a plethora of external data sources regarding (disruption) events. However, they are often overwhelmed on how to employ this data to foster adequate decision-making. Thus, they are in need of a practical and fast solution that guides them on how resilient a supplier is and will be in the future. Machine learning approaches can tackle this challenge and detect previously unknown relationships in large data sets well [18]. We therefore conclude that the intelligent use of this data for the practical operationalization of resilience to inform supplier decisions is a relevant topic of investigation, resulting in the following research question:

RQ: How can the resilience of a supplier be accurately predicted with the help of primary ERP and publicly available secondary data?

From a scientific perspective, this study aims to contribute to closing the gap between qualitative and quantitative resilience research and simplify resilience operationalization and prediction. From a practical perspective, the study aims to make use of easily accessible data to empower procurement professionals with an automated supplier resilience assessment solution to intelligently and proactively inform supplier decisions.

In the following sections, related work to resilience and especially its operationalization is presented. Afterwards, the methodology including data preparation and model development are described. The results are presented, benchmarked against the current approach, and discussed, before we summarize and conclude our study with a future outlook.

2. Related Work

2.1. Supply Risks and Uncertainty

Supply chain risks and uncertainty can cause events that disrupt the flow of material and goods in the supply chain [19, 20]. While some use both concepts interchangeably, a clear distinction is needed [9].

Supply chain risks have been studied since 1980 to assess the expected costs for a supply chain failure [20]. Through risk assessment, events such as data loss or natural disasters are associated with a concrete probability of occurring and the expected economic impact [21, 22]. To calculate the probability distribution of a risk, historical data or expert knowledge can be used [23, 24].

Uncertainty, on the flip side, entails unpredictable events such as a pandemic or a tsunami [25], that have not been encountered yet or that fall outside past experience [9]. Traditional risk assessment approaches have difficulties portraying this uncertainty, as they lack historical information and thus cannot predict the impact of control actions [26]. Events like the Japanese tsunami, with high impact and low probability (HI/LP), are driven by uncertainty and require decision-makers to design supply chains that withstand unplanned disruptions [9]. To complement risk assessment and enable handling of HI/LP-events, companies can leverage solutions to build and track resilience [6].

2.2. Resilience

The term resilience can be found in many different disciplines (e.g., psychology and biology), however the focus of this study lies on resilience of companies within their supply chains. One of the most accepted definitions is by Fiksel (2006), who defines resilience as “the capacity of an enterprise to survive, adapt and grow in the face of turbulent change” [8]. Complementary to traditional risk management [6], resilience allows companies to cope with uncertain environments filled with HI/LP events [10]. Resilience may be perceived from two perspectives: proactive and reactive. Proactive resilience follows a business continuity narrative [3] and aims to keep company performance as high as possible at all times, agnostic to specific disruption events [27]. In contrast, reactive resilience focuses on disaster management by suggesting case-by-case strategies for quickly and effectively recovering after disruptions [28]. While reactive resilience is more idiosyncratic towards individual cases, it is not suited for a usage on daily basis [29].

However, resilience is still an abstract term difficult to grasp to many researchers and practitioners alike [13]. Thus, in order to operationalize it, authors have proposed single metrics [30, 31, 32, 33], compared multiple metrics [4, 34, 35] and created indices to aggregate various correlated metrics [3, 14, 36, 37, 38].

In general, resilience metrics in literature are often based on company performance in terms of economic output. Performance metrics take into account disruption severity and recovery time [36] and measure the impact of disruptions on a company’s performance [10, 33]. Furthermore, authors have proposed various measurements for performance-based resilience, like lead time deviation [30, 32], on-time delivery [4, 16, 33, 39], sales volume [40] or expected availability [31]. As decision-makers are interested not only in resilience but also in the economic impact of their actions, many models additionally include cost metrics [30, 31, 32, 33].

Aside single metric approaches, authors also built models with multiple values to measure resilience. For example, Ivanov (2020) predicted the impacts of epidemic outbreaks and compared production-inventory dynamics, estimated lead time service level, financial performance and lead-time performance [4]. Analyzing eleven indicators, Cardoso et al. (2015) suggest network design indicators for active planning of resilient supply chain and centralization indicators for reporting the dynamics of the network [34]. Three robustness measure and one recoverability measure are proposed by Li et al. (2020), who argue that decision-makers should focus on key characteristics of the supply chain [35].

Composite indices based on linear models are suited for supply chain resilience, as long as performance indicators are without synergy and conflict effects [36]. The SCRAM-Tool developed by Pettit et al. (2013) can be used to measure the supply chain resilience of a company, aggregating the resilience gaps proposed by Pettit et al. (2010) [3, 6]. Using a Bayesian network approach, Hosseini & Ivanov (2019) combine vulnerability and recoverability of companies in a supply network to a resilience index with ripple effect considerations [14].

According to Sheffi & Rice (2005), resilience is measured by the performance of suppliers [10]. It is along with quality and delivery history the most important factor for supplier selection [41, 42]. While currently multiple indices exist for measuring supplier resilience, most companies do not make use of them, as they are impractical in terms of length and input effort or extremely case-specific. The go-to solution in companies nowadays is a subjective supplier evaluation, where soft factors are intuitively rated on a scale and averaged to achieve a score. These solutions often lack speed, objectivity, and accuracy for tracking suppliers and further decision-making. In the literature, we have not found a speedy and accurate supplier resilience measurement solution based on real-world data that is simple to use and requires little manual input. This study focuses on a practical, rapid, and automated approach to operationalize resilience for procurement professionals beyond subjective supplier evaluation. Therefore, this study measures resilience through the supplier service indicated by on-time deliveries, as proposed by Carvalho et al. (2011) and Cavalcante et al. (2019) [16, 33]. However, their models lack real-world application by using simulated supplier data based on assumptions and statistical distributions.

3. Methodology

In order to address the research question, the study uses a supervised machine learning approach divided in three phases (Figure 1). The first phase consists of data collection and exploratory analysis of company-internal supplier data and complementary secondary data, aggregated to a feature set of independent variables. From the order history, a ternary supplier resilience metric is calculated, which constitutes the dependent variable. Following the data collection phase, the second phase includes the development, training and application of a multi-variable supervised machine learning model, aiming to predict the resilience of the current and new suppliers. The final phase is dedicated to compete the performance of our model against the current subjective supplier resilience evaluation.

The data for this study is provided from a key partner within a research consortium from the Swiss manufacturing industry. With a revenue of 150 million Swiss Francs, the company is one of the global leaders for switch gear development as well as control technology for electrical infrastructure products. Their purchasing volume of 90 million Swiss Francs is designated for product groups including raw materials (i.e., copper and aluminum) as well as electrical components. In general, the company has not been strongly affected from the COVID-19 crisis and is seeking for a simple solution for daily use. A profile of the case study company is presented in Table 1.

Revenue (2020)	CHF 150 million
Purchasing volume	CHF 90 million
Number of suppliers	436
% suppliers in Switzerland	66%
% suppliers in Europe (no CH)	32%
% suppliers rest	1%

Table 1. Company profile

While 66 percent of their suppliers are located in Switzerland, 32 percent are located in Europe (excl. CH) and 1 percent in the rest of the world.

3.1. Preparing the data

In the first phase, we analyzed the supplier data and order history data from our partner company. The order data included a list of all orders made by the purchasing department over the past 3 years. It contained 27’144 historical entries from 436 unique suppliers.

In a next step, we systematically selected resilience elements from literature, which could be represented by quantitative metrics. To achieve generalization and application of our model beyond our partner, we agreed

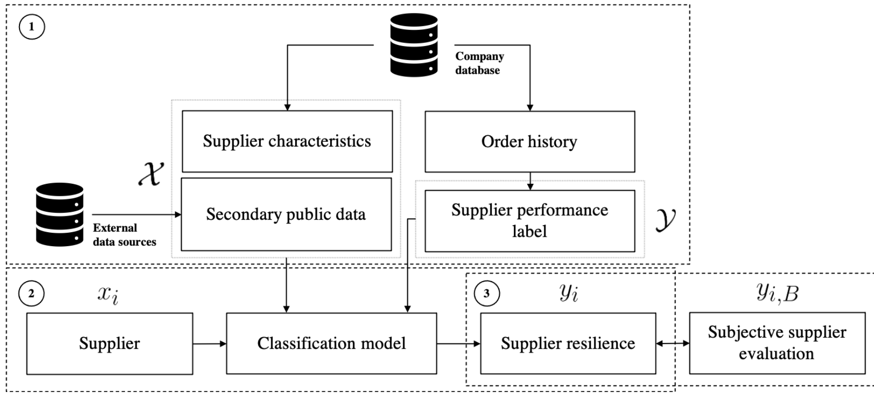

Figure 1. Proposed method pipeline for predicting supplier resilience.

on three kinds of data within the research consortium: Data already stored in internal ERP systems by all the consortium members (e.g., revenue of the supplier, use of multiple sourcing), data that can automatically be fetched from external data sources (e.g., currency volatility and country risk data, such as the one year volatility [43] and the World Risk Index [44], as well as subjective evaluations on the supplier, obtained during negotiation and first contact. The responsible procurement employees rated the reachability [6, 45, 46, 47, 48], friendliness [6, 48], price in relation to market price [6, 45], the political stability [6] and flexibility [6, 45, 46, 47, 49] of the supplier on a four-point Likert scale. This allows the model to profit from human expertise and obtain company specific information.

Our aim was to build a set of independent features:

$$\mathcal{X} = \{x_1, ..., x_n\},\ x_i \in \mathbb{R}^m, \tag{1}$$

where n is the number of suppliers and m is the number of features. After merging both the supplier data set inferred from the order history and the external data sources, a final data set could be prepared. We extracted the procurement volume of a supplier [6], if the delivered product is success-critical [6, 45, 46, 47], if multiple sourcing for the supplier's product is available [6, 45, 46, 47], the supplier's revenue [6], and the language of correspondence with the supplier [6, 45, 47]. Additionally, the location of the supplier and the used currency allowed us to derive the world risk index [6] and the currency volatility [6] through automated access to external data sources. Lastly, the company evaluated the 98 suppliers with highest revenue to the above listed subjective features. All these features map resilience elements mentioned in literature, as illustrated in Table 2.

3.2. Preparing the resilience metric

Let the dependent target variable set be

$$\mathcal{Y} = \{y_1, ..., y_n\},\ y_i \in \{0, 1, 2\}, \tag{2}$$

where 0 stands for ”not resilient”, 1 stands for ”partially resilient”, and 2 stands for ”resilient”. Following a similar approach as [50, 51], we compare the results to a human benchmark in the final phase and therefore choose a discrete ternary classification set.

While recent studies suggest to measure resilience solely according to the service level and fault quantity [4, 16, 33], we propose the additional penalization of delays by the severity. High delays of success-critical products have led to production stops in the past and are used to assess supplier reliability by numerous studies [4, 16, 33, 39]. Thus, from the order history, we extracted the service level in terms of planned and actual delivery dates per order, such that each for each supplier $i \in \{1, ..., n\}$ the following sequence can be defined:

$$\mathcal{S}_i = \{\sigma_1^{(i)}, ..., \sigma_{p(i)}^{(i)}\} \in \mathbb{Z}, \tag{3}$$

where $\sigma_j^{(i)}$ is the deviation in days from the agreed delivery date of the kth delivery from supplier i and $p(i)$ is the total number of deliveries from a supplier i. $\sigma_j^{(i)} > 0$ implies a late delivery, whereas $\sigma_j^{(i)} < 0$ implies an early delivery. We define as the subset of $\mathcal{S}_i$ the subset of all non-zero elements:

$$\mathcal{S}_i^{\pm} = \{\sigma_j^{(i)} \in \mathcal{S}_i \mid \forall j \in \{1, ..., p(i)\},\ \sigma_j^{(i)} \neq 0\}. \tag{4}$$

The service level of a supplier i can then be calculated as:

$$\alpha_i = 1 - \frac{\|\mathcal{S}_i^{\pm}\|}{\|\mathcal{S}_i\|} \tag{5}$$

Feature	Number set	Resilience elements
Procurement volume	$\mathbb{Z}^+$	Complexity, Capacity [6]
Success-critical product	$\mathbb{B}$	Collaboration [6, 45, 46], Security [6, 45], Integration [47]
Multiple sourcing	$\mathbb{B}$	Flexibility in sourcing [6, 46] Adaptability [6], Robustness, Redundancy [45, 47]
Revenue	$\mathbb{R}^+$	Financial strength [6]
Currency volatility	$\mathbb{R}_0^+$	External pressures, Sensitivity [6]
Political stability	$\{0, 1, 2, 3\}$	Turbulence, Security [6]
Reachability	$\{0, 1, 2, 3\}$	Anticipation [6], Visibility [6, 45, 46, 47, 48], Information sharing [46, 47]
Friendliness	$\{0, 1, 2, 3\}$	Organization [6], Trust [48]
Market price	$\{0, 1, 2, 3\}$	Market position [6, 45], External pressures [6]
German correspondence	$\mathbb{B}$	Collaboration [6, 45, 47]
Flexibility	$\{0, 1, 2, 3\}$	Flexibility in order fulfillment, Dispersion [6], Agility [45, 46, 47, 49]
World Risk Index	$[0, 100]$	Deliberate threats [6]

Table 2. Overview of features of the data set $\mathcal{X}$ and associated resilience elements.

Being an element in $[0, 1]$, it includes the quotient between the cardinality of all non-zero elements and the cardinality of the total set. This definition of service level is currently widely used [52]. Therefore a high number of non-zero elements is penalized, resulting in low α_i, an indication of a low service level performance. In order to take into account the extent of delay, we penalize the service level with the help of the deviation from the agreed delivery date in days, following Kamalahmadi & Parast (2015) [53]. Therefore, comparing two suppliers with the same service level but with different average deviations, the one with the higher deviation is penalized more heavily. We define the delay penalty of a supplier i as follows:

$$\beta_i = 1 - \frac{\sum_{z \in \mathcal{S}_i^{\pm}} |z| - \min_i \sum_{z \in \mathcal{S}_i^{\pm}} |z|}{\max_i \sum_{z \in \mathcal{S}_i^{\pm}} |z| - \min_i \sum_{z \in \mathcal{S}_i^{\pm}} |z|} \quad (6)$$

It sums the absolute number of deviation days and normalizes this to a scale $[0, 1]$. Note that both positive (late) and negative (early) deviations are penalized. While late deliveries may result in production halts and customer dissatisfaction, early deliveries may pose challenges in terms of storage, shelf life, and early payment requests. We then multiply the service level with the delay penalty to arrive at the continuous target variable:

$$\tilde{y}_i = \alpha_i \beta_i \in [0, 1]. \quad (7)$$

The company provided us with a supplier evaluation list, in which the responsible purchasers have rated the most important existing suppliers with values in the set $\mathcal{Y}_B = \{0, 1, 2\}$. The evaluation is done yearly, and is based on subjective evaluation and heuristics. Afterwards, the team takes action on poorly rated supplier to improve their score in the future. This evaluation is neither based on hard performance metrics nor on scientific methods, but rather on ”gut feeling” of decision-makers. In order to make our target variable y_i directly comparable to the human baseline benchmark, which includes a ternary subjective evaluation of each supplier, we transform the continuous variable $\tilde{y}_i$ into a discrete one by introducing the thresholds t_0 and t_1:

$$y_i = \begin{cases} 0 & \text{if } \tilde{y}_i < t_0 \\ 1 & \text{if } t_0 \leq \tilde{y}_i < t_1 \\ 2 & \text{if } t_1 \leq \tilde{y}_i \end{cases} \quad (8)$$

Internal thresholds that allow variable discretization are often used (e.g., in credit scoring systems) to enable easy interpretation of results and justify specific decisions [54]. The thresholds t_0 and t_1 were chosen to match the distribution of the subjective evaluation and confirmed subsequently in an interview with the company. This hints to the fact that anyone with a penalized service level of around 93 percent is highly reliable, whereas the range from 89 percent to 93 percent is fairly reliable, and anything under that is considered ”bad performance”.

Due to lacking order history or erroneous order entries, y_i was only available for 75 samples. The remaining 23 labels were then generated through Multivariate Imputation through Chained Equations (MICE), the most common and generalizable imputation method to make up for missing data [55]. In Figure 2, the distribution of the service level α_i, the delay penalty β_i, and the target variable $\tilde{y}_i$ is demonstrated.

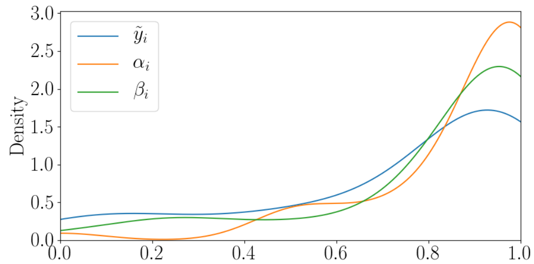

Figure 2. The distribution of the service level α_i, the delay penalty β_i, and $\tilde{y}_i$.

3.3. Developing and training the model

After developing $\mathcal{X}$ and $\mathcal{Y}$, we could define the sequence

$$\mathcal{M} = ((x_1, y_1), ..., (x_n, y_n)) \tag{9}$$

of pairs from $\mathcal{X} \times \mathcal{Y}$. Our aim was to find a function $f : \mathcal{X} \rightarrow \mathcal{Y}$, which predicts $y_i \in \mathcal{Y}$ for arbitrary $x_i \in \mathcal{X}$. The discrete target variable $\tilde{y}_i$ allowed us to formulate this as a classification problem.

The real-world data noise, the low samples-to-features ratio [56], and the presence of few categorical features led us to tree-based classification algorithms that can also be applied to small data sets. We applied the random forest (RF) algorithm, a supervised machine learning method, which consists of a combination of decision trees to provide accurate and stable predictions, as it frequently outperforms other classification methods such as logistic regression and k nearest neighbors [57]. In order to further increase the robustness of our results, RF was applied within a bootstrapping framework [58]. One big advantage of tree-based models is their high resistance against irrelevant variables, in comparison to neural networks or kernel methods [49]. During model performance analysis, we evaluated the inclusion of features by considering the variable importance, which measured how effective a feature is at reducing uncertainty in the model [59].

As the data with a sample size of $n = 98$, the distribution of labels revealed highly imbalanced data, i.e., many samples leaning toward $y_i = 0$ or $y_i = 2$. This may bias prediction results and have an effect on how well the model predicts resilience. Using Synthetic Minority Oversampling Technique (SMOTE), we balanced the label distribution of classes by generated samples for the minority classes in order to reach $n = 132$ [60].

Since this set was still rather small, results may suffer from low generalizability. To mitigate this, we applied a method known as bootstrapping, a resampling technique, where samples are drawn randomly with replacement from our data set to build a model. By repeating this step many times, the distribution of the results got closer to the real distribution [58]. A summarization of all data manipulation steps are demonstrated in Table 3. In our case, $n = 132$ instances were drawn randomly with replacement from the data set.

#	Description	n_x	n_y
1	Cleaned data set	98	75
2	MICE for missing labels	98	98
3	SMOTE for balancing classes	132	132
4	Bootstrapping during training	132	132

Table 3. Steps to augment training set.

In order to prevent the model from overfitting (i.e., an inherent lack of generalizability), regularization steps were introduced. We thus implemented a grid search on hyperparameters such as the number of trees, the maximum depth of trees, the maximum number of features, and the maximum number of samples. Moreover, during training of the model, a k-fold cross validation was performed, randomly splitting the data in training and testing sets in each iteration, over which model performance metrics were averaged.

3.4. Benchmarking the results

In the last phase, we tested whether our model beats the current supplier evaluation. We therefore directly benchmarked the ternary predictions against the subjective supplier assessment by the purchasers, in terms of accuracy, precision, and recall.

4. Results

We performed a total of 500 bootstrapping iterations, each time drawing 132 samples with replacement from the generated data set and performing a 5-fold cross validation on the data set. Thereby, in each iteration, we split the samples in five sets. The model is trained on four sets (80 percent) and tested on the remaining data (20 percent). Each bootstrapping iteration, consists of five training and testing phases, so that each of the five sets is used for testing once. The mean scores over all bootstrapping iterations of the five cross validation steps were stored and afterwards evaluated.

First, in order to measure the predictive performance of our approach, we calculated several metrics on the testing data sets. For investigating overall model performance, we first calculated the accuracy of our predictions (the percentage of suppliers that were correctly classified as ”resilient”, ”partially resilient” or ”not resilient”). We also calculated the precision, recall, and the AUC (”area under the curve”). The overall accuracy of our prediction models is 79 percent. However, high precision values (83 percent), high recall (70 percent) and high AUC (82 percent) are more important for evaluating the model.

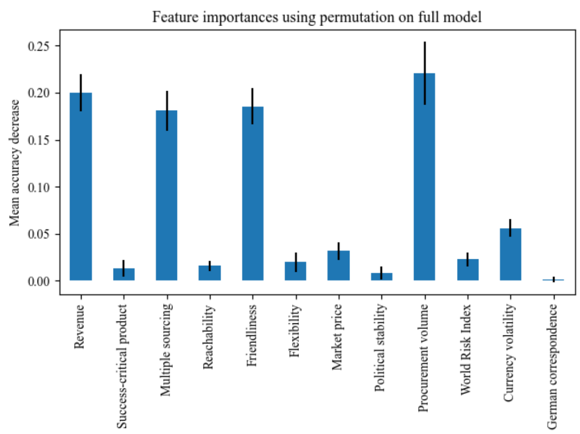

Figure 3. Variable importance of each feature.

Second, to get a more detailed understanding of the role of individual variables in predicting supplier resilience, we investigated the importance of each variable for the prediction model (see Figure 3). The term “variable importance” reflects the relative contribution of each measure to the overall prediction model. Thus, the higher this value, the higher the variable’s importance for making accurate predictions. Figure 3 shows that the variable measuring the procurement volume and revenue of a supplier is the most important one for predicting supplier resilience. At the same time, multiple sourcing and friendliness rank similarly high. Political stability, currency volatility, and the World Risk Index are of relatively lower importance. We elaborate the theoretical implications of these results in the discussion section.

Last, we benchmarked our results against the subjective evaluations of purchasers. In terms of Mean Absolute Percentage Error, our model achieved a 139 percent performance improvement over human baseline when considering the service level with an accuracy of 79 percent vs. 33 percent, precision of 83 percent vs. 29 percent and recall of 70 percent vs. 29 percent.

5. Discussion

Corroborative with [4, 16, 33, 39], we estimated the service level for suppliers as a direct indicator for resilience, as it is key in managing both uncertainties and unexpected risks. While [61] solely used service level as the resilience metric, we extended this definition to include the quality of service by including β_i [53]. As managers require performance metrics to monitor supply chain operations [62] and to understand the resilience of increasing complex supply networks [31], we suggest a single metric instead of a metric set. This is confirmed by the company, as, in their words, ”a KPI overload would lead quickly to overwhelming purchasers”. Furthermore, as mentioned by [63], on-time delivery is a suitable KPI to measure and track security, flexibility, knowledge management, and collaboration, which are key resilience factors [3, 45, 47]. This implicates that the proposed value is of high relevance in terms of operationalizing resilience.

As demonstrated in Figure 3, the procurement volume and revenue were the most decisive factors in the model. As hard financial metrics, these determine the size of a supplier and the dependence of the company toward this supplier, respectively. The financial dependency on a supplier thus plays a decisive role in predicting their resilience [6]. In agreement to the findings of [6, 45, 46, 47], single/multiple sourcing plays a significant role in estimating the resilience of the supplier. Most subjective factors (i.e., reachability, flexibility, market price, political stability) have scored low importance, apart from friendliness. Thus we assume that friendliness – an indicator for trust [48] and organization [6] – is the factor, which can be best estimated subjectively on a phone call or during a meeting. Surprising to us was that flexibility played a relatively low role. This may stem from the human bias in the decision, as for many the score did not reflect how suppliers reacted in case of a plan change, but rather how flexible they ”seemed”.

In general, our model recommends a higher consideration of the set of features with high importance (i.e., revenue, multiple sourcing, friendliness, and procurement volume). The less important factors might be more relevant when we consider reactive resilience in the disaster management. Since our data contains relatively few disruptions, as the company has not been strongly affected from the COVID-19 crisis, our model seems to be more suitable for proactive, daily use. We can thereby categorize the current study into the proactive resilience research stream. Further efforts may be made to extend the model for reactive disaster management cases, especially for those companies that have been more strongly affected.

Furthermore, country-specific features including the political stability, currency volatility, and the World Risk Index played a minor role toward resilience in our model, which may be due to the high percentage of suppliers in Switzerland, a highly politically stable country. Low-cost country sourcing would offer a different perspective [42].

Applying machine learning algorithms on real-world data – often incomplete and underrepresented – inherently comes with challenges. We have therefore introduced pre-processing two steps to augment and improve the data set. MICE is a method to impute incomplete data by Fully Conditional Specification (FCS) [55]. Unlike other methods such as Joint Modeling [64], FCS is the best suited method for our approach, as it does not require assumptions about the distribution of missing data, but compares missing data on a variable-by-variable basis [55].

We observed that the category of suppliers, who are ”partially resilient”, i.e., neither ”resilient” nor ”not resilient”, has potential for supplier development and should be tracked, such that they do not slip into the ”not resilient” category. As the given label distribution was inherently imbalanced, we used SMOTE to oversample this minority class [60]. Due to the lack of sample data, the undersampling of the majority classes – often used in literature [65, 66] – would not have been an alternative. Through the generation of synthetic minority data, our model can be further generalized ”to carve broader decision regions” [60].

5.1. Limitations

Naturally, the results of this study come with limitations and constraints. Many companies from the manufacturing industry currently struggle with proper digital documentation, especially in procurement, one of the major barriers to digitalization (i.e., ”garbage in garbage out”). Whereas studies like [4, 16, 33] use purely simulated data, we managed to clean and augment real-world data, increasing the practicability of the approach. When comparing the model and feature set to existing supervised machine learning approaches to predict resilience (e.g., [14, 16]), however, our model is one of the first ones that is based on real-world supplier data. As with all real-world data in the order history and the supplier evaluation may be incomplete or erroneously entered manually. Through exploratory data analysis, these pitfalls have been identified and considered, such as filtering statistically insignificant variables and cleaning erroneous entries.

Additionally, the small sample size drove us to apply resampling techniques and other statistical approaches. Using techniques like MICE, SMOTE, and bootstrapping during training allowed us to fill gaps and increase the sample size. However, these techniques may come with a few caveats, such as artificial bias.

6. Conclusion and future research

This study presents the first practical, easy-to-use, and fast supplier resilience prediction model based on real-world data in the Swiss manufacturing industry. Through data mining of primary ERP data (supplier list and order history) as well as publicly available secondary data and no further insights into the supply chain, we built and cleaned a data set estimating the resilience of each supplier through performance. We augmented and completed the data set through MICE, SMOTE, and bootstrapping, achieving a model accuracy score of 79 percent of predicting correctly, whether a supplier is resilient, partially resilient, or not resilient. This is coupled with high precision (83 percent), recall (70 percent) and high AUC (83 percent). The model outperforms the human benchmark by 139 percent. It is a bridge between quantitative and qualitative research streams promoting proactive resilience. It includes a multiplicative performance metric to penalize service level of suppliers and develop a simple model, which requires little data for relatively high performance. With the help of the model's prediction, purchasers can automatically pre-assess new suppliers in the future, when looking to source new products or switch suppliers, with previously unused ERP data.

In future research endeavors, the feature set could be augmented by the additional integration of more company- and supply-chain specific resilience elements stemming from external data sources, such as proprietary risk data, sustainability factors, and creditworthiness of suppliers. Moreover, the implementation of APIs to ERP systems would allow a real-time and seamless transfer of data to further automate the estimations.

References

[1] P. Trkman and K. McCormack, “Supply chain risk in turbulent environments—a conceptual model for managing supply chain network risk,” *International Journal of Production Economics*, vol. 119, no. 2, pp. 247–258, 2009.

[2] V. Jacob, “The economic aftermath of hurricane katrina,” *The Journal of Economic Perspectives*, vol. 22, no. 4, pp. 135–154, 2008.

[3] T. J. Pettit, K. L. Croxton, and J. Fiksel, “Ensuring supply chain resilience: Development and implementation of an assessment tool,” *Journal of Business Logistics*, vol. 34, no. 1, pp. 46–76, 2013.

[4] D. Ivanov, “Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak (covid-19/sars-cov-2)

case," *Transportation Research Part E: Logistics and Transportation Review*, vol. 136, p. 101922, 2020.

[5] J. Fiksel, M. Polyviou, K. Croxton, and T. Pettit, "From risk to resilience: Learning to deal with disruption," *MIT Sloan Management Review*, vol. 56, pp. 79–86, Dec. 2015.

[6] T. Pettit, J. Fiksel, and K. Croxton, "Ensuring supply chain resilience: Development of a conceptual framework," *Journal of Business Logistics*, vol. 31, pp. 1–21, 2010.

[7] E. P. Dalziell and S. T. McManus, "Resilience, vulnerability, and adaptive capacity: implications for system performance," *International Forum for Engineering Decision Making (IFED)*, 2004.

[8] J. Fiksel, "Sustainability and resilience: toward a systems approach," *Sustainability: Science, Practice and Policy*, vol. 2, no. 2, pp. 14–21, 2006.

[9] S. Melnyk, D. J. Closs, S. Griffis, C. Zobel, and J. Macdonald, "Understanding supply chain resilience," *Supply Chain Management Review*, vol. 18, pp. 34–41, 2014.

[10] J. Yossi Sheffi, James B. Rice, "A supply chain view of the resilient enterprise," *MIT Sloan Management Review*, vol. 47, no. 1, pp. 41–48, 2005.

[11] D. Ivanov and A. Dolgui, "A digital supply chain twin for managing the disruption risks and resilience in the era of industry 4.0," *Production Planning & Control*, vol. 32, no. 9, pp. 775–788, 2021.

[12] M. Ungar, "Qualitative contributions to resilience research," *Qualitative Social Work*, vol. 2, pp. 85–102, 03 2003.

[13] J. Pires Ribeiro and A. Barbosa-Povoa, "Supply chain resilience: Definitions and quantitative modelling approaches – a literature review," *Computers & Industrial Engineering*, vol. 115, pp. 109–122, 2018.

[14] S. Hosseini and D. Ivanov, "A new resilience measure for supply networks with the ripple effect considerations: a bayesian network approach," *Annals of Operations Research*, 2019.

[15] D. Ivanov, "Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak (covid-19/sars-cov-2) case," *Transportation Research Part E: Logistics and Transportation Review*, vol. 136, p. 101922, 2020.

[16] I. M. Cavalcante, E. M. Frazzon, F. A. Forcellini, and D. Ivanov, "A supervised machine learning approach to data-driven simulation of resilient supplier selection in digital manufacturing," *International Journal of Information Management*, vol. 49, pp. 86–97, 2019.

[17] S. Parkash and V. K. Kaushik, "Supplier performance monitoring and improvement (spmi) through sipoc analysis and pdca model to the iso 9001 qms in sports goods manufacturing industry.," *LogForum*, vol. 7, no. 4, 2011.

[18] O. M. Araz, T.-M. Choi, D. L. Olson, and F. S. Salman, "Data analytics for operational risk management," *Decision Sciences*, vol. 51, no. 6, pp. 1316–1319, 2020.

[19] J. Blackhurst, C. W. Craighead, D. Elkins, and R. B. Handfield, "An empirically derived agenda of critical research issues for managing supply-chain disruptions," *International Journal of Production Research*, vol. 43, no. 19, pp. 4067–4081, 2005.

[20] P. R. Kleindorfer and G. H. Saad, "Managing disruption risks in supply chains," *Production and Operations Management*, vol. 14, no. 1, pp. 53–68, 2005.

[21] M. Christopher and H. Peck, "Building the resilient supply chain," *The International Journal of Logistics Management*, vol. 15, no. 2, pp. 1–14, 2004.

[22] C. Harland, R. Brenchley, and H. Walker, "Risk in supply networks," *Journal of Purchasing and Supply Management*, vol. 9, no. 2, pp. 51–62, 2003.

[23] I. Manuj and J. T. Mentzer, "Global supply chan risk management," *Journal of Business Logistics*, vol. 29, no. 1, pp. 133–155, 2008.

[24] S. M. Wagner and C. Bode, "An empirical examination of supply chain performance along several dimensions of risk," *Journal of Business Logistics*, vol. 29, no. 1, pp. 307–325, 2008.

[25] E. Simangunsong, L. C. Hendry, and M. Stevenson, "Supply-chain uncertainty: a review and theoretical foundation for future research," *International Journal of Production Research*, vol. 50, no. 16, pp. 4493–4523, 2012.

[26] J. G. A. J. van der Vorst and A. J. M. Beulens, "Identifying sources of uncertainty to generate supply chain redesign strategies," *International Journal of Physical Distribution & Logistics Management*, vol. 32, no. 6, pp. 409–430, 2002.

[27] J. Wang, R. Dou, R. Muddada, and W. Zhang, "Management of a holistic supply chain network for proactive resilience: Theory and case study," *Computers & Industrial Engineering*, vol. 125, pp. 668–677, 2018.

[28] L. T. T. Dinh, H. Pasman, X. Gao, and M. S. Mannan, "Resilience engineering of industrial processes: Principles and contributing factors," *Journal of Loss Prevention in the Process Industries*, vol. 25, no. 2, pp. 233–241, 2012.

[29] D. Ivanov, "An adaptive framework for aligning (re)planning decisions on supply chain strategy, design, tactics, and operations," *International Journal of Production Research*, vol. 48, no. 13, pp. 3999–4017, 2010.

[30] T. Wu, J. Blackhurst, and P. O'grady, "Methodology for supply chain disruption analysis," *International Journal of Production Research*, vol. 45, no. 7, pp. 1665–1682, 2007.

[31] A. Zavala, D. Nowicki, and J. E. Ramirez-Marquez, "Quantitative metrics to analyze supply chain resilience and associated costs," *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, vol. 233, no. 2, pp. 186–199, 2019.

[32] H. Carvalho, A. P. Barroso, V. H. Machado, S. Azevedo, and V. Cruz-Machado, "Supply chain redesign for resilience using simulation," *Computers & Industrial Engineering*, vol. 62, no. 1, pp. 329–341, 2012.

[33] H. Carvalho, A. Barroso, V. Machado, S. Azevedo, and V. Cruz-Machado, "Supply chain resilience: A simulation study: proceedings of the international conference on instrumentation," *Measurement, Circuits and Systems, ICIMCS2011*, pp. 12–13, 2011.

[34] S. R. Cardoso, A. Paula Barbosa-Póvoa, S. Relvas, and A. Q. Novais, "Resilience metrics in the assessment of complex supply-chains performance operating under demand uncertainty," *Omega*, vol. 56, pp. 53–73, 2015.

[35] Y. Li, C. W. Zobel, O. Seref, and D. Chatfield, "Network characteristics and supply chain resilience under conditions of risk propagation," *International Journal of Production Economics*, vol. 223, p. 107529, 2020.

[36] A. Barroso, V. Machado, H. Carvalho, and V. C. Machado, "Quantifying the supply chain resilience," *Applications of Contemporary Management Approaches in Supply Chains*, pp. 13–32, 2015.

[37] A. Munoz and M. Dunbar, "On the quantification of operational supply chain resilience," *International Journal of Production Research*, vol. In Press, 2015.

[38] R. Rajesh, "A novel advanced grey incidence analysis for investigating the level of resilience in supply chains," *Annals of Operations Research*, 2020.

[39] S. M. Wagner and C. Bode, "An empirical examination of supply chain performance along several dimensions of risk," *Journal of business logistics*, vol. 29, no. 1, pp. 307–325, 2008.

[40] D. Ivanov, A. Pavlov, A. Dolgui, D. Pavlov, and B. Sokolov, "Disruption-driven supply chain (re)-planning and performance impact assessment with consideration of pro-active and recovery policies," *Transportation Research. Part E, Logistics and Transportation Review*, vol. 90, pp. 7–24, 2016.

[41] G. W. Dickson, "An analysis of vendor selection systems and decisions," *Journal of Purchasing*, vol. 2, no. 1, pp. 5–17, 1966.

[42] A. Chen, C.-Y. Hsieh, and H. M. Wee, "A resilient global supplier selection strategy—a case study of an automotive company," *The International Journal of Advanced Manufacturing Technology*, vol. 87, no. 5, pp. 1475–1490, 2016.

[43] P. Della Corte, T. Ramadorai, and L. Sarno, "Volatility risk premia and exchange rate predictability," *Journal of Financial Economics*, vol. 120, no. 1, pp. 21–40, 2016.

[44] T. Welle and J. Birkmann, "The world risk index – an approach to assess risk and vulnerability on a global scale," *Journal of Extreme Events*, vol. 02, no. 01, p. 1550003, 2015.

[45] A. Ali, A. Mahfouz, and A. Arisha, "Analysing supply chain resilience: integrating the constructs in a concept mapping framework via a systematic literature review," *Supply Chain Management*, vol. 22, no. 1, pp. 16–39, 2017.

[46] S. Azevedo, H. Carvalho, and V. Cruz-Machado, "Larg index: A benchmarking tool for improving the leanness, agility, resilience and greenness of the automotive supply chain," *Benchmarking*, vol. 23, no. 6, pp. 1472–1499, 2016. cited By 45.

[47] C. Roberta Pereira, M. Christopher, and A. Lago Da Silva, "Achieving supply chain resilience: the role of procurement," *Supply Chain Management: An International Journal*, vol. 19, no. 5/6, pp. 626–642, 2014.

[48] U. Soni, V. Jain, and S. Kumar, "Measuring supply chain resilience using a deterministic modeling approach," *Computers & Industrial Engineering*, vol. 74, pp. 11–25, 2014.

[49] M. M. Ahmed and M. Abdel-Aty, "Application of stochastic gradient boosting technique to enhance reliability of real-time risk assessment: use of automatic vehicle identification and remote traffic microwave sensor data," *Transportation research record*, vol. 2386, no. 1, pp. 26–34, 2013.

[50] H. F. O'Neil Jr, Y. Ni, A. Jacoby, and K. M. Swigger, "Human benchmarking for the evaluation of expert systems," *Echnology Assessment in Software Applications*, p. 13, 2013.

[51] W. Xiong, J. Droppo, X. Huang, F. Seide, M. Seltzer, A. Stolcke, D. Yu, and G. Zweig, "Achieving human parity in conversational speech recognition," 2017.

[52] F. T. S. Chan, "Performance measurement in a supply chain," *The International Journal of Advanced Manufacturing Technology*, vol. 21, no. 7, pp. 534–548, 2003.

[53] M. Kamalahmadi and M. Parast, "Developing a resilient supply chain through supplier flexibility and reliability assessment," *International Journal of Production Research*, pp. 1–20, 09 2015.

[54] P. Mironchyk and V. Tchistiakov, "Monotone optimal binning algorithm for credit risk modeling," tech. rep., Utrecht: Working Paper, 2017.

[55] S. v. Buuren and K. Groothuis-Oudshoorn, "mice: Multivariate imputation by chained equations in r," *Journal of statistical software*, pp. 1–68, 2010.

[56] J. Hua, Z. Xiong, J. Lowey, E. Suh, and E. R. Dougherty, "Optimal number of features as a function of sample size for various classification rules," *Bioinformatics*, vol. 21, pp. 1509–1515, 11 2004.

[57] R. Couronné, P. Probst, and A.-L. Boulesteix, "Random forest versus logistic regression: a large-scale benchmark experiment," *BMC bioinformatics*, vol. 19, no. 1, pp. 1–14, 2018.

[58] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. No. 57 in Monographs on Statistics and Applied Probability, Boca Raton, Florida, USA: Chapman & Hall/CRC, 1993.

[59] M. Loecher, "Unbiased variable importance for random forests," *Communications in Statistics - Theory and Methods*, vol. 0, no. 0, pp. 1–13, 2020.

[60] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, p. 321–357, Jun 2002.

[61] S. Hosseini, N. Morshedlou, D. Ivanov, M. D. Sarder, K. Barker, and A. A. Khaled, "Resilient supplier selection and optimal order allocation under disruption risks," *International Journal of Production Economics*, vol. 213, pp. 124–137, 2019.

[62] U. S. Bititci, A. S. Carrie, and L. McDevitt, "Integrated performance measurement systems: a development guide," *International Journal of Operations & Production Management*, vol. 17, no. 5, pp. 522–534, 1997.

[63] A. A. Karl, J. Micheluzzi, L. R. Leite, and C. R. Pereira, "Supply chain resilience and key performance indicators: a systematic literature review," *Production*, vol. 28, 2018.

[64] J. L. Schafer and R. M. Yucel, "Computational strategies for multivariate linear mixed-effects models with missing values," *Journal of Computational and Graphical Statistics*, vol. 11, no. 2, pp. 437–457, 2002.

[65] N. Japkowicz, "The class imbalance problem: Significance and strategies," in *Proc. of the Int'l Conf. on Artificial Intelligence*, vol. 56, Citeseer, 2000.

[66] F. Provost and T. Fawcett, "Robust classification for imprecise environments," *Machine learning*, vol. 42, no. 3, pp. 203–231, 2001.