

Burstein, Guy; Zuckerman, Inon

Article

Deconstructing risk factors for predicting risk assessment in supply chains using machine learning

Journal of Risk and Financial Management

Provided in Cooperation with:

MDPI – Multidisciplinary Digital Publishing Institute, Basel

Suggested Citation: Burstein, Guy; Zuckerman, Inon (2023) : Deconstructing risk factors for predicting risk assessment in supply chains using machine learning, Journal of Risk and Financial Management, ISSN 1911-8074, MDPI, Basel, Vol. 16, Iss. 2, pp. 1-16, <https://doi.org/10.3390/jrfm16020097>

This Version is available at:

<https://hdl.handle.net/10419/275165>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>

Article

Deconstructing Risk Factors for Predicting Risk Assessment in Supply Chains Using Machine Learning

Guy Burstein and Inon Zuckerman * 

Department of Industrial Engineering & Management, Ariel University, Ariel 40700, Israel

* Correspondence: inonzu@ariel.ac.il

Abstract: Risk management is an ongoing process that includes several stages of mapping and identification, analysis, and evaluation, planning, and implementation to reduce risks and ensure ongoing control. Risk management along the supply chains has become more significant in recent years due to an increased complexity of the relationships between components in the chain as well as various disruptions such as climate change, COVID-19, or geo-political scenarios. The current literature alongside the increase in complexity and frequency of risk events, leads us to the single, most prominent challenge in risk management today: the auditor's subjectivity in determining the risk levels. Simply stated, two different auditors may assess a given situation differently due to their specific history and experience. Specifically, it seems to be extremely difficult to find cases in which different auditors, working on the same organization, made the same risk assessment. With that in mind, this research aims to reduce the human subjectivity bias and reach a risk evaluation that is as objective as possible, by using the machine learning approach. For this aim the paper introduces a new risk assessment framework based on factors analysis and artificial neural network as the predictive model. We first introduced a new approach of deconstructing the risk factors into their basic elements and analyzing them as a feature vector. Next, we collected unique, real-world data of risk surveys and audit reports from 60 industrial companies of various industries (from plastic and metal factories to logistic and medical devices companies). Lastly, we constructed a neural network to predict the risk levels of operational processes in the industry. We trained our model on 42 samples and managed to achieve a R^2 score of 0.9 on the test set of 18 samples. Our model was validated and managed to predict the risk accuracy with $R = 0.95$ in accordance with the human auditor results.

Keywords: supply chain; risk management; operational risk



Citation: Burstein, Guy, and Inon Zuckerman. 2023. Deconstructing Risk Factors for Predicting Risk Assessment in Supply Chains Using Machine Learning. *Journal of Risk and Financial Management* 16: 97. <https://doi.org/10.3390/jrfm16020097>

Academic Editors: Ștefan Cristian Gherghina and David Liu

Received: 28 December 2022

Revised: 27 January 2023

Accepted: 1 February 2023

Published: 6 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Operational risk management is an ongoing process that includes several stages designed to reduce risks and ensure ongoing controlled processes. In the industry, risk can be viewed as the effect of uncertainty on achieving one's goals. Thus, risk is the probability that any event that affects the target quantitatively or qualitatively might occur. The practical effect of the risk is a joint function of its occurrence probability and its potential damage. In practice, it is also customary to refer to elements that reduce damage by applying protection and control mechanisms.

Risk management in the industry and supply chains has become more significant in recent years due to the increased complexity of the relationships between different components in the chain. While in previous years mass production required fewer components and modifications, today's supply chain includes many more participating parties (suppliers, customers, regulators, competitors), and therefore is more vulnerable to disruptions and malfunctions.

In addition to the complexity of the production processes, we are currently witnessing an increase in various types of disruptions causing substantial losses, for example, climate change, the COVID-19 pandemic, and varied geo-political scenarios (Russia–Ukraine

war). Key factors affecting vulnerable points include volatility and disruptions on the demand side by customers; supply-side concerns (e.g., fluctuations in raw material prices, supply failures, logistical difficulties); regulatory and bureaucratic issues (e.g., political changes, tax systems, administrative barriers); infrastructure deficiencies (in production and information systems); and catastrophes such as terrorist attacks, natural disasters and pandemic (Wagner and Bode 2009). Another work (Qazi et al. 2016) proposed classifying the risk categories into those arising from non-coordination between supply and demand, and risks arising from disruption of activity. Alternatively, the risks can be divided into accidents, operational catastrophes, and strategic uncertainty. For example, the recent COVID-19 pandemic caused a worldwide shortage of containers, reduced flights, and higher rate of worker absences due to lockdowns causing billions of dollars in damages (Pereira Gaspar et al. 2020). This example demonstrates the issue of reliability and accuracy in delivery dates and the importance of a structured risk management process to reduce exposure and disruptions in the flow of supplies.

The current literature alongside the abovementioned increase in complexity and frequency of risk events, leads us to the single, most prominent challenge in risk management today: the auditor's subjectivity in determining the risk levels. Simply stated, two different auditors may assess a given situation differently due to their specific history and experience. While the Delphi method can be used to assess risk and reduce subjectivity, it requires averaging the assessments of a large panel of auditors (e.g., Perera et al. 2014), which obviously is not an applicable practice due to financial costs to the firms. As of today, it seems to be extremely difficult to find cases in which even two different auditors, working on the same organization, made the same risk assessment and evaluation.

This challenge is worth investigating as a subjective risk assessment process might cause various problems for industries and organizations: first, low risk assessment results might damage the companies in their preparation for mitigating these risk factors, which might lead to severe monetary losses. Second, subjective assessment leaves too much room for managers to "cherry pick" their risk evaluator in a way that will not interfere with their managerial agendas or strategic plans for the organization. Lastly, the subjectivity and freedom in the risk assessment operation also reduces the accountability from the auditors themselves, which consequently might lead to poor professional expertise.

With that in mind, this research aims to reduce the human subjectivity bias and reach a risk evaluation that is as objective as possible, by using a machine learning approach. Such an algorithm will be the first stage in a holistic AI tool that will support managers' strategic decision making, as it is well established that increasing supply chain resilience could be very influential and important for organizations (Belhadi et al. 2021).

Specifically, in this work we developed a new methodology to automate and objectify the risk assessment process as follows. First, we introduce a new approach of deconstructing the risk factors into their basic, atomic elements and analyze them as a feature vector, that can be utilized as an input layer in a machine learning model. The vector represents the actual status of each sub-process as a numeric parameter convention. Next, we collected unique, real-world data of risk surveys and audit reports from 60 Israeli industrial companies of various industries (from plastic and metal factories to logistic and medical devices companies). These reports were manually cleansed and converted from textual written reports to our predefined set of 17 deconstructed risk factors. This rare data resulted in a small but highly valuable data set for training a machine learning model. Noteworthy, this transformation from written text to predefined categorical data reduces the amount of noise in the data as it creates a well-defined reference point for the human auditors to consider. Lastly, we constructed a neural network-based algorithm to predict the risk levels of operational processes in industry. We trained our model on 42 samples and managed to achieve an R^2 score of 0.9 on the test set of 18 samples. We validated our model on two additional samples and managed to verify that similarity of the real risk assessment and the predicted risk assessment as was output by our model.

Our work presents the first step in moving the risk assessment practice from a subjective, qualitative assessment, to a more objective, quantitative one. The method comprised the entire risk assessment process: the auditor will approach the firm with a well-defined set of parameters to review, his numeric report will then be fed into a trained machine learning model, and a risk value output will be given by the model. To summarize, our suggested method can be deployed as a useful tool to companies and organizations that are interested in evaluating their risk management process in an unbiased approach.

2. Literature Review

For many years, mathematical tools have been used to study risk management in supply chains. In recent years, research has begun to consider additional tools based on artificial intelligence and multi-agent systems. Aspects covered include (1) supply chain and production risk management classifications, (2) modern tools for dealing with risk management challenges in our complexed environment, and (3) specific solutions for isolated parts of the process. Despite the improvements these offer, they do not suggest a general and systematic approach to risk management problems.

2.1. Definitions and Categories

A relevant paper was an extensive literature review of 224 international articles on supply chain risk management, published between 2003 and 2013 (Ho et al. 2015). The review classified the papers into the categories of “definitions,” “types,” “factors,” and “methods” for managing risks in the supply chain. Additionally, it also classified risk management methods as either quantitative or qualitative according to their main processes, including risk detection, risk assessment, risk reduction, and risk monitoring. Moreover, together with perspectives gleaned from other researchers, a list of potential factors affecting the five common types of risk in demand, supply, manufacturing, transport, and information was also developed. This was a starting point for creating a supply chain risk index model and was consequently used in our work as an additional resource for defining the deconstructed risk factors. The review paper also points out ten gaps in the literature and proposed corresponding research directions, which the authors believe will help researchers conduct influential studies in supply chain risk management. Another recent literature review (Pournader et al. 2020) gave a thorough insight into recent years and current research in the field of SCRM and some future directions. Initial findings of the literature review show that data analysis for supply chains is a growing issue, and a range of predictive and synthetic documents emphasize its promise and importance. The literature pays scant attention to the use of machine learning for supply network management, except for few predictive analyses of demand. A literature survey on risk management and artificial intelligence (Baryannis et al. 2019a) revealed that 53% of the examined studies do not provide decision-making ability, while 39% do not offer administrative insights to assist decision-making. Only 8% provide a decision support system, mostly in agent-based categories, machine learning, and big data. Hence, most studies are limited in their ability serve as the basis for a decision-making framework. Based on these findings, we now consider the case of predicting supply disruptions, which is very topical. Over the past decade, researchers have noted that the increased complexity in large-scale global supply chains has made them vulnerable to network-related environmental, organizational, and organizational risks, causing supply chain risk management to become a fundamental problem.

2.2. Digitization and Big Data

Industry and supply chain risk management has evolved in the era of information systems and digitization. A study reviewing the impact of digitization, Industry 4.0, and big data concluded that adopting new technologies may contribute to reducing the domino effect in supply chain risk (Ivanov et al. 2019). However, it explicitly states that more research is required before offering decision-supporting systems for risk management, including the use of agent and machine learning algorithms to optimize stochastic systems.

Another research (Fan et al. 2015) offers a model for managing supply chain risks in the era of big data, emphasizing planning and process control. It collects and analyses extensive data from many sensors and internal/organizational and external sources to provide decision makers with a report on the situation, which could then be used to redesign the chain. The authors (Giannakis and Louis 2016) show that supply chain data analysis can also be implemented in purchasing and managing supply risks and supplier performance, allowing global supply chains to adopt an initiative instead of a reactive response to supply chain risks. However, in all the above-mentioned work there is no predictive model based on the collected data, but only a practical data fusion for ease of presentation.

2.3. New Tools and Frameworks

Several recent papers offer solutions for specific parts of the supply chain. For example, using analytical tools in big data systems to detect fraud by rogue suppliers to reduce risks in the supply chain and increase its reliability (Zage et al. 2013). The ability of machine learning techniques and methods to predict the duration of activities in projects was demonstrated in (Braga et al. 2008), using several methods including decision tree, support vector machines, bagging-bootstrap aggregating predictors, and neural networks. The study concludes that using integrated machine learning techniques improved the reliability and accuracy of project duration forecasts. In our opinion, some of these techniques should also be applied to the supply chain risk management framework. Agent-based approaches and networks can predict and learn, and one study (Giannakis and Louis 2016) offers a multi-agent system for managing risks in the supply chain capable of learning. Preliminary tests have been conducted of a few machine learning techniques; however, they were left as a theoretical practice without any usage of real data. Another recent study (Baryannis et al. 2019b) provides conceptual frameworks for managing supply chain risks using machine learning techniques and big data but did not continue to implement and evaluate this framework. Despite there only being a few studies that implement learning, they conclude that it should be possible to rely on AI to uncover new knowledge that decision makers can combine with their expertise to make optimal decisions in the supply chain risk management process.

The untapped potential of big data analysis and AI in supply chain risk management has also been identified by (Ivanov et al. 2019), whose digitalization framework presents a cyber supply chain built on traditional supply chain analysis. Their cyber supply chain relies on harnessing the power of big data, the Internet of Things, the cloud, and blockchain technologies. The authors emphasize the need for research that examines the relationship between these technologies and risk management in the supply chain, a need their current survey confirms. Throughout the literature, it is evident that different AI techniques are partially applicable to the various stages of risk management in the supply chain (e.g., mathematical programming). Approaches succeed in reducing risks but are unable to support automated decision-making and automated learning. Agents and machine learning techniques are less effective in modelling complex systems such as supply networks.

That said, there may be limitations on adopting machine learning to evaluate risks in industry. To date, a thorough review of machine learning applications for engineering and risk assessment is yet to be published (Hegde and Rokseth 2020). Therefore, a structured examination is needed to answer research questions concerning the machine learning methods for risk assessment, including the best methods suited for the purpose; industry leaders in adopting machine learning; applying and testing machine learning; geographical trends; types of data used to develop algorithms; what stage risk assessments are best assisted by machine learning; and what trends can be expected. The extensive literature review covering a variety of industries provides an up-to-date snapshot in the field. Descriptive analysis is widespread, however, except for demand forecasting, few present cases show how data analysis can be used in supply chains.

Indeed, this extensive review reveals no literature on the use of machine learning or new methodologies to improve risk management processes and accuracy as we propose in

this framework. Therefore, we propose exploring integration, modelling, and optimization combined with AI techniques capable of making automated decisions based on prediction and learning for effective management of risks in the supply chains. In this research we selected the main risk factors based on the reviewed papers and their relevancy to supply chain and production companies. From our point of view, the selected risk factors reflect the relationships between companies in the supply chain as well as operations within the company. We mainly focused on the major production part of the chain and on its interfaces with suppliers and customers.

3. Materials and Methods

Understanding the risk management process is important for mapping the current gaps and challenges prior to applying a new computerized model that improves the accuracy of analyses made by decision makers. In this section, we briefly describe the workflow commonly applied by auditors when conducting risk assessment and managing risk. Our methodology is based on minimizing the human bias by systematic deconstruction of the risk factors and evaluating them in high resolution to improve accuracy. In this research we used an artificial network model with input layer of risk factors and output of risk result (probability of the risk) to train a model for predicting future risk. Substantial materials were reviewed and used by the authors (risk surveys, audit reports, and international standards and regulations) to establish a generic set of risk factors that will serve as the skeleton of the research on deconstructing risk elements.

3.1. Risk Management Process Method in Practice Today

The risk management process is an ongoing effort that includes identifying, analyzing, and assessing the risks and providing responses to each scenario on strategic and operational management levels. During the risk identification stage, the possible causes of disruptions and the impact of the events are examined for both external and internal factors. The work begins with data collection, interviews, and observations of the processes. Next, the process of risk analysis includes determining the risk level of each event given the probability, results, additional effects, and control mechanisms for reducing the risk and their effectiveness. Finally, a comprehensive assessment of the risk considers the previous stages and consequently decisions can be made with respect to addressing each risk. Decisions may range from taking no action and accepting the full risk (for low-risk events) to taking a range of preventive actions; most severely, a target might be abandoned due to the high risk. The flow chart in Figure 1 shows the risk management process, which is also accepted in the industry and is commonly used as a standard method.

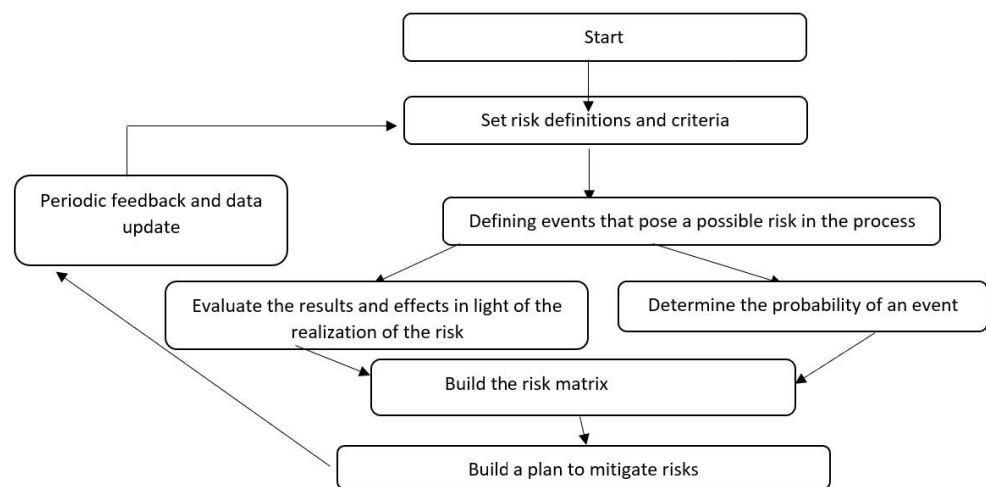


Figure 1. Risk assessment process flow chart.

The risk matrix (RM) model is a commonly used method to describe risks in the industry (Thomas et al. 2013). This methodology, presented in Figure 2, shows that a higher risk level, (the red squares in the RM model) is the product of elements very likely to occur and the severity of the potential damage. Higher risk levels call for an immediate response, while lower risk levels (the green squares) are risks that can be accepted without taking prompt significant actions.

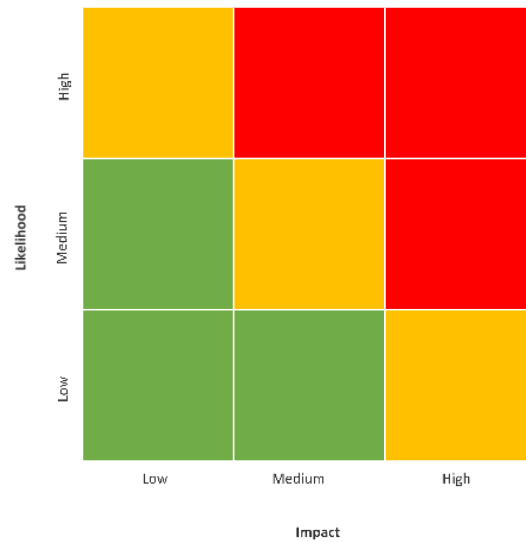


Figure 2. Risk matrix model.

Another interesting model was proposed in a research (Paltrinieri et al. 2019) that contains an additional dimension (the z-axis) that describes the experts’ knowledge of the probabilities that are related to the uncertainties in the model, as shown in Figure 3. Risk factors that have a poor expert knowledge may increase their overall risk assessment. To counter it, the decision makers might decide to gather additional knowledge on that risk factor, which will clarify its potential effect. Similarly, we strive to reduce the uncertainties in the risk assessment process by injecting additional expert knowledge into the framework via machine learning algorithms.

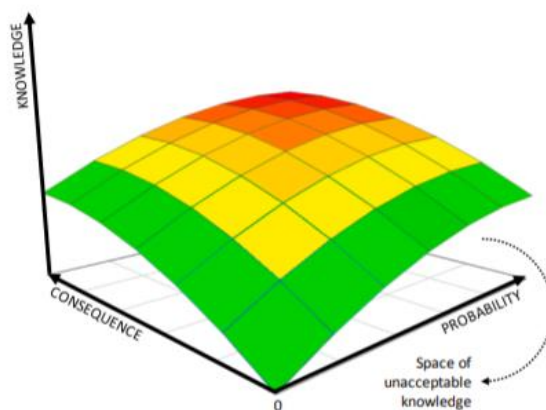


Figure 3. Risk matrix with knowledge dimension.

3.2. Implementing Neural Networks in the Risk Assessment Process

To begin developing our suggested methodology, we first created an exhaustive list of parameters that are commonly examined by risk auditors and built a standard checklist that covers most of the operational risks. Next, we converted the findings of 60 genuine risk survey reports in accordance with that checklist. We then fitted the relevant total

operational risk of each survey, according to actual findings, as a dataset to train a machine learning model to predict such risks.

3.2.1. Data Gathering and Pre-Processing

To meet the goal of applying machine learning tools, sufficient data need to be collected and preprocessed. Risk assessment work is unique because the data available for the auditors are usually highly classified business information that is not publicly available. Moreover, human auditors do not use commonly defined methods and scales for the risk assessment process. Therefore, attaining data for the purpose of our goal was challenging.

3.2.2. Risk Factors—The Selection and Design Phase

- The selection of the most relevant risk factors is based on analyzing over 700 audit reports, risk surveys and research papers, produced in 2004–2022, in the field of operational risk management. This step is a unique opportunity to gain insights into the best practice process using confidential business information.
- Thorough review of relevant standards and regulations in the risk management such as ISO standards ([ISO 31000:2018\(En\)](#), [Risk Management—Guidelines n.d.](#)), SOX–Sarbanes–Oxley Act ([Lander 2002](#)), COSO model ([COSO 2013](#)), and COBIT framework ([De Haes and Van Grembergen 2009](#)). These standards were used by the authors in their experience in the industry.
- Applying the Pareto 80-20 principle ([Harvey and Sotardi 2018](#)) to select the most commonly used risk factors in the supply chains and operations world. In the processing and selection phase we neglected unique and specific risk factors that were not covering most industries, to produce the commonly applied factors.
- Analyzing each of the selected risk factors and breaking them down into their sub-elements in which the auditors may easily assess, as well as will fit as an input vector into our ML algorithm.

It is important to note that the current state of the risk assessment practice is that the human auditors do not have a well-defined and agreed upon set of parameters that they investigate. Some industry standards do exist (as shown above), but many others, those of a more qualitative nature, do not have an agreed upon procedure for their assessment. For instance, reviewing what is known as “corporate culture” might yield different practices. Some auditors will look at ethical codes, others at HR protocols, and yet others might review management practice regarding whistleblowers. Each auditor has its own way of doing things, and the academic literature does not have insights into that matter. Our selected parameters (see below) are not claimed to be optimal or exhaustive. They provide a first step of common parameters that are well agreed upon, and in the future specific industries should strive to provide a more fine-tuned list for themselves.

3.2.3. Methodology for Numerical Evaluation

Each parameter was given a value: 1 = an appropriate level of performance, 0 = a neutral level, and –1 = material weakness. Table 1 shows the complete list of these 17 parameters. In addition to these parameters, each compiled report included risk probability value, in accordance with the human auditor assessment. To accurately determine the values for each parameter {–1, 0, 1}, each parameter was graded based on a set of common sub-parameter values.

Table 1. List of parameters in the input layer.

No.	Parameter
1	Organizational culture and corporate governance
2	Document processes and procedures
3	New vendor handling process
4	Purchase orders
5	Inventory control
6	Payment control
7	Sole supplier dependency
8	Sole raw material dependency
9	Long lead time
10	Information systems supporting the processes
11	Frequency of the action
12	Frequency of failure in the process
13	Ability to stop the action in case of hazard
14	Level of mission's complexity
15	Dependency of accumulated success factors (non-parallel)
16	Dependency of accumulated failure factors (parallel)
17	Susceptibility to random and unstable factors on the process completion

For example, the first parameter, “Organizational culture and corporate governance” was compiled from the following sub-parameters:

- **High level of organizational responsibility**—Is there an ethical code of conduct for the employees? Do they receive periodic training regarding this? Is there a culture of debriefing and learning?
- **Commitment to the Organization**—Is there a high level of employee turnover? Are the employees and managers loyal to the system? Were there issues of breach of contract or procedures by the staff? Were there integrity problems with the staff?
- **Culture of exploring and learn lessons**—Are there procedures regarding audits and learned lessons? Is it well documented?
- **Culture of correcting defects**—Are there procedures regarding corrective actions? Is there a plan for testing the implementation of correct procedures?
- **Credibility and integrity of the management team and staff**—Are there procedures for employee screening? Does the organization conduct credibility tests? Are employees required to sign NDA?
- **High level of employee turnover**—Is the rate of turnover relatively high? What is the mean seniority period?
- **Employee training and certification processes**—Are there special training sessions for new workers? What is the certification procedure for employees? Is there an annual training plan?
- **Collaborations between employees**—What is the level of collaboration between the teams and between individuals?
- **Human backing and preservation of organizational knowledge**—Are there documented procedures and thorough documentation of the important know-how information?
- **Organizational structure appropriate to needs**—Are there sufficient employees in the existing structure to enable proper operation? Are the employees as professional as required?

Here are some detailed examples of the sub-parameters:

- **Procedures and process documentation**
 - Procedures are updated and well documented
 - Detailed and approved procedures
 - Distributed procedures to relevant staff
 - Procedures are implemented in practice
- **Vendor handling**
 - Checking new suppliers and confirming procedure
 - Existence of ISO certification for the suppliers
 - Carrying out supplier surveys in the field
 - Annual supplier evaluation procedure
 - Control of changes in supplier details
- **Establishment of purchase orders**
 - Approval of requirements by a qualified entity
 - Receiving multiple quotes and lowering costs
 - Arranging a purchase order is organized
 - Order confirmation and budget approval
- **Receivers to inventory**
 - Warehouse inspection procedures
 - Testing and counting records
 - The inventory system is updated
- **Payment control**
 - Control over orders and receipts
 - Existence of control between the certificate of receipt and the invoice
 - Adequate control over suppliers and payments
- **Dependence on sole supplier and relationships with suppliers**
 - Dependence on major suppliers
 - Long-term relationships with reliable suppliers
- **Dependence on a major raw material**
 - Dependence on main raw material or special spare parts
- **Long turnaround times and items phase out equipment**
 - Long lead time items
- **Information systems support**
 - High data integrity
 - Back up procedures
 - Professional IT staff
 - Data security and BCP (business continuity plan) procedures

4. Results

4.1. Applying the New Model on Real Database

The authors managed to acquire real operational risk data from surveys conducted by different professional auditors. These data were made available because of the personal expertise and access of the principal author, a senior risk analyst, and his colleagues. Real risk assessment reports from the last ten years were gathered and compiled by quantifying the free form text and numerical assessments according to seventeen parameters we defined as relevant risk factors.

To apply our model to real world companies, we used risk surveys prepared by professional auditors at 60 selected companies. The first task was to fit the narrative and findings of each process into our parameters and relevant checklists to create a stable database for the model. For example, we chose two companies: company #2, a metal recycling company in northern Israel that had many operational problems including working accidents and

frauds; and company #3, a manufacturer of plastic pipes in eastern Israel, which is well run and fully controlled. These reports were prepared and edited over a period of three months to provide the boards of directors with insights into the operational risks of each company to reduce risk and improve efficiency. Table 2 shows how the text from the risk surveys was converted into a table of parameters. Then, each parameter in each company was graded {−1, 0, 1} according to the adequacy of its execution of the said operation, as a sum of its preliminary sub-parameters.

Table 2. Example of two companies from the sample.

Parameter#	Company 2	Company 3	Details
Organizational culture and corporate governance	−1	1	Company #2 has high turnover and personnel problems; and few scams and frauds were discovered. Company #3 has very low staff turnover; a good code of conduct, and ethical procedures are circulated by management
New vendor handling process	−1	1	Company #2 has poor procedures and documents regarding selecting and appraising suppliers. Company #3 has adequate procedures for selecting suppliers and well-established annual appraisals.
Purchase orders	0	1	Company #2 has poor procedures and documents for purchase order confirmation and control. Company #3 has an adequate ERP system that applies rules to control purchase processes.
Inventory control	0	1	Company #2’s inventory has no ERP control and the inventory count shows many inventory differences. Company #3 has very good inventory control and reports show that inventory integrity is high.
Information systems supporting the processes	−1	1	Company #2 has much software, but without interfaces and connectivity between them, the IT staff has many missions but does not respond promptly due to shortage of professional staff. Company #3 has a strong IT team and very good support on time to service calls.
Total risk probability	0.7	0.4	

The total operational risk probability calculated using the authors’ methodology corresponds directly to the findings of the risk survey reports. The probabilities by the human auditor were given in a matrix or in a text that reflect their assessment as low, medium, or high. We transformed the assessment to numerical values, where 0–0.3 represent low probability, 0.4–0.7 represent medium probability, and 0.8–1 refer to high-risk probability. In each company, the narrative and the risk matrix correlated with the risk score and faults and weaknesses that were found. Company #2 received a probability value of 0.7, which represents relatively high risk. Company #3 has stronger controls and minor weaknesses, and was graded 0.4, which is relatively low. In this preliminary project, we set the total risk probability as a target, to train the model.

4.2. Set-Up Phase and Training

With the newly available data, we are now faced with a supervised predication problem. There are data from 60 observations, each with 17 parameters (valued {−1,0,1}), and a score in the range [0,1] that represents the total risk assessed for that specific organization. Appendix A presents a sample of this dataset. We then constructed a neural network with 17 input neurons, 20 hidden neurons in the next layer and a single output neuron which outputs a value in the [0,1] range. This architecture is not claimed to be optimal for the problem at hand and was constructed ad hoc via trial and error. To train the network we divided the dataset into 42 training samples, 9 validation samples, and 9 testing samples.

The model was constructed in MATLAB and the network was trained using two different algorithms, Levenberg–Marquardt and Bayesian regularization. The latter yielded a higher level of accuracy.

4.3. The Trained Neural Network Results

After training the neural network and completing the validation and testing, the findings fit well. The R correlation coefficient fit between output data and the target of the model was high ($R = 0.951$), which represents a significant positive linear relationship and allows proper prediction with relatively low error rate. Table 3 shows the sets’ structures and results after the runtime. The graph in Figure 4 was created at the end of the process runtime and shows the parameters resulting from the network.

Table 3. Sample size and composition.

Set	Samples	MSE	R
Training	42	0.00244	0.941
Validation	9	0	0
Testing	9	0.00256	0.952
Total	60		

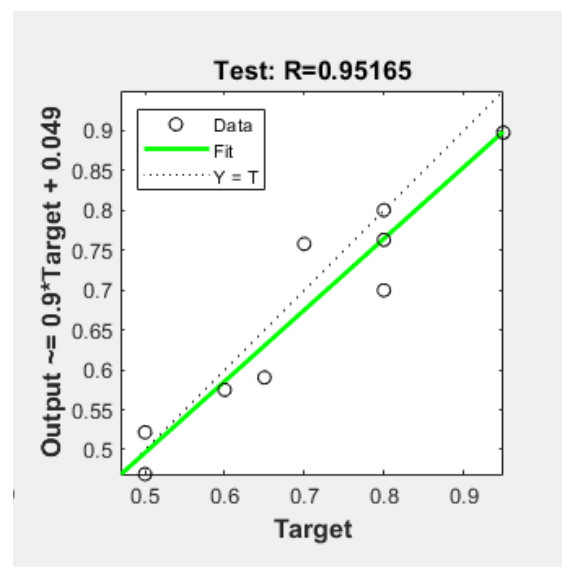


Figure 4. Output results of the ANN model. $R^2 = 0.90563$. Output = $0.9 \times \text{Target} + 0.049$.

4.4. Validation Phase

To validate our new model, we used two additional, real-life test cases and predicted their risk probabilities, without introducing them to our model in the training or testing phases. Our goal was to predict the risk levels computed by the algorithm and comparing it to the score of the auditors’ assessment in the field.

As described above, our proposed prediction method using a new model requires preparation, preprocessing the data and formatting for the input layer structure, following construction of the neural network that produces the computed risk probability. We analyzed the findings for two plants, retrieved raw data from audit reports, organized it by parameters and graded each item. Plant A had poor internal controls; its risk survey and audit showed many weaknesses. For example, it received (−1) for “Document processes and procedures” due to the lack of written work procedures on the production floor and no standards or protocol distribution to employees. Plant B had better management and working procedures, therefore receiving much high scores (1) due to its appropriate

practices. Finally, we entered the 17 parameters for each plant into the neural network and executed the predict operation to receive the calculated output of the model.

After running the new dataset, the risk probabilities computed by the model showed very good fit with the auditor's conclusions in their field reports. The model, as shown in Table 4, assigned plant A, with its weak procedures, a high-risk probability and received a poor rating by the auditor (output = 0.8738), while plant B, with robust internal controls, strong working procedures, and well-managed staff that also received a good rating by the auditor (that reflects a low-risk probability) yielded a low-risk probability by our model (output = 0.4366). For research purposes, we are satisfied by these initial results, and believe that the proposed development will improve the fit, thereby reducing errors in the network and application of our new decision-support tool.

Table 4. Parameters of two plants for validating the model.

Parameter	Plant A	Plant B
Organizational culture and corporate governance status	−1	1
Document processes and procedures level	−1	1
New vendor handling process	0	1
Purchase orders control	−1	0
Inventory security and safety levels control	0	0
Payment control	−1	1
Sole supplier dependency risk	−1	0
Sole raw material dependency risk	0	0
Long lead time risks	0	1
Information systems supporting the processes	−1	0
Frequency of the risky actions	0	1
Frequency of failure in the process	0	1
Ability to stop the action in case of hazard	−1	0
Level of mission's complexity	−1	1
Dependency of accumulated success factors (non-parallel)	−1	1
Dependency of accumulated failure factors (parallel)	0	0
Susceptibility to random and unstable factors on the process completion	−1	1
Computed risk probability by the ANN	0.8738	0.4366

5. Discussion

In this work we developed a new methodology to automate and objectify the risk assessment process. First, we introduced a new approach of deconstructing the risk factors into their atomic elements and analyzing them as a feature vector, that can be utilized as an input layer in a machine learning model. The vector represents the actual status of each sub-process as a numeric parameter convention. Next, we collected unique, real-world data of risk surveys and audit reports from 60 mid-size (between 100 and 600 employees) manufacturing companies from various sectors (from plastic and metal factories to logistic and medical devices companies with suppliers, production, and customer-related operations). These reports were manually cleansed and converted from textual written reports into our predefined set of deconstructed risk factors. This rare data resulted in a small but highly valuable data set for training a machine learning model. Noteworthy, this transformation from written text to predefined categorial data reduces the amount of noise in the data as it creates a well-defined reference point for the human auditors to consider. Lastly, we constructed a neural network to predict the risk levels of operational

processes in the industry. Specifically, we trained our model on 42 samples and managed to achieve an R^2 score of 0.905 on the test set of 18 samples. We validated our model on two additional samples to verify that the predicted risk probability that was output by our model is like that of the human auditors.

There are additional theoretical and practical implications that can be drawn from this research. This new framework can be applied to predict risk probability in the same context as the prediction of activity duration in projects as was suggested in (Braga et al. 2008), using neural networks tools. Their study concluded that using integrated machine learning techniques improved the reliability and accuracy of project duration forecasts. Another implication that can be deduced from this work is of harnessing the power of big data in the risk management process (Ivanov et al. 2019). This can be done, for example, by establishing a universal database of elementary risk factors and their associated probabilities, by many auditors in their specific industries. The practical usage of the wisdom of the crowd as a vast input vector to the machine learning algorithm will increase the model's reliability and can be customized to specific sectors due to their unique risk factors. Another example in the same context is of automatic data collection from various sensors (either human or automated), which would enable a real-time adaptation of the risk assessment matrices. For instance, upon sensing low inventory levels or missing human resources forms, such automated assessment would adapt its parametric values and the overall risk assessment will revise automatically. Our parametric model is general enough to be adapted in full or using specific parts of it according to data availability.

To the best of our knowledge, our work suggested here is the first of its kind that tries to formally conceptualize the whole practice of the risk assessment process. Interviews with expert auditors in their fields revealed that there is no formal education for that job and no professional accountability. This, in turn, leads to a reality in which different auditors look at different aspects of the firms and present different risk matrices. The lack of professional accountability is one of the main factors for subjectivity in the analysis and lack of formal methodology in conducting the work. Our work presents the first step in moving that practice from a subjective qualitative assessment, to a more objective and quantitative one. We suggested a method that comprised the entire process: the auditor will approach the firm with a well-defined set of parameters to review, his numeric report will then be fed into a trained machine learning model, and a risk value output will be given by the model. To summarize, our suggested method can be deployed as a useful tool by companies and organizations that are interested in evaluating their risk management process in an unbiased approach. While there are still many challenges to tackle (some of which can be seen in the next section below), the main contribution of our work is that of the presentation of the framework itself.

6. Conclusions

The overarching goal of the research was to create a methodology and a basis for an automated tool for risk assessment use in the industry. The findings of our initial work shows that machine learning is a highly viable and effective paradigm for both shortening the time needed to prepare a risk report and increasing the accuracy and objectivity of the system that calculates the weighted risk. The high level of correlation found in the research allows us to take the basic model a step further, and attempt to expand it by introducing the ability to integrate automatic input from various sensors that provide data from the factory's production, logistical, and computer systems. With that in mind, our vision for future work may continue in three different but complementary directions: a methodological framework, theoretical risk arithmetic, and finally, a basis for a stand-alone application for risk assessment in the field.

Our study takes the first step in an overall objective risk assessment process; however, it is not free of limitations. First, the dataset comprised 60 manufacturing-oriented companies. It might be the case that other types of companies, such as research and development-oriented firms will require additional parameters that are finer-tuned to

their intrinsic characteristics. Second, the process of transforming the written reports into categorized parameters still includes some form of subjective bias. We hope that in the future this bias will be further reduced by working according to a formal protocol. Lastly, the machine learning model that was constructed is most probably not the optimal one for the dataset. Extensive research should be done to assess the problem parameters and fit the best possible algorithm and meta-parameters to achieve better classification results. Specifically, such optimization is of utmost importance as the size of the data set is small with respect to the large amount of input neurons.

In respect to future work, first, with our newly constructed model that can predict and quantify risk probability based on a set of predefined data parameters, we can proceed to construct a methodological framework to unify auditors' processes for working with the model. The main goal of this methodology is to decrease the variability in assessments by different auditors due to subjective bias in the assessment process by using structured checklists containing few sub-levels. Next, we aim to lay the foundations for a theoretical analysis for objectively measuring the risk probabilities. We will try to develop a systematic approach to deconstruction of the risk factors into their predecessor factors and view them as vectors in an array of elements. Our intuition is that it should be possible to view a single operational risk as a compound value, according to its basic sub-factors and the structure of their relationships as vectors. From our perspective, we can also indicate a research direction for the relations between some risk factors that have low risk probability and impact but produce a significant effect when combined. The last phase may offer an automated decision support system based on our algorithm, enabling the management team to obtain available RM updates on an ongoing basis. This framework may be a significant practical contribution of our new model that will reduce bias and dependency of the human auditor and introduce automation to the risk assessment process, and it is not limited to supply chain risks only. Another aspect in future directions will be to conduct interviews with the auditors themselves to obtain their takes and insights on the new methodology and maybe use that information to construct a valid and established work procedure for risk assessment.

Author Contributions: Both authors contributed equally to all processes. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

In this appendix we present a small fragment of 6 companies out of 60 that were used to train and test the ANN model. Each company (marked 1–6) has detailed reports and risk surveys in our records. Obviously, we must give the companies a serial number to keep disclosure of discrete information. Our convention of scoring each parameter is by allocating a numeric value: 1 = an appropriate level of integrity/low risk, 0 = a neutral risk level, and –1 = material weakness/high risk as shown in Table A1:

Table A1. Part of the dataset used to train the neural network.

Parameter\Company Number	1	2	3	4	5	6
Organizational culture and corporate governance	0	-1	1	0	0	1
Document processes and procedures	-1	0	0	-1	0	0
New vendor handling process	0	-1	1	-1	0	1
Purchase orders	1	0	1	1	1	1
Inventory control	0	0	1	0	0	1
Payment control	1	1	1	0	0	1
Sole supplier dependency	-1	0	1	-1	-1	0
Sole raw material dependency	0	0	0	-1	-1	1
Long lead time	0	-1	-1	0	-1	1
Information systems supporting the processes	0	-1	1	0	0	1
Frequency of the action	0	-1	0	-1	0	0
Frequency of failure in the process	0	-1	-1	0	0	0
Ability to stop the action in case of hazard	1	0	1	-1	0	0
Level of mission’s complexity	1	0	0	0	1	0
Dependency of accumulated success factors (non-parallel)	1	1	1	1	0	1
Dependency of accumulated failure factors (parallel)	0	0	1	-1	0	0
Susceptibility to random and unstable factors on the process completion	0	0	1	-1	0	0
Final risk probability	0.6	0.7	0.4	0.65	0.6	0.4

References

Baryannis, George, S. Dani, and Grigoris Antoniou. 2019a. Predicting Supply Chain Risks Using Machine Learning: The Trade-off between Performance and Interpretability. *Future Generation Computer Systems* 101: 993–1004. [CrossRef]

Baryannis, George, Sahar Validi, Samir Dani, and Grigoris Antoniou. 2019b. Supply Chain Risk Management and Artificial Intelligence: State of the Art and Future Research Directions. *International Journal of Production Research* 57: 2179–202. [CrossRef]

Belhadi, Amine, Venkatesh Mani, Sachin S. Kamble, Syed Abdul Rehman Khan, and Surabhi Verma. 2021. Artificial Intelligence-Driven Innovation for Enhancing Supply Chain Resilience and Performance under the Effect of Supply Chain Dynamism: An Empirical Investigation. *Annals of Operations Research* 1–26. [CrossRef] [PubMed]

Braga, Petronio L., Adriano L. I. Oliveira, and Silvio R. L. Meira. 2008. Software Effort Estimation Using Machine Learning Techniques with Robust Confidence Intervals. Paper presented at the 7th International Conference on Hybrid Intelligent Systems (HIS 2007), Kaiserslautern, Germany, September 17–19; pp. 352–57.

COSO. 2013. COSO Internal Control-Integrated Framework Frequently Asked Questions. *Committee of Sponsoring Organisation of the Treadway Commission* 6: 1–11.

De Haes, Steven, and Wim Van Grembergen. 2009. Enterprise Governance of Information Technology. Available online: <https://link.springer.com/book/10.1007/978-3-030-25918-1> (accessed on 4 April 2022).

Fan, Yingjie, Leonard Heilig, and Stefan Voß. 2015. Supply Chain Risk Management in the Era of Big Data. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9186: 283–94.

Giannakis, Mihalis, and Michalis Louis. 2016. A Multi-Agent Based System with Big Data Processing for Enhanced Supply Chain Agility. *Journal of Enterprise Information Management* 29: 706–27. [CrossRef]

Harvey, H. Benjamin, and Susan T. Sotardi. 2018. The Pareto Principle. *Journal of the American College of Radiology* 15: 931. [CrossRef] [PubMed]

Hegde, Jeevith, and Børge Rokseth. 2020. Applications of Machine Learning Methods for Engineering Risk Assessment—A Review. *Safety Science* 122: 104492. [CrossRef]

Ho, William, Tian Zheng, Hakan Yildiz, and Srinivas Talluri. 2015. Supply Chain Risk Management: A Literature Review. *International Journal of Production Research* 53: 5031–69. [CrossRef]

ISO 31000:2018(En), Risk Management—Guidelines. n.d. Available online: <https://www.iso.org/obp/ui/#iso:std:iso:31000:ed-2:v1:en> (accessed on 4 April 2022).

Ivanov, Dmitry, Alexandre Dolgui, and Boris Sokolov. 2019. The Impact of Digital Technology and Industry 4.0 on the Ripple Effect and Supply Chain Risk Analytics. *International Journal of Production Research* 57: 829–46. [CrossRef]

Lander, Guy P. 2002. The Sarbanes-Oxley Act of 2002. *Journal of Investment Compliance* 3: 44–53.

Paltrinieri, Nicola, Louise Comfort, and Genserik Reniers. 2019. Learning about Risk: Machine Learning for Risk Assessment. *Safety Science* 118: 475–86. [CrossRef]

Pereira Gaspar, Priscila Gomes, Paula Santos Ceryno, Ana Luiza Ferrer, and Antônio Márcio Tavares Thomé. 2020. Phases and Tools for Supply Chain Risk Management: A Systematic Literature Review. *Gestao e Producao* 27: 1–21.

- Perera, B. A., R. Rameezdeen, N. Chileshe, and M. R. Hosseini. 2014. Enhancing the Effectiveness of Risk Management Practices in Sri Lankan Road Construction Projects: A Delphi Approach. *International Journal of Construction Management* 14: 1–14. [[CrossRef](#)]
- Pournader, Mehrdokht, Andrew Kach, and Srinivas Talluri. 2020. A Review of the Existing and Emerging Topics in the Supply Chain Risk Management Literature. *Decision Sciences* 51: 867–919. [[CrossRef](#)] [[PubMed](#)]
- Qazi, Abroon, John Quigley, Alex Dickson, and Konstantinos Kirytopoulos. 2016. Project Complexity and Risk Management (ProCRiM): Towards Modelling Project Complexity Driven Risk Paths in Construction Projects. *International Journal of Project Management* 34: 1183–98. [[CrossRef](#)]
- Thomas, Philip, Reidar B. Bratvold, and J. Eric Bickel. 2013. The risk of using risk matrices. *SPE Economics & Management* 6: 56–66.
- Wagner, Stephan M., and Christoph Bode. 2009. Dominant Risks and Risk Management Practices in Supply Chains. *International Series in Operations Research and Management Science* 124: 271–90.
- Zage, David, Kristin Glass, and Richard Colbaugh. 2013. Improving Supply Chain Security Using Big Data. Paper presented at the IEEE ISI 2013–2013 IEEE International Conference on Intelligence and Security Informatics: Big Data, Emergent Threats, and Decision-Making in Security Informatics, Seattle, WA, USA, June 4–7; pp. 254–59.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.