12th Conference on Learning Factories, CLF2022

Training Module on Probabilistic Forecasting of Production and Intralogistics Turbulences within Learning Factories

Jan Schuhmacher[a,*], Vera Hummel[a]

[a]*ESB Business School, Reutlingen University, Reutlingen, Germany*

Abstract

Industrial practice is characterized by random events, also referred to as internal and external turbulences, which disturb the target-oriented planning and execution of production and logistics processes. Methods of probabilistic forecasting, in contrast to single value predictions, allow an estimation of the probability of various future outcomes of a random variable in the form of a probability density function instead of predicting the probability of a specific single outcome. Probabilistic forecasting methods, which are embedded into the analytics process to gain insights for the future based on historical data, therefore offer great potential for incorporating uncertainty into planning and control in industrial environments. In order to familiarize students with these potentials, a training module on the application of probabilistic forecasting methods in production and intralogistics was developed in the learning factory "Werk150" of the ESB Business School (Reutlingen University). The theoretical introduction to the topic of analytics, probabilistic forecasting methods and the transition to the application domain of intralogistics is done based on examples from other disciplines such as weather forecasting and energy consumption forecasting. In addition, data sets of the learning factory are used to familiarize the students with the steps of the analytics process in a practice-oriented manner. After this, the students are given the task of identifying the influencing factors and required information to capture intralogistics turbulences based on defined turbulence scenarios (e.g. failure of a logistical resource) in the learning factory. Within practical production scenario runs, the students apply probabilistic forecasting using and comparing different probabilistic forecasting methods. The graduate training module allows the students to experience the potentials of using probabilistic forecasting methods to improve production and intralogistics processes in context with turbulences and to build up corresponding professional and methodological competencies.

Peer Review statement: Peer-review under responsibility of the scientific committee of the 12th Conference on Learning Factories 2022.

Keywords: Intralogistics; Probabilistic Forecasting; Turbulences; Analytics; Machine Learning;

1. Introduction

Manufacturing companies are subject to numerous unforeseen events (so-called turbulences) in their daily operations, which hinder the execution of the target-optimized production plan (e.g. in the area of sequence planning, machine assignment, etc.) [1,2]. Turbulences can be differentiated with regard to their places of origin into internal and external turbulences. Internal turbulences originate in unexpected changes within the company (e.g. machine breakdowns, process delays), while external turbulences (e.g. delay of a delivery of a supplier, change of an order by the customer) arise due to reasons external to the company [3]. The planning process of companies is particularly influenced by medium-term external (e.g. changed demand quantities) and internal (e.g. new product variants) turbulences [3,4]. Short-term turbulences, such as machine breakdowns or process delays, are mainly facing the target-oriented control of the production and intralogistics processes in terms of unexpected deviations from the plan [5]. In general, the detection and especially the forecasting of turbulence offers great

* Corresponding author. Tel.: +49 7121 271 3112
E-mail address: jan.schuhmacher@reutlingen-university.de

potential for improving production planning and control to maintain an target-optmized (cost and performance) production supported by appropriate intralogistics processes [6].

For this purpose, a training module was developed for the learning factory "Werk150" of the ESB Business School (Reutlingen University), which provides students a fundamental and practice-oriented insight into the field of probabilitic forecasting of production and intralogstics turbulences embedded in a holistic data analytics process.

2. Analytics and probabilistic forecasting

In the following the scientific fields of "Analytics" and "Probabilistic forecasting" are analysed, which are essential for the developed training module on probabilistic forecasting for production and intralogistics.

2.1. Analytics

Following *Schuh et al.* [7] and *Hill & Berry* [6], the term "analytics" refers to the scientific process of mathematically and logically transforming historical data into insights to explain the past, as well as using those insights for better decision making in the future. With increasing maturity of analytical skills, a distinction is made between descriptive (*What happened?*), diagnostic (*Why did it happen?*), predictive (*What will happen (with which probability)?*) and prescriptive analytics (*How can we make it happen?*), which all build on each other [6–8].

The descriptive analytics sets the starting point for analysis by analyzing what has happened based on statistics and instruments as tables and charts to organize, summarize and visualize data [6,9]. So descriptive analytics is used to determine from the data set how often corresponding turbulences have occurred. After the descriptive analytics of the data, the diagnostic analytics of the dataset to find potential patterns, root causes or links between attributes is conducted. Therefore data-driven machine-learning algorithms as unsupervised learning methods as k-means clustering can be applied [6]. These are used to analyse correlations and interactions between the attributes, such as the influence of a turbulence on the process time. The next step, which also sets the focus of this paper, is the predictive analytics of the data to perform forecasts based on the datasets. Therefore methods and models are applied to perform forecasts on the occurance of turbulences based on identified patterns of the data set [6,7]. Probabilistic forecasting methods represent a specific type of these forecasting methods, which will be discussed in more detail in the next section. For the final step of the analytics process of prescriptive analytics simulation and optimization approaches are applied in order to positively influence the future through targeted decisions, e.g. to prevent turbulences through targeted measures like the initiation of an additional intralogistics material transport in order to avoid a shortage of material at a work station.

With regard to the analytics approach, a distinction has to be made between model-driven and data-driven approaches [10,11]. In classical statistics, model-driven approaches are used, while data mining and machine learning approaches are data-driven approaches. Model-driven approaches consider the data generating process to be known and explicit stochastic behavior is assumed for the process. In data-driven approaches, the data-generating process is considered unknown and thus no assumption of an explicit stochastic model (e.g., multivariate normal distribution) is made for the process in question. Instead, algorithms are used that use input data to pursue specific goals [11].

For the developed training module to forecast turbulences in the learning factory, a data-driven approach has been chosen, since the data-generating turbulence-triggering process is considered unknown and likewise actual data from production scenarios from the learning factory are to be used.

2.2. Probabilistic forecasting

A probabilistic forecast represents an estimate of the corresponding probability of all possible future outcomes of a random variable. Unlike single value forecasts, such as time series averages or quantile forecasts, probabilistic forecasts represent a probability density function considering future quantities or events [12,13]. For the prediction of turbulence events in the production system, these probabilistic forecasting methods are particularly suitable, since they allow the determination of the probabilities of occurrence of all turbulence events considered and not only one particular event as for single value forecasts. By using simulators, a probabilistic forecast can be also carried by using originally deterministic forecast methods by varying the attribute values and/or combining different forecasting scenarios as in the case of ensemble forecasts.

Common data-driven probabilistic forecasting methods are supervised learning methods as ensemble methods, deep learning methods, methods based on Bayes' theorem and methods using regression analysis.

Among the most popular ensemble methods are Random Forests and Gradient Boosted Trees, which can be extended from their deterministic origin to probabilistic forecasting methods. These combine many weak deterministic predictors (decision trees in the case of Random Forests and Gradient Boosted Trees) to create a

superior deterministic predictor. To turn the ensemble method into a probabilistic forecasting method, the mixing or combining process is adapted to include probabilities instead of a single aggregated outcome [12].

Deep Learning methods are based on a multi-layer feedforward artificial neural network that is trained with stochastic gradient descent [14]. The input data is processed through the neuronal network and the created output is then compared to the desired output. The result of the comparison is an error which is fed back into the neural network to adjust the weights for each node aiming on minizing the error with each iteration [15]. Deep Learning models are probabilistic, because the gradient descent converges faster when the loss function reflects a probabilistic forecast [12].

The method of the naïve bayes classifier is a probabilistic classification method, which uses the Bayes' theorem. The main assumption is that given the value of the class, the value of any attribute is conditionally independent of the value of any other attribute [16,17]. Although this independence between input attributes and target attributes is only given in rare cases, this method often allows good (forecast) results even in cases where this precondition is not met and only little data is available [16,17].

Regression analysis methods, such as Logistic Regression, Support Vector Machines and Fast Large Margin can also be used for probabilistic forecasts. By their very nature, they produce particularly good results when there are strong relationships or correlations between a categorical target variable and one or more input variables [15].

2.3. Tool selection

To select a suitable tool for data analysis as well as data-driven pattern recognition and probabilistic prediction, a qualitative comparison of different tools was conducted. Based on the preliminary work of *Flückiger* [18], the method libraries for performing descriptive, diagnostic, predictive and prescriptive analytics, the effort required for data import and model creation and the usability of the tools RapidMiner, SPSS, Python (Pandas), R (CRAN-Project) and Minitab have been compared. The aspects of user-friendliness and a low effort for data import and model creation are particularly important for the developed training module, since the students do not yet have any specific previous experience in the field of data mining. Due to the high usability, a low effort for the data import and the creation of the (forecast) models as well as a consistent coverage of the required analytics functionalities with corresponding method libraries, the decision was made to use the tool RapidMiner for the training module. RapidMiner also provides a wide range of forecasting methods and models (Ensemble, Deep Learning, Regression Analysis, etc.). In addition, the Auto Model function of RapidMiner allows to automatically create a process and model based on the input data, which enables the comparison of different forecasting methods in a very short time in order to identify the most suitable method(s) to forecast intralogistics and production turbulences.

3. Training module on probabilistic forecasting

To familiarize students with the potentials of probabilistic forecasting for production and intralogistics in a practice-oriented manner, a training module has been developed in the learning factory “Werk150” at ESB Business School (Reutlingen University) which will be described in the following.

3.1. Didactical concept and training module

The didactical concept of the developed training module follows the approach of learning factories comprising formal, informal and nonformal learning based on the active involvement and own actions of the students in the physical learning factory environment [19]. The teaching methods used for the training module are following the four-step method according to *Riffelmacher* [4]. In the first step, students are taught the theoretical foundations of data analytics and probabilistic forecasting in particular in the setting of a classical lecture. In the second step, students are introduced to the application of analysis and forecasting methods using examples from various disciplines. In the third step, the students receive a prepared example in the form of an event log file from previous production runs in the learning factory, which contains various production-related turbulence events. In the fourth step, students apply what they have learned independently in the course of a practical production scenario with intralogistics turbulences (failure of a transport system, delay of a transport) triggered by the instructor.

3.2. Theoretical introduction and examples from other disciplines

The theoretical introduction to the training module builds on the students' prior knowledge of classical statistics and summarizes the essential basics (e.g. probability, correlation, stability, standard deviation) as a recap. This is followed by an introduction to the topics of analytics in general and probabilistic forecasting in particular including common methods as described in section 2. The practical relevance and wide range of applications of probabilistic

forecasting is introduced to the students with examples of model-driven weather forecasting to predict wind speed for renewable energies and a data-driven forecast to predict electricity consumption for a city.

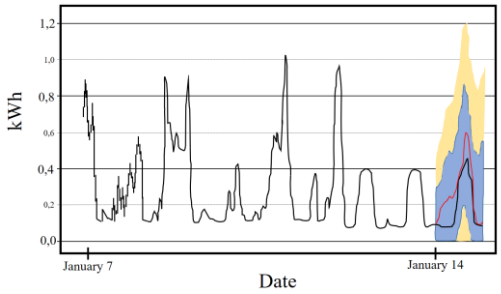


Fig. 1. Probabilistic forecasting of electricity consumption

Fig. 1 displays an example of a data-driven forecast of electricity consumption with the 70% quantile (blue), 95% quantile (yellow), the point forecast of average simulated consumption (red line) and the actual electricity consumption (black line). Considering the 70% quantile of the electricity consumption forecast it can be estimated with a probability of 70% that the energy consumption will be between 0,2 and 0,82 kWh at peak time. This allows the students to understand the advantages of probabilistic forecasting for reasoned decision making in comparison to single value forecasting.

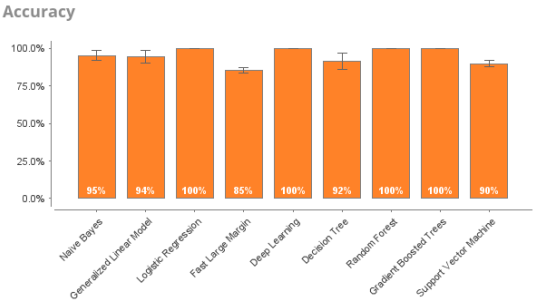
3.3. Exercise on probabilistic forecasting of production turbulences

After the theoretical introduction, students receive an introduction to the production scenario to be considered at Werk150, which serves as the basis for the predictive and prescriptive data analyses to be performed by the students. The students are also introduced to the selected tool RapidMiner and are asked to analyze a sample event log file of an assembly workstation following the steps of descriptive, diagnostic, predictive and prescriptive analytics aiming on the investigation of production-related turbulences. After each of these steps, the results are discussed in class. The used event log file has been generated based on aggregated information out of the ERP and MES and contains information as order-ID, product variant, workstation, worker experience in hours, actual process time in minutes and if a turbulence has been detected. The considered production-side turbulences are resource breakdowns (Robot in error state), process delay (Target vs. actual time deviation at workstation) and missing parts at a workstation.

The students are starting with the import of the event log file into RapidMiner to perform a descriptive and diagnostic analysis of the sample data. Amongst other things, the descriptive, statistical analysis shows that the event log contains in total 375 lines of order data of which the majority (70.67%) were processed without the influence of turbulence at the considered workstation. Diagnostic analysis of the data set using the k-Means cluster algorithm shows, among other things, that the data set has 5 clusters corresponding to the considered turbulence events and that there is a correlation between the turbulence events and the process times.

For the predictive analytics, the students select the attributes to be considered for the forecast of turbulences on their own and compare the results of different models (Naïve Bayes, Logistic Regression, Deep Learning, etc.) based on the classification error/accurancy, standard deviation and total time. In order to ensure a certain comparability of the results, 60% of the data set is used for setting up and training the model and 40% for testing the model to check the forecast accuracy. The students also note that due to the data-driven forecasting approach, the forecasting quality deteriorates when the data set is reduced in size, especially when numerous input attributes are to be included for forecasting. The results of the students regarding the used attributes, models and achieved results are then collected, compared and discussed.

Fig. 2 shows the results of the performed forecast of turbulence events considering the attributes of process time and worker experience. In total 4 out of the 9 forecasting models considered achieved a forecasting accuracy of 100 %. One reason for the very good forecasting results of the models of Logistic Regression, Deep Learning as well as the ensemble methods of Random Forests and Gradient Boosted Trees is the correlation of an increase in process time with the occurrence of turbulence, which were particularly well detected by these methods. Nevertheless, the Logistic Regression and Deep Learning models are the most favorable forecasting methods considering the forecast accuracy of 100% and a total time of 3 seconds.



Model	Accuracy	Standard Deviation	Gains	Total Time
Naive Bayes	95.3%	± 3.2%	50	3 s
Decision Tree	91.5%	± 5.3%	46	3 s
Generalized Linear Model	94.4%	± 4.0%	40	3 s
Logistic Regression	100.0%	± 0.0%	60	3 s
Fast Large Margin	85.2%	± 1.9%	30	6 s
Deep Learning	100.0%	± 0.0%	64	3 s
Random Forest	100.0%	± 0.0%	64	17 s
Gradient Boosted Trees	100.0%	± 0.0%	64	17 s
Support Vector Machine	89.8%	± 1.9%	48	9 s

Fig. 2. Comparison of forecasting models to forecast production-related turbulences

The probabilities of certain turbulence events for different attribute value constellations can be determined in RapidMiner via a corresponding simulator with graphical output of the probabilities and influencing factors that support or contradict the occurrence of the turbulence event (see Fig. 3). To perform a prescriptive analysis, for example to find out how to prevent turbulence in the future, an optimization based on an Evolutionary Algorithm can be performed using the corresponding prediction model. For the example shown in Fig. 3, it can thus be determined that the occurrence of the event "No turbulence" can be supported by a higher worker experience. In addition, it can be analysed that especially the probability of the turbulence of a process delay due to a jammed screw can be greatly reduced by a higher worker experience.

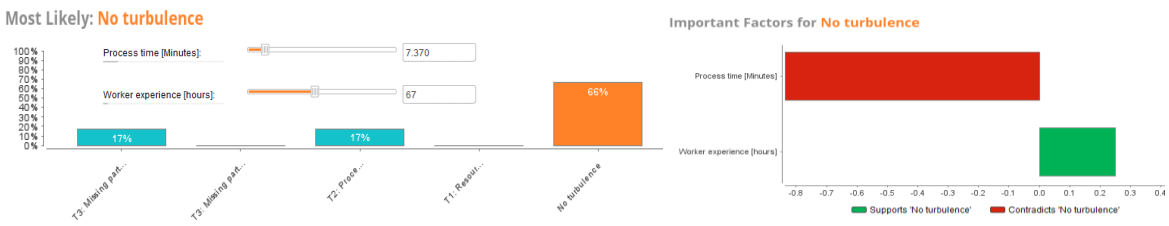


Fig. 3. Probabilistic forecast of production turbulences using Deep Learning (left) and relevant influencing factors for turbulences (right)

3.4. Practical production scenario run with intralogistics turbulences at Werk150

For the practical production scenario run, the students are given the task of defining the input information required to forecast the intralogistic turbulences "Delayed material delivery" and "Resource breakdown of an automated guided vehicle (AGV)" in addition to the production-related turbulences of the introductory example and to determine how this information can be captured. Afterwards, the students are producing in the production system of the Werk150 under the influence of logistic turbulences triggered by the instructors. For example, the AGV is switched off a few times for a certain period of time (Turbulence "Resource breakdown of an AGV") and the delivery of a transport vehicle with material to the workplace is delayed for a short period of time (Turbulence "Delayed material delivery"). Subsequently, the students are given the task to analyze the resulting event log file and determine the appropriate methods for predicting the logistical turbulence in combination with the production-related turbulences.

To capture the turbulence events, the status data of the AGV (Turbulence "Resource breakdown of an AGV") and time information (Turbulence "Delayed material delivery") from the MES are used for the event log data set. For the analysis of the data set, students again go through the steps of descriptive, diagnostic, predictive and prescriptive analytics. When analyzing the data (diagnostic analytics), it becomes apparent that there are causal relationships between the production-related turbulences and the intralogistics-related turbulences. The intralogistics turbulences of a delayed material delivery or the failure of an AGV thus leads to the turbulence events of missing parts in production, which also results in increasing process times in production.

Fig. 4 (left) shows a comparison of the models used to forecast intralogistics turbulences considering the input attributes of process time and turbulence events of production. Due to the direct relationship (correlation) between the considered turbulences of production and intralogistics, the comparison of the forecasting accuracy of the considered models shows that the regression analysis methods of Logistic Regression, Support Vector Machines as well as the Fast Large Margin model provide the best forecasting results. Since the most favorable forecasting models differ in comparison to the introductory example from chapter 3.3, it has to be highlighted that a parallel comparison of different forecasting methods is of great importance in addition to a general knowledge of the specifics of the forecasting methods.

Model	Accuracy	Standard Deviation	Gains	Total Time
Naive Bayes	90.7%	± 3.4%	12	2 s
Generalized Linear Model	88.8%	± 7.3%	8	1 s
Logistic Regression	100.0%	± 0.0%	32	3 s
Fast Large Margin	100.0%	± 0.0%	32	4 s
Deep Learning	90.7%	± 3.4%	12	2 s
Decision Tree	90.7%	± 3.4%	12	2 s
Random Forest	90.7%	± 3.4%	12	12 s
Gradient Boosted Trees	90.7%	± 3.4%	12	12 s
Support Vector Machine	100.0%	± 0.0%	32	5 s

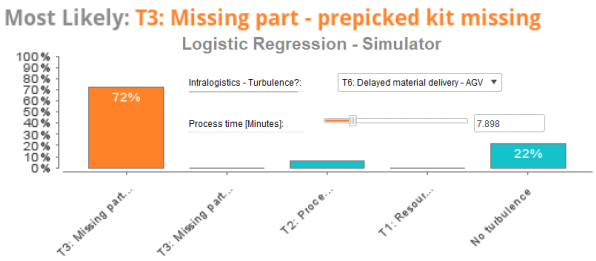


Fig. 4. Comparison of models for probabilistic forecasting of logistics turbulences (left) and probabilitic forecast of turbulences (right)

In order to determine with which probabilities an intralogistics turbulence (e.g. delayed material delivery) leads to a turbulence in production, a probabilistic forecast of production turbulence can be carried out on the basis of the generated forecasting models. For example it can be determined by using the Logistic Regression model that the probability of occurrence of the production-related turbulence "Missing part - prepicked kit missing" is highest at 72 % if there is a delayed material delivery by the AGV (Turbulence “Delayed material delivery – AGV”), which then leads to a specific process delay in production (also see Fig. 4, right). This probabilistic information can then be used in a next step to adapt production and logistics processes in a predictive manner to these changed conditions at short notice. For example, in the event of (potential) turbulence-related process delays in production, subsequent orders can be allocated to alternative workstations and the logistics processes can be adapted to supply an alternative workstation.

4. Conclusion and outlook

The developed training module allows the students to gain a practical and application-oriented insight into the potential of probabilistic forecasting embedded in the concept of analytics. Through the intended integration of this training module into the graduate study programs, the data basis can be continuously expanded and the suitability of the corresponding forecasting methods can also be verified in the connection with a rising amount of data. This will generate further scientific insights into the suitability of specific forecasting methods and models for forecasting turbulences in production and intralogistics. These probabilistic information can then be integrated into production planning and control in order to be able to deal with (potentially) occurring turbulences in the production system in a predictive and target-optimized (cost and performance) manner.

References

[1] F. Frankenbach, A. Riess, P. Michaeli, G. Reinhart, Kontextbasierte Rekonfiguration von Fabriksystemen, ZWF 115 (2020) 751–754.
[2] R. Kupers (Ed.), Turbulence: A corporate perspective on collaborating for resilience, Amsterdam University Press, Amsterdam, 2014.
[3] E. Westkämper, E. Zahn, Wandlungsfähige Produktionsunternehmen: Das Stuttgarter Unternehmensmodell, Springer-Verlag, Berlin, Heidelberg, 2009.
[4] P. Riffelmacher, Konzeption einer Lernfabrik für die variantenreiche Montage, Disseration, 2013, Fraunhofer-Verl., Stuttgart, 2013.
[5] H.-H. Wiendahl, Auftragsmanagement im turbulenten Umfeld: Teil 1–Anforderungen, WT Werkstatttechnik online 96 (2006) 183–189.
[6] R. Hill, S. Berry, An Introduction to Industrial Analytics, in: R. Hill, S. Berry (Eds.), Guide to Industrial Analytics, Springer International Publishing, Cham, 2021, 3–20.
[7] G. Schuh, M. Blum, J. Reschke, M. Birkmeier, Der Digitale Schatten in der Auftragsabwicklung, ZWF 111 (2016) 48–51.
[8] Gartner, Inc., Gartner Says Advanced Analytics Is a Top Business Priority, https://www.gartner.com/en/newsroom/press-releases/2014-10-21-gartner-says-advanced-analytics-is-a-top-business-priority.
[9] D.L. Olson, G. Lauhoff, Descriptive Data Mining, 2nd ed. 2019, Springer, Singapore, 2019.
[10] L. Breiman, Statistical modeling: The two cultures, Statistical science 16 (2001) 199–231.
[11] M. Freitag, M. Kück, A. Ait Alla, M. Lütjen, Potentiale von Data Science in Produktion und Logistik: Teil 1 – Eine Einführung in aktuelle Ansätze der Data Science, Industrie 4.0 Management 31 (2015) 22–26.
[12] J. Vermorel, Probabilistic Forecasting (Supply Chain), [April 26, 2021], https://www.lokad.com/probabilistic-forecasting-definition.
[13] T. Gneiting, M. Katzfuss, Probabilistic Forecasting, Annu. Rev. Stat. Appl. 1 (2014) 125–151.
[14] H2O.ai, Deep Learning (Neural Networks), https://docs.h2o.ai/h2o/latest-stable/h2o-docs/data-science/deep-learning.html.
[15] R.V. McCarthy, M. McCarthy, W. Ceccucci, L. Halawi, Applying predictive analytics: Finding value in data, Springer, Cham, 2019.
[16] RapidMiner, Operator Reference Guide - Naive Bayes, [January 03, 2022], https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/bayesian/naive_bayes.html.
[17] O.Z. Maimon, L. Rokach, Decomposition methodology for knowledge discovery and data mining: Theory and applications, World Scientific, Hackensach, NJ, 2005.
[18] I. Flückiger, Data Analytics in Organisations and Business: 5.5 Selection of Software Tools, [January 04, 2022], https://stat.ethz.ch/education/semesters/as2015/analytics/Data_Analytics_2015-11-25.pdf.
[19] E. Abele, J. Metternich, M. Tisch, Learning Factories, Springer International Publishing, Cham, 2019.