



**HAL**  
open science

# Prediction of Cold Chain Transport Conditions Using Data Mining

Clayton Gerber Mangini, Nilsa Silva Lima, Irenilza De Alencar Nääs

► **To cite this version:**

Clayton Gerber Mangini, Nilsa Silva Lima, Irenilza De Alencar Nääs. Prediction of Cold Chain Transport Conditions Using Data Mining. IFIP International Conference on Advances in Production Management Systems (APMS), Aug 2020, Novi Sad, Serbia. pp.616-623, 10.1007/978-3-030-57997-5\_71 . hal-03635641

**HAL Id: hal-03635641**

**<https://inria.hal.science/hal-03635641v1>**

Submitted on 20 Jun 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Prediction of Cold Chain Transport Conditions using Data Mining

Clayton Gerber Mangini<sup>1[0000-0001-9107-5435]</sup>, Nilsa Duarte da Silva Lima<sup>1[0000-0002-1284-7810]</sup>  
and Irenilza de Alencar Nääs<sup>1[0000-0003-0663-9377]</sup>

<sup>1</sup> Graduate Studies in Production Engineering, Paulista University,  
São Paulo, SP 04026-002, Brazil  
clayton.mangini@gmail.com, irenilza@gmail.com,  
nilsa.lima@stricto.unip.br

**Abstract.** Ensuring the delivery of temperature-controlled products in transportation is an increasing challenge, especially in countries with continental extension and tropical climate such as Brazil. Products with this type of specificity generally have a higher added value and involve specialized equipment and labor. Thus, route mapping is necessary for the logistics of the cold chain. The study aimed to predict the transport conditions in the cold chain. The data set analyzed includes the temperature of the loads and the route information (Southeast to Northeast and South of Brazil). The classification of temperature excursions considered data below 15°C or above 30°C. The Naïve Bayes and Multilayer Perceptron algorithms are used to predict the optimal temperature excursion model. The Multilayer Perceptron algorithm proved to be the most suitable for a thermal route mapping model. With this identified standard, logistics decision making can be improved to reduce waste and ensure product integrity with less recourse.

**Keywords:** Classifiers, Naïve Bayes, Multilayer Perceptron.

## 1 Introduction

The cold chain begins in the production phase. The raw materials needed to produce a specific product are processed, followed by the transport of the product at low temperatures under strict conditions, refrigerated storage and its subsequent distribution by land, sea or air, finally, is acquired by the patient [1-2]. The cold chain is part of the supply chain, where temperature monitoring plays a vital role in the system. At the same time, automatic data collection is an essential aspect of cold chain management [3-7,2].

Cold chain logistics refers to a systematic operation where perishable products require a low-temperature environment. They are collected in an environment with specific temperature ranges in each process before reaching the final consumer. It is necessary temperature management in the production, transportation, storage, distribution, and marketing of these drugs, which not only guarantee a high-quality product but reduce waste in transportation and reduce cost [1-2,5,7]. Therefore, the manufacturer must balance the two situations and choose the best product distribution strategy to

avoid degradation and maintain product quality [4,5,8]; and increased public health [9,10,11].

In this way, continuous temperature monitoring and the availability of reliable reports for these products are essential [12,13]. Thus, the relevance of this theme is justified because, despite technological advances, temperature monitoring in the cold chain of pharmaceutical products is still a challenge for the value chain of pharmaceutical products. This study aimed to classify the transport conditions of pharmaceutical products using data mining techniques.

## 2 Material and Methods

The study included temperature excursion trials in two Brazilian regions during the summer and winter of 2019.

### 2.1 Parameters tested during transport tests

The drugs were conditioned in thermal packaging under adequate environmental conditions and transported land ways by truck to the Northeast and South regions of Brazil, the parameters included in the tests to determine the optimal temperature excursion are presented in Table 1.

**Table 1.** Temperature excursion parameters set by manufacturers.

Excursions	Optimal temperature excursion conditions
15 a 30 °C	Yes
< 15 a > 30 °C	No

Source: [14].

### 2.2 Prediction of optimal temperature excursion conditions

The study data were collected at a pharmaceutical distribution company in São Paulo and contain anonymous records regarding transportation and temperature excursion. Data mining applies to this study, through techniques (algorithms), to support the transportation planning of pharmaceutical products in search of strategic information that allows the extraction of implicit information existing in the databases, contributing to the process identify and classify new standards [15-17]. The steps of the data mining process are selection, pre-processing, data mining, and post-processing for knowledge discovery [16, 17]. The results obtained with data mining can be used in information management, information request processing, decision making, and process control. The data contained in the databases are used to learn a particular target concept [15-18]. Tasks performed by data mining techniques (such as classification) allow the building of models that can be applied to unclassified data. Such a procedure categorizes

them into classes, to relate the meta attribute (whose value will be predicted) and a set of prediction attributes [15-18].

Data pre-processing of the data was performed in Microsoft Excel® spreadsheets for further processing in Weka® data mining software version 3.8 [19-22]. The attributes used for the construction of the predictive model in data mining by the classification technique (*modeling classification*) were: "seasons," "period," "forecasted-routes," "temperature," and "optimal-temperature-excursion." A pre-processing stage was performed in the data to obtain the best result in the classification. In that stage, we performed a discretization of the attributes "period" (late, night, dawn and morning) and "optimal-temperature-excursion" (yes and no, Table 1). Discretization reduces and simplifies data, making learning faster, and results in denser [15].

The algorithms Naïve Bayes (`weka.classifiers.bayes.NaiveBayes`) and Multilayer Perceptron (`weka.classifiers.functions.MultilayerPerceptron`) were applied to the construction of a model of rules for the prediction of the temperature excursion depending on the route (northeast, south), the period (dawn, morning, afternoon, night) and the season (summer, winter). The validation of the model was parameterized using *cross-validation*. The validation was applied both for the training set with 75% (5,308 instances) of the total data. The testing was done using 25% (1,770 instances) of the total data (7,078 instances).

The classification process was performed in two stages. In the first step, a training model (training set) was constructed that describes a predetermined set of data classes. In the second step, the model is used for sorting. The validation method is a simple technique that uses a labeled class test set. The test samples are randomly selected and are independent of the training samples. The filter was used *Stratified Remove Folds* to stratifying the data set, to separate the training set and the test set. All classification algorithms used in this study used the cross-validation test option (*cross-validation*).

### 2.3 Evaluation of the prediction model

The last step is to evaluate the prediction; this analysis was made based on the performance values obtained through the prediction model test [23]. The performance evaluation measures of prediction models used were confusion matrix, accuracy, precision, sensitivity (recall), and Matthews correlation coefficient (MCC). The Kappa statistic measures the learning capacity of the algorithm. The confusion matrix is a matrix with results obtained during the model test step, widely used in models that use classification algorithms. It is also called the confusion matrix of a hypothesis, this hypothesis offers a valid measure of the classification model, by showing the number of correct classifications versus the predicted classifications for each class, over a set of examples. The number of hits, for each class, is on the main matrix diagonal, and the other elements represent errors in the classification (Table 2). The confusion matrix of an ideal classifier has all these elements equal to zero since it does not make mistakes.

**Table 2.** Confusion matrix

Prediction	Observed True	Observed False
Expected true	TP	FP
Expected false	FN	TN

TP = true positive; FN = false positive; FP = false negative; TN = true negative. Source: [23].

### 3 Results and Discussion

The mean metrics for unbalanced classes of the Naïve Bayes algorithm presented the following results: the correctly classified instances were equal to 86.61%, instances classified incorrectly equal to 13.39%, with Kappa statistics equal to 0.64. Considering the average accuracy detailed by class, the true positive rate was equal to 0.87, the false positive rate 0.21, accuracy 0.87, sensitivity 0.87 (recall), and the MCC was equal to 0.64. For the Multilayer Perceptron, the average metrics for unbalanced classes were 93.73% for correctly classified instances, 6.27% for incorrectly sorted instances, 0.82 for Kappa statistics. For the average accuracy detailed by class, the results were equal to 0.94 for true positive rate, 0.16 for false-positive rate, 0.94 for accuracy, 0.94 for sensitivity (recall), and 0.82 to MCC. The results of the comparison of Naïve Bayes and Multilayer Perceptron algorithms comparing unbalanced and balanced classes are presented below in Table 3.

**Table 3.** Classifier performance for unbalanced class prediction models

Class imbalanced	Classifier model - Stratified cross-validation					
	Naïve Bayes			Multilayer Perceptron		
Metrics						
Correctly classified instances	86.61%			93.73%		
Incorrectly classified instances	13.39%			6.27%		
Kappa statistic	0.64			0.82		
Detailed accuracy by class	Class			Class		
	Yes	No	WA*	Yes	No	WA*
TP Rate	0.90	0.75	0.87	0.98	0.79	0.94
FP Rate	0.24	0.10	0.21	0.21	0.02	0.16
Precision	0.92	0.71	0.87	0.94	0.94	0.94
Recall	0.90	0.75	0.87	0.98	0.79	0.94
MCC	0.64	0.64	0.64	0.82	0.82	0.82

\*WA = Weighted average.

The performance of the Multilayer Perceptron, within the unbalanced classes, was higher than the Naïve Bayes. That includes the lower false-positive rate, the higher percentage of correctly classified instances, lower percentage of ranked instances incorrectly and the Kappa statistic close to 1 (0.82) indicating the proper performance of

the algorithm to classify the temperature excursion to the northeast and southern regions of Brazil (Table 3).

The metrics for balanced classes of the Naïve Bayes algorithm presented mean values of 73.48% for correctly classified instances, 26.52% for instances classified incorrectly, with Kappa statistics equal to 0.47. The average accuracies detailed by class presented values equal to 0.73 for the true positive rate, 0.26 for the false-positive rate, 0.75 for accuracy, 0.74 for sensitivity (recall), and 0.48 for the MCC. The Multilayer Perceptron presented average performance metrics for balanced classes, higher than the Naïve Bayes with 96.07% of correctly classified instances, 3.93% of instances sorted incorrectly, including 0.92 for Kappa statistics. The results were equal to 0.96 for true positive rate considering the average accuracy detailed per class; 0.04 for false-positive rate, 0.96 for accuracy, 0.96 for sensitivity (recall), and 0.92 to MCC. The results of the comparison of algorithms comparing unbalanced and balanced classes for the Naïve Bayes and Multilayer Perceptron algorithms are presented below in Table 4.

**Table 4.** Classifier performance for balanced class prediction models

Class balanced	Classifier model - Stratified cross-validation					
	Naïve Bayes			Multilayer Perceptron		
Metrics						
Correctly classified instances	73.48%			96.07%		
Incorrectly classified instances	26.52 %			3.93%		
Kappa statistic	0.47			0.92		
	Class			Class		
Detailed accuracy by class	Yes	No	WA*	Yes	No	WA*
TP Rate	0.62	0.85	0.73	0.98	0.94	0.96
FP Rate	0.85	0.38	0.26	0.06	0.02	0.04
Precision	0.80	0.69	0.75	0.94	0.98	0.96
Recall	0.62	0.85	0.74	0.98	0.94	0.96
MCC	0.48	0.48	0.48	0.92	0.92	0.92

\*WA = Weighted average.

In class balancing, Multilayer Perceptron performance outperformed Naïve Bayes, including lower false-positive rate, the higher percentage of correctly classified instances, lower percentage of instances misclassified and Kappa statistics. These class performance metrics presented values close to 1 (TP Rate, Precision, Recall, MCC) that indicates the proper performance of the algorithm to classify the temperature excursion condition for the northeast and southern regions of Brazil (Table 4).

The Naïve Bayes algorithm correctly classified the "yes" class for optimal temperature excursion (true positive = TP), with an average rate of 87% for unbalanced classes and 73% for balanced classes. The class "no," for an excursion outside the temperature parameters presented a rate of 75.47% for unbalanced classes and 84.81% for balanced classes (true negative = TN) indicating a proper performance of the algorithm in the classification task (Table 5).

However, the Multilayer Perceptron algorithm correctly classified the "yes" class for optimal temperature excursion (true positive = TP). With an average rate of 94% for unbalanced classes and 96% for balanced classes, and the "no" class for an excursion outside the temperature parameters had a rate of 79.21% for unbalanced classes and 94.16% for balanced classes (true negative = TN) indicating better performance than Naïve Bayes in the classification task (Table 6). The results obtained confirm the possibility of predicting the optimal temperature specifications in the transport of pharmaceutical products.

**Table 5.** Naïve Bayes classification confusion matrix

Class imbalanced - Naïve Bayes			
a	b	Total	Classified as
1,210	132	1,342	a = yes
105	323	428	b = no
1,315	455	1,770	
Class balanced - Naïve Bayes			Classified as
549.99	335.01	885	a = yes
134.40	750.0	885	b = no
684.39	1085.61	1,770	

**Table 6.** Multilayer Perceptron Classification Confusion Matrix

Class imbalanced - Multilayer Perceptron			
a	b	Total	Classified as
1,320	22	1,342	a = yes
89	339	428	b = no
1,409	361	1,770	
Class balanced - Multilayer Perceptron			Classified as
867.19	17.81	885	a = yes
51.69	833.31	885	b = no
918.88	851.12	1,770	

When connecting the results of the present study with the advantages of real-time cargo tracking, we can build rules for reducing logistics operations costs. We might also manage expenses, minimize product value loss, and maximizing the total value of items. Other advantages are to contribute to decision-making related to distribution and storage planning, finding possible quality failures, and identifying risks for real-time action to mitigate such risks [3,6,24].

The cold chain refers to the transportation of perishable products, which are temperature sensitive throughout a supply chain through thermal and refrigerated packaging methods, and planning logistics to protect the integrity of shipments [1-2,25,26]. The

products can be damaged in the complex transport operations; the strategic use of environmental sensors in transportation and storage is essential. Damage can be mainly by shock or temperature variations. Perishable products maintain chemical reactions whose rate can be attenuated by low temperatures. However, delays and poor planning can have negative consequences or reduce the efficiency of the shipping shipment [27,28].

## 4 Conclusions

The Multilayer Perceptron algorithm integrated with performance metrics determined better results with superior stability to solve problems and predict temperature excursion within the optimal standard of transport specification regardless of the use of class balancing filter.

**Acknowledgment.** The first author wishes to thank the Coordination of Superior Studies (Capes) for the scholarship.

## References

1. Campos Y, Villa JL (2018) Technologies applied in the monitoring and control of the temperature in the Cold Chain. In: 2<sup>nd</sup> IEEE Colombian Conference on Robotics and Automation (CCRA) IEEE Press, New York pp. 1-6 doi:10.1109/CCRA.2018.8588118
2. Rodrigue JP (2016) The cold chain and its logistics. The Geography of Transport Systems. New York: Routledge. Taylor & Francis.
3. Bogataj M, Bogataj L, Vodopivec R (2005) Stability of perishable goods in cold logistic chains. *Int J Prod Econ* 93: 345-356.
4. Ammann C (2011) Stability studies needed to define the handling and transport conditions of sensitive pharmaceutical or biotechnological products. *AAPS PharmSciTech* 12: 1264-1275.
5. Chen KY, Shaw YC (2011). Applying backpropagation network to cold chain temperature monitoring. *Adv Eng Inform* 25: 11-22.
6. Lee H, Jo SK, Lee N, Lee HW (2016) A method for co-existing heterogeneous IoT environments based on compressive sensing. In: 18<sup>th</sup> International Conference on Advanced Communication Technology (ICACT) IEEE Press, New York pp. 206-209 doi:10.1109/ICACT.2016.7423330
7. Kim K, Kim H, Kim SK, Jung JY (2016) i-RM: An intelligent risk management framework for context-aware ubiquitous cold chain logistics. *Expert Syst Appl*, 46: 463-473 doi.org/10.1016/j.eswa.2015.11.005.
8. Sharma S, Pai S (2015) Analysis of operating effectiveness of a cold chain model using Bayesian networks. *Bus Process Manag J* 21: 722-742 doi:10.1108/BPMJ-10-2014-0105
9. Ou CM, Tu JF (2018) The WSN and 3G/NFC embedded into IoV (Internet-of-Vehicle) fulfill cold chain logistics. *Microsyst Technol* 24: 3977-3983 doi:10.1007/s00542-017-3594-3
10. Öcal M, Kaya İA (2015) Food safety and GIS applications. In: 4<sup>th</sup> International Conference on Agro-Geoinformatics pp 85-90.



11. Li CM, Nien CC, Liao JL, Tseng YC (2012) Development of wireless sensor module and network for temperature monitoring in cold chain logistics. In 2012 IEEE International Conference on Wireless Information Technology and Systems (ICWITS) (pp. 1-4). IEEE.
12. Lippi G, Lima-Oliveira G, Nazer SC, Moreira MLL, Souza RFM, Salvagno GL, Montagnana M, Scartezini M, Picheth G, Guidi GC (2011) Suitability of a transport box for blood sample shipment over a long period. *Clinical biochemistry*, 44(12), 1028-1029.
13. Kulkarni NS, Niranjana S (2013) Multi-echelon network optimization of pharmaceutical cold chains: a simulation study. In: Proceedings of the 2013 winter simulation conference: simulation: making decisions in a complex world. IEEE Press, p. 3486-3498, 2013.
14. The US Pharmacopeia. 2018. Packaging and storage requirements. Rockville. 6801.
15. Larose DT, Larose CD (2015) Wiley Series on Methods and Applications in Data Mining. Data mining and predictive analytics. John Wiley & Sons. Hoboken, New Jersey, USA.
16. Frank E, Hall M, Holmes G, Kirkby R, Pfahringer B, Witten IH, Trigg L (2009) Weka-a machine learning workbench for data mining. In: Data mining and knowledge discovery handbook (pp. 1269-1277). Springer, Boston, MA.
17. Ngai EWT, Xiu L, Chau DCK (2009) Application of data mining techniques in customer relationship management: a literature review and classification. *Expert Systems with Application*, 36(2), 2592-2602.
18. Naik A, Samant L (2016) Correlation Review of Classification Algorithm Using Data Mining Tool: Weka, Rapidminer, Tanagra, Orange and Knime. *Procedia Computer Science*, 85, 662-668.
19. Holmes G, Donkin A, Witten IH (1994) Weka: A machine learning workbench.
20. Witten IH, Frank E, Trigg LE, Hall MA, Holmes G, Cunningham SJ (1999) Weka: Practical machine learning tools and techniques with Java implementations. LNCS Homepage, <http://www.springer.com/lncs>, last accessed 2016/11/21.
21. Markov Z, Russell I (2006) An introduction to the Weka data mining system. In *ACM SIGCSE Bulletin*, 38 (3) 367-368.
22. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1), 10-18.
23. Bowes D, Hall T, Gray D (2012) Comparing the performance of fault prediction models which report multiple performance measures: recomputing the confusion matrix. In Proceedings of the 8<sup>th</sup> Int. Conf. on Predictive Models in Software Engineering (pp. 109-118).
24. Gessner GH, Volonino L, Fish LA (2007) One-up, one-back ERM in the food supply chain. *Information Systems Management*, 24(3), 213-222. doi:10.1080/10580530701404561
25. Uthayakumkar R, Priyan S (2013) Pharmaceutical supply chain and inventory management strategies: optimization for a pharmaceutical company and a hospital. *Operations Research for Health Care*, 2(3), 52-64.
26. Raab V, Petersen B, Kreyenschmidt J (2011). Temperature monitoring in meat supply chains, *British Food Journal*, Vol. 113 No. 10, pp. 1267-1289.
27. Kumar N, Jha A (2017) Temperature excursion management: A novel approach of quality system in pharmaceutical industry. *Saudi Pharmaceutical Journal*, 25(2), 176-183.
- Montanari R (2008) Cold chain tracking: a managerial perspective. *Trends in Food Science Technology*, 19(8), 425-431.