Predicting the Tear Strength of Woven Fabrics Via Automated MachineLearning: An Application of the CRISP-DM Methodology

Rui Ribeiro [1,2], André Pilastri [1], Carla Moura [3], Filipe Rodrigues [4], Rita Rocha [4] and Paulo Cortez [2]

[1] EPMQ - IT Engineering Maturity and Quality Lab, CCG ZGDV Institute, Guimarães, Portugal
[2] ALGORITMI Centre, Dep. Information Systems, University of Minho, Guimarães, Portugal
[3] Riopele, Pousada de Saramagos, Portugal
[4] CITEVE - Centro Tecnológico das Indústrias Têxtil e do Vestuário de Portugal, Famalicão, Portugal

Keywords: Fabrics, Tear Strength, Industry 4.0, Regression, Automated Machine Learning.

Abstract: Textile and clothing is an important world industry that is currently being transformed by the adoption of the Industry 4.0 concept. In this paper, we use Data Mining (DM) technology and the CRoss-Industry Standard Process for DM (CRISP-DM) methodology to model the textile testing process, which assures that products are safe and comply with regulations and client needs. Real-world data were collected from a Portuguese textile company, which has the goal to reduce the number of attempts they take in order to produce a woven fabric. Thus, predicting the outcome of a given test is beneficial to the company because it can reduce the number of physical samples that are needed to be produced when designing new fabrics. In particular, we target two important textile regression tasks: the tear strength in warp and weft directions. To better focus on feature engineering and data transformations, we adopt an Automated Machine Learning (AutoML) during the modeling stage of the CRISP-DM. Several iterations of the CRISP-DM methodology were employed, using different data preprocessing procedures (e.g., removal of outliers). The best predictive models were achieved after 2 (for warp) and 3 (for weft) CRISP-DM iterations.

1 INTRODUCTION

The textile and clothing industry is one of the largest industrial sectors in the world (Shishoo, 2012). However, in a highly competitive market as the textile one, companies must find a way to differentiate themselves from the competitors, thus are compelled to improve their production processes, improving the quality of the products and/or lowering the production prices. Under this context, the Industry 4.0 concept, which assumes the digitalization of the productive processes (e.g., digital sensors with connectivity capabilities) and a stronger usage of Information Technology (Lasi et al., 2014) is thus an opportunity for the textile industry enhancement.

In this industry, in order to create the final product, the raw materials undergo a series of processes, where fibers are combined into yarns and the combination of these yarns creates a fabric, which receives a series of textile treatments, creating the final product that is delivered to costumers. Automation technologies can be utilized in several parts of this productive process. In particular, a large amount of data is created and stored, such as the properties of each yarn (e.g., color, thickness), the configuration of each machine used in the creation process (e.g., spinning, weaving) (Mozafary and Payvandy, 2014), and the results of the specific tests that the company executes. All these data can be processed by Data Mining (DM) and Machine Learning (ML) methods, allowing the discovery of valuable knowledge in order to improve the textile manufacturing process (Yildirim et al., 2018).

This paper presents an implementation of a DM project using Automated ML (AutoML) and the CRoss-Industry Standard Process for DM (CRISP-DM) methodology (Wirth and Hipp, 2000) to predict the results of tear strength test (warp and weft directions) on fabrics. The data were collected from a Portuguese textile company, aiming to reduce the number of attempts required to produce a fabric that complies with the industry standards and fulfills the client requirements. The paper is structured as follows. Section 2 introduces the concepts of fabric testing, DM and AutoML. Then, Section 3 presents the data and methods used, including the three CRISP-DM iterations. The obtained results are discussed in Section 4. Finally, Section 5 presents the main conclusions and future work.

2 BACKGROUND

2.1 Fabric Testing

Every time a textile company creates a new woven fabric, it will typically execute a series a tests. These tests have a crucial role in evaluating the product quality (Hu, 2008). The International Organization for Standardization (ISO) has launched several standards related to fabric testing for a whole series of tests that examine the physical, mechanical and chemical properties of fabrics. Some tests can be made to test two different aspects, such as the tensile strength in warp and weft direction (Dimitrovski et al., 2004) or one aspect, such as pilling (Beltran et al., 2006).

Each time a new fabric is created, the tests are made using a sample that the company must produce. The sample is then analyzed and several results are possible: the sample can pass all the tests and met the requirements, thus the fabric is read for massive production; the sample did not met the client requirements, so the developer must start again the fabric design and change some of the characteristics; or the fabric did not pass the standard tests and it must be verified if the flaw is in conception phase or in the production phase. This process can be repeated several times until all the requirements are made, resulting in additional time and costs. It should be noted that the loom that manufactures the fabric sample needs to produce a minimum of several meters of a fabric in each design attempt. Thus, the dematerialization of this process, by means of a DM predictive modeling, can potentially improve the analysis of fabric design parametrization changes and reduce the number of physical fabric sample productions, saving time and costs.

2.2 CRISP-DM and AutoML

The CRISP-DM is a open analytic process standard for increasing the success of DM projects. The methodology is based on a hierarchical process model, described at four levels of abstraction: phase, generic task, specialized task, and process instance (Wirth and Hipp, 2000). Overall, CRISP-DM provides an overview to the life cycle of a data mining project, with iterations of several phase sequences, as shown in Figure 1. The iterative execution of the methodology also assumes an interaction between the business experts and the DM analysts.

During the Modeling phase of CRISP-DM, Machine Learning (ML) algorithms are often used to extract valuable knowledge from the data. Due to the relevance of ML, several algorithms have been proposed is the last decades, each one presenting its advantages. Examples of popular regression algorithms include (Witten et al., 2016): Regression Trees, Linear Regression, Generalized Linear Models, Support Vector Machines, Ensembles (including Boosting and Random Forest) and Neural Networks (including Deep Learning).

In practice, the ML model creation process tends to involve a highly iterative exploratory process. In this sense, an effective ML modeling process requires

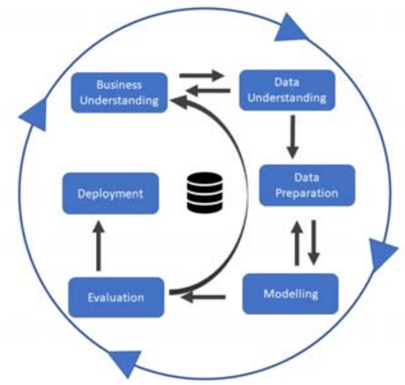


Figure 1: Phases of the CRISP-DM model, adapted from (Wirth and Hipp, 2000).

solid knowledge and understanding of the different types of ML algorithms and their hyperparameter adjustment (Maher and Sakr, 2019). In effect, the selection of the best ML algorithm is often performed using a trial-and-error procedure, which can be guided by the analyst expert knowledge or heuristics (Gibert et al., 2018). Such iterative and explorative nature of the modeling process is commonly tedious and time-consuming. Moreover, the quality of the ML results is also dependent of data and feature engineering aspects (e.g., feature selection, outlier detection) (Domingos, 2012) that are typically performed on the Data Understanding and Data Preparation CRISP-DM stages (Gibert et al., 2016).

To focus on these aspects, in this work we use Automated Machine Learning (AutoML) (Feurer et al., 2015), during the Modeling stage of CRISP-DM. AutoML systems are developed to automate this challenging and time-consuming process. These intelligent systems increase the accessibility and scalability of various ML applications by efficiently solving an optimization problem to discover pipelines that yield satisfactory outcomes, such as prediction accuracy (Le et al., 2019). Therefore, AutoML allows the DM analysts to focus their effort in applying their expertise in other important components, such as feature and data engineering, model validation and deployment.

2.3 Data Mining Applied to Fabrics

The process of fabric manufacturing generates large amounts of data. DM techniques started being used in textile engineering during recent years, aiming to solve the difficulties of classical mathematical and statistics in modeling the complex relationships present in the data. Most DM applications to the textile industry involve classification tasks, such as quality control (e.g., textile image inspection) (Yildirim et al., 2018). The application of DM to test areas is more scarce, in particular regarding the prediction of tear strength.

The tear strength is usually a measure of the force (tensile stress) required to propagate a tear and is often used to give a direct assessment of the serviceability of the fabric (Teli et al., 2008). Tear strength can be tested in both warp and weft directions and it is considered one of the most important performance attributes of woven textiles (Malik et al., 2011). Most of the tear strength prediction studies employ linear models, which are rather rigid and thus fail when nonlinear relationships exist among the data attributes. In (Kotb, 2009), linear regression models were used to predict the fabric tearing force based on 9 identified input features, concluding that tearing force is largely affected by the type and number of weft yarns, weft density, ground structure, and ground yarns, while the shape of the pile and the change in pile designation have minor effects. In another study, the linear regression was also used to predict the fabric tear strength in warp and weft direction for woven wool fabrics, obtaining a Pearson correlation between the actual and the predicted strength for warp and weft of 0.976 and 0.975, respectively (Malik et al., 2011). The same linear regression model was used in (Eltayib et al., 2016) to predict the relationship between fabric tear strength and other independent variables, such as yarn tensile strength, yarn count and fabric linear density. In (Zeydan, 2010), a flexible nonlinear model, based on an Evolutionary Artificial Neural Network, was proposed to predict the tensile strength in a woven fabric, outperforming a linear regression model.

In this paper, we use recent data, collected by a Portuguese textile company, aiming to predict the tear strength test, at both warp and weft directions, of fabrics. Within our knowledge, this is the first textile industry study that employed an AutoML procedure, which automatically tested five families of flexible regression algorithms during the Modeling stage of CRISP-DM. Such an automatic selection of the best ML method allowed us to perform more quickly different CRISP-DM iterations (described in Sections 3.2, 3.3 and 3.4), after obtaining feedback from the textile company and aiming to explore different data and feature engineering approaches.

3 MATERIALS AND METHODS

In this paper, we use recent data, collected by a Portuguese textile company, aiming to predict the tear strength test, at both warp and weft directions, of fabrics. This Portuguese textile company creates and produces fabrics for fashion and clothing collections from diverse customers. The current fabric design is based on the designer experience and intuition and several trial-and-error fabric sample production experiments. When designing new fabrics, in order to meet the requirements of the client, the company produces several small sample attempts. In each attempt, several laboratory tests are used to verify if the fabric complies with quality goals. If this sample is not approved, the design process must be repeated, which is translated into more time and costs for the company. The fabric design process generates data that is related to the several components of the fabric, as well as the quality test results.

Within our knowledge, this is the first textile industry study that employed an AutoML procedure, which automatically tested five families of flexible regression algorithms during the Modeling stage of CRISP-DM. Such an automatic selection of the best ML method allowed us to perform more quickly different CRISP-DM iterations (described in Sections 3.2, 3.3 and 3.4), after obtaining feedback from the textile company and aiming to explore different data and feature engineering approaches.

3.1 Computational Environment

All executed experiments were conducted in two different open source computational environments: the R statistical tool and its rminer package, for data manipulation and ML result analysis, including the computation of the regression metrics and regression error characteristic (REC) curves (Cortez, 2010); and H2O, which implements an easy to use AutoML algorithm (Landry et al., 2018). The AutoML was configured to automatically select the regression model and its hyperparameters based on the best Mean Absolute Error (MAE) over a validation set, using a 10-fold cross-validation that is applied over the whole training data. A total of five different regression families were automatically compared by the AutoML. These include three individual base learners, namely Generalized Linear Models (GLM), Gradient Boosting Machines (GBM) and distributed Random Forests (XRF), and two stacking ensembles, one using all trained models (Stacking All) and other using just the best model per ML base algorithm (Stacking Best).

3.2 First CRISP-DM Iteration

In this iteration, we performed the first five phases of the CRISP-DM, Business Understanding to Evaluation, aiming to predict the two fabric tear strength targets (warp and weft).

3.2.1 Business Understanding

During this phase, the textile company expressed the need to reduce the number of attempts that were necessary to produce a fabric sample. The two fabric tear strength numeric attributes (warp and weft directions) were also identified as relevant prediction targets, thus setting two regression tasks. We also selected the computational tools (R and H2O), as detailed in Section 3.1.

3.2.2 Data Understanding

First, we analyzed the textile company two main fabric data sources: the Enterprise Resource Planning (ERP), which included the 88,653 fabric main data records, and the laboratory testing database, which contained the fabric quality tests performed between February 2012 to March 2019. After merging the two data sources, the resulting dataset had 12,088 examples for the warp test and 12,143 for weft. Table 1 summarizes the initial set of input attributes, as suggested by the textile company. Most attributes are numeric and the exceptions are the type of fabric and yarn code. The last 5 rows are related with yarn attributes. We note that the each fabric can include several types of yarns, which is a relevant issue that is handled in Section 3.2.3. Figure 2 shows the box plot distribution of the fabric weft and warp tear strength.

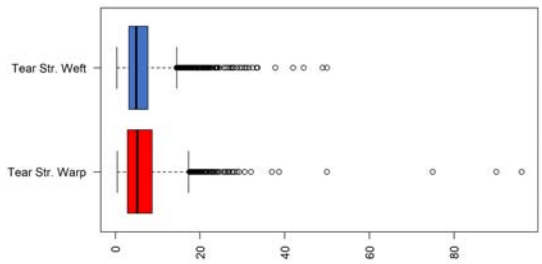


Figure 2: Box plot of the test targets used in First CRISP-DM iteration.

3.2.3 Data Preparation

A Data Warehouse system was implemented, in which an Extraction, Transform, Load (ETL) process was used to merge the ERP and laboratory test databases and preprocessing some data records. The preprocessing included the removal of fabric records with missing components (e.g., with not registered yarns). Also, in some cases it was detected that the same fabric had different quality test values, related with repeated tests conducted at different fabric production stages. In order to have a single test value per fabric, the distinct test values for the same fabric were averaged. The resulting preprocessed data included 8,453 observations for the warp test and 8,423 examples for the weft tear strength.

Each fabric can include several types of yarns. In this work, we propose a novel input combination of features in which we include the sequence of all possible yarns (up to 9 in our dataset), for both warp and weft. Since each yarn is represented by 5 features (Table 1), the regression models are fed with $12+9\times5\times2$ (warp and weft)=102 input variables. A zero padding (i.e., addition of zero values to missing elements) was performed on all fabrics that had less than 18 yarn codes. Finally, before feeding the data to the ML algorithms, the numeric input attributes were standardized to a zero mean and one standard deviation, while the nominal variables were transformed using the one-hot binary encoding, which sets one binary variable per possible level.

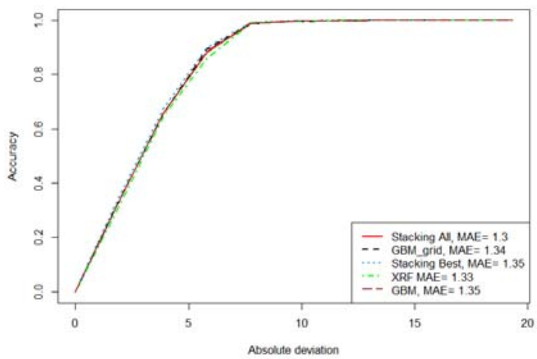
3.2.4 Modeling

To evaluate the predictive models, an external holdout split was executed, in which the data was randomly divided into training (75%) and test (25%) data. The quality of the predictions is measured by using (Cortez, 2010; Witten et al., 2016): the Mean Absolute Error (MAE), Adjusted R2 (Adj. R2) and classification Tolerance. For MAE, the lower the values, the better are the predictions. Regarding Adj. R2 and Tolerance, higher the values indicate better predictions. Adj. R2, known as the adjusted coefficient of determination, is often used in multiple linear regression and it ranges from 0 to 1. The Tolerance value is based on the REC analysis and it measures the percentage of correctly classified examples when assuming a fixed absolute error tolerance (Bi and Bennett, 2003). In this paper, three tolerance values were set: 5%, 10% and 20%. We note that the percentage of error tolerance is computed by considering the range of the true values.

Using only training data, the AutoML procedure was applied, as described in Section 3.1. Figure 3 shows the REC curves and respective MAE values, computed using validation data, for the best five ML algorithms that were obtained when using the internal 10-fold procedure for the warp and weft tear strength predictions. The REC curve shows the error tolerance on the x-axis versus the percentage of correctly predicted points within the tolerance on the y-axis. In all AutoML experiments conducted in this study, and for both targets, the selected ML algorithm was a stacked ensemble that used all trained models (Stacking All).

Table 1: List of input attributes used for regression.

Name	**Description (data type)**	**Min.**	**Max.**	**Average**
T_cm	Number of finished threads per centimeter (numeric)	18	1,321	115.60
P_cm	Number of finished picks per centimeter (numeric)	7	510	88.87
weight/m^2	Weight (in grams) per square meter (numeric)	22	1,690	241.70
finished width	Width in centimeters (numeric)	90	168	140.00
weave design	Weave pattern of the fabric (nominal with 21 levels)	-	-	-
reed width	Width of the reed in centimeters (numeric)	30	242	188.80
denting	Number of the reed dents per centimeter (numeric)	0	252	126.80
ends/dent	Number of yarns per dent (numeric)	0	88	2.30
n_picks	Number of picks on loom per centimeter (numeric)	0	81	16.50
weft code	Identification code of the weft (nominal with 6,883 levels)	-	-	-
warp code	Identification code of the warp (nominal with 5,353 levels)	-	-	-
warp total ends	Total number of threads on the warp (numeric)	477	21,858	6,950.00
yarn code	Identification code of the yarn (nominal with 11,020 levels)	-	-	-
n_folds	Number of single yarns twisted (numeric)	1	12	1.60
yarn count	Mass per unit length of the yarn	2	268	47.70
yarn usage	If the yarn is used in warp or weft (binary)	-	-	-
yarn repetitions	Number of yarn repetitions in warp or weft	1	8	1.42



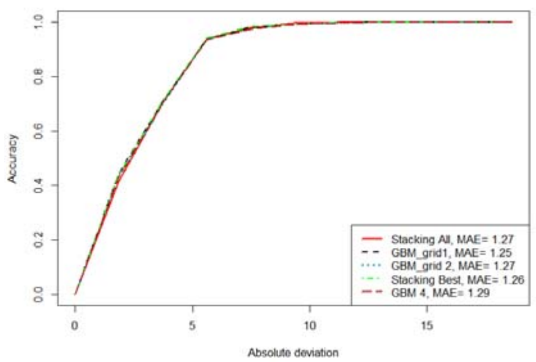


Figure 3: AutoML validation REC curves for the warp (top) and weft (bottom) tear strengths.

3.2.5 Evaluation

The obtained predictions, computed over the 25% test data, are shown in Figure 4, in terms of the predicted (y-axis) versus real (x-axis) values. The scatter plots show an interesting initial fit, with most points being close to the perfect prediction, represented by the red diagonal line, although there are high errors, particularly when the real target values increase. This behavior alerted the textile company experts for the need to discard outliers, which was addressed in the second CRISP-DM iteration.

3.3 Second CRISP-DM Iteration

In order to improve the previous results, a new iteration of CRISP-DM was defined. During a new Business Understanding phase execution, the textile company provided a list business normal ranges for the tear strength values. Thus, all test values that were inferior to 0.4 and superior to 25 were discarded in a new Data Preprocessing stage, since these tests were considered outliers (e.g., related with special uncommon military fabrics). The resulting dataset included 8,431 observations for the warp shear strength target and 8,399 examples for weft one. Figure 5 presents the box plot values of the two analyzed targets. The Modeling and Evaluation phases were then executed, similarly to what is detailed in Section 3.2. When the predictive results (presented in Section 4) were shown to the textile company, they were considered more satisfactory. However, the textile experts suggested a new CRISP-DM iteration, which would test the im-

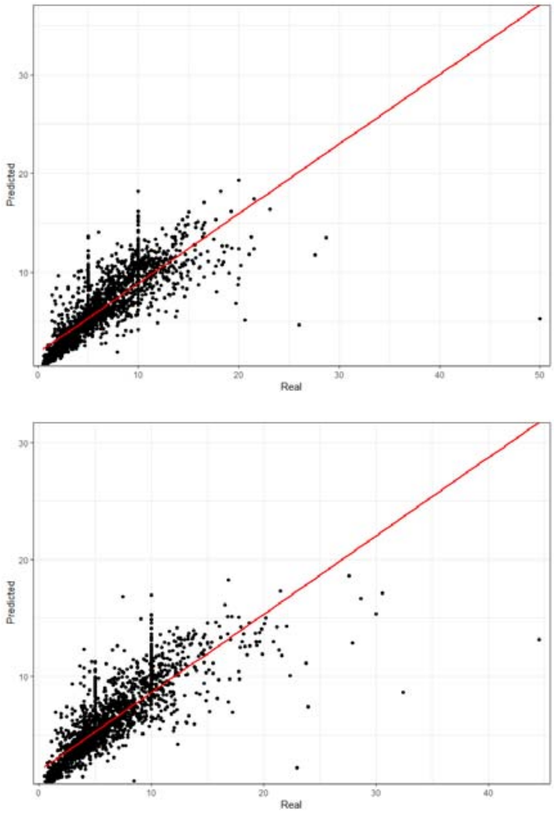


Figure 4: Regression scatter plot for the first CRISP-DM iteration warp (top) and weft (bottom) tear strength predictions.

pact of using the overall composition of the fabric as an useful and extra input element.

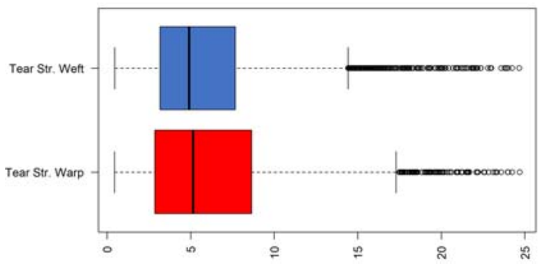


Figure 5: Box plot of the test targets used in Second CRISP-DM iteration.

3.4 Third CRISP-DM Iteration

In the third CRISP-DM iteration, we tested if addition of the final composition of the fabric (e.g., overall percentage of cotton and polyester), as an extra input feature, could improve the quality of the tear strength predictions. During a new Data Understanding stage, we collected the final composition attribute. The attribute, which contained 1,164 distinct levels, was treated as nominal, being thus preprocessed using the one-hot transform. The remaining CRISP-DM iteration was executed similarly to the second CRISP-DM iteration (e.g., with outlier removal), except that the predictive models used a total of 103 input variables (and not 102).

4 RESULTS

Table 2 presents the overall predictive results for the test data (25%) and the three CRISP-DM iterations. For comparison purposes, we also tested a baseline method that is equivalent to the first CRISP-DM iteration except that it uses the classical multiple linear regression model, as implemented in the rminer R package (Cortez, 2010). In all three CRISP-DM iterations, and as previously explained, the AutoML selected model was the ensemble that included all searched AutoML models (Stacking All).

The analysis of the tear strength warp results shows an improvement from the first to second CRISP-DM iteration but not from the second to the third one. In effect, the best predictive results (for all regression metrics) were achieved during the second CRISP-DM iteration, showing that outlier removal is beneficial when predicting the warp test, although there is no gain in including the final fabric composition as an input variable. Regarding the tear strength weft, the results confirms the progress of the CRISP-DM iterations, where each iteration resulted in a lower MAE value. Also, the Adj. R^2 values improved in a similar way. Thus, the best prediction results were obtained in the third iteration, which also corresponds to the best classification Tolerance for all 5%, 10% and 20% values. This confirms that removing outliers and using the final fabric composition is valuable for improving the weft quality predictions.

As for the baseline results, they are clearly worst when compared to the AutoML method and for both prediction goals, confirming that the regression tasks are nonlinear. The MAE differences are higher when comparing the linear method with the AutoML results for the first CRISP-DM iteration than when comparing different AutoML CRISP-DM iterations (e.g., the differences are 0.65, 0.08 and 0.02 for the weft test), which clearly backs the AutoML as an interesting modeling method.

To complement this analysis, Figure 6 plots the REC curves for the predictive models from Table 2. The plots include also the Normalized Regression Error Characteristic (NAREC) value for each curve (the higher, the better). The REC curves confirm the best

Table 2: Overall predictive results for the test data (best values in **bold**)

Test	Iteration	Target Interval	Regression Metrics				
			MAE	Tol. 5%	Tol. 10%	Tol. 20%	Adj. R^2
Tear warp	Baseline	[0.52,50.00]	2.06	2%	4%	7%	0.44
	1st	[0.52,50.00]	1.30	6%	12%	23%	0.68
	2nd	[0.44,24.00]	**0.70**	**8**%	**16**%	**29**%	**0.92**
	3rd	[0.44,24.00]	1.20	6%	14%	23%	0.75
Tear weft	Baseline	[0.50,44.52]	1.92	2%	4%	8%	0.49
	1st	[0.50,44.52]	1.27	5%	**12**%	20%	0.69
	2nd	[0.56,24.25]	1.18	5%	11%	22%	0.71
	3rd	[0.56,24.25]	**1.16**	**6**%	**12**%	**21**%	**0.72**

performance of the third CRISP-DM iteration model for tear weft and second CRISP-DM iteration model for tear warp. The quality of the best model predictions can be visualized in Figure 7. The regression scatter plots show that the predictions are more closer to the real values when compared with the scatter plots of the first CRISP-DM iteration (Figure 4). In particular, a high quality regression was achieved for the warp tear strength prediction (top of Figure 7). These regression results were shown to the textile company experts, which provided a positive feedback.

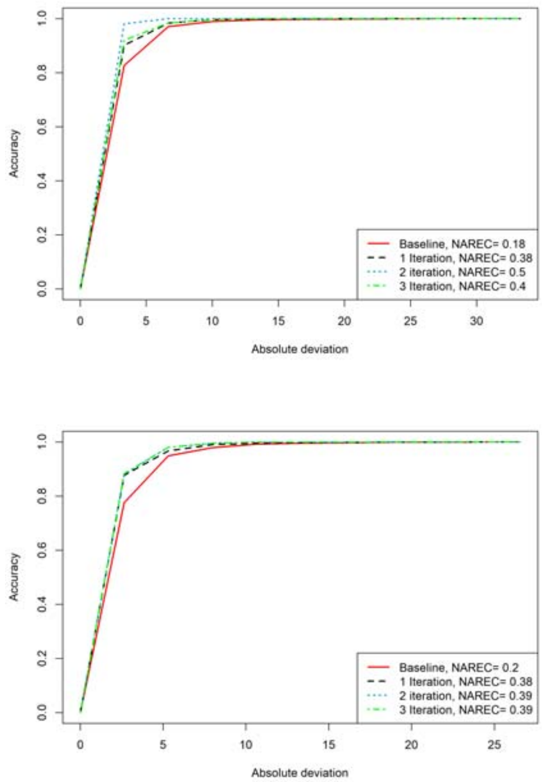


Figure 6: REC curves for the warp (top) and weft (bottom) tear strength prediction models.

5 CONCLUSIONS

In this paper, a DM approach guided by the CRISP-DM methodology was used to predict the results of fabric tear strength tests in warp and weft directions. A total of three CRISP-DM iterations were executed, aiming to explore distinct data preprocessing operations (e.g., outlier removal, inclusion of additional inputs) for the two regression tasks. The data were collected from a Portuguese textile company and it included an initial database of thousands of fabric records from 2012 to 2019. A Data Warehouse was created, allowing to clean and merge these records with the laboratory test data, resulting in a warp and weft test datasets with around 8,400 examples. During the Modeling stage of CRISP-DM, an AutoML was adopted, automatically tuning and selecting the best ML model for a particular dataset. The AutoML tool always selected a stacking ensemble that included all tested ML models and that obtained much better regression results when compared with a linear regression baseline model.

The best warp tear strength predictions were achieved at the second CRISP-DM iteration, which involved outlier removal, while the best weft test results were obtained at the third CRISP-DM iteration, which included the final fabric composition as an extra input feature. The best predictive results were analyzed as valuable by the textile company experts. In future work, we intend to apply a similar approach in the prediction of other fabric quality tests, such as pilling. Moreover, we intend to apply the best prediction models in a real textile environment, aiming to reduce the number of fabric sample creation attempts.

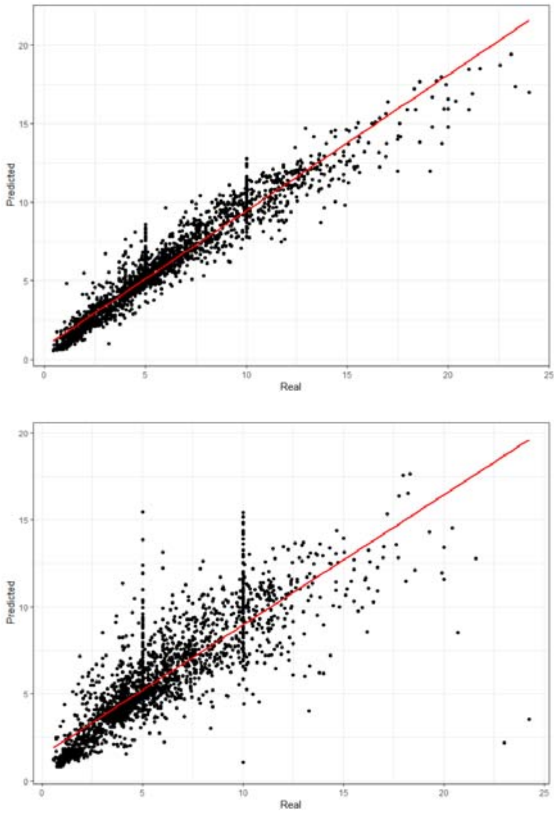


Figure 7: Regression scatter plot of best models for the warp (top) and weft (bottom) tear strength predictions.

REFERENCES

Beltran, R., Wang, L., and Wang, X. (2006). Predicting the pilling tendency of wool knits. *The Journal of The Textile Institute*, 97(2):129–136.

Bi, J. and Bennett, K. P. (2003). Regression error characteristic curves. In *Proceedings of the 20th international conference on machine learning (ICML-03)*, pages 43–50.

Cortez, P. (2010). Data mining with neural networks and support vector machines using the r/rminer tool. In *Industrial Conference on Data Mining*, pages 572–583. Springer.

Dimitrovski, K., Gabrijelčić, H., Kovačević, S., and Nikolić, M. (2004). The influence of weft yarn characteristics on tensile strength of woven fabrics in warp direction. In *Magic World of Textiles*.

Domingos, P. M. (2012). A few useful things to know about machine learning. *Commun. acm*, 55(10):78–87.

Eltayib, H. E., Ali, A. H., and Ishag, I. A. (2016). The prediction of tear strength of plain weave fabric using linear regression models. *International Journal of Advanced Engineering Research and Science*, 3(11).

Feurer, M., Klein, A., Eggensperger, K., Springenberg, J., Blum, M., and Hutter, F. (2015). Efficient and robust automated machine learning. In *Advances in neural information processing systems*, pages 2962–2970.

Gibert, K., Izquierdo, J., Sànchez-Marrè, M., Hamilton, S. H., Rodríguez-Roda, I., and Holmes, G. (2018). Which method to use? an assessment of data mining methods in environmental data science. *Environmental modelling & software*, 110:3–27.

Gibert, K., Sànchez-Marrè, M., and Izquierdo, J. (2016). A survey on pre-processing techniques: Relevant issues in the context of environmental data mining. *AI Communications*, 29(6):627–663.

Hu, J. (2008). *Fabric testing*. Elsevier.

Kotb, N. (2009). Engineering of tearing strength for pile fabrics. *Journal of Textile and Apparel, Technology and Management*, 6(1).

Landry, M., Bartz, A., Aiello, S., Eckstrand, E., Fu, A., and Aboyoun, P. (2018). Machine Learning with R and H2O: Seventh Edition. Technical Report September.

Lasi, H., Fettke, P., Kemper, H.-G., Feld, T., and Hoffmann, M. (2014). Industry 4.0. *Business & information systems engineering*, 6(4):239–242.

Le, T. T., Fu, W., and Moore, J. H. (2019). Scaling tree-based automated machine learning to biomedical big data with a feature set selector. *Bioinformatics*. btz470.

Maher, M. M. M. Z. A. and Sakr, S. (2019). SmartML: A Meta Learning-Based Framework for Automated Selection and Hyperparameter Tuning for Machine Learning Algorithms. In *EDBT: 22nd International Conference on Extending Database Technology*, Lisbon, Portugal.

Malik, Z. A., Malik, M. H., Hussain, T., and Arain, F. A. (2011). Development of models to predict tensile strength of cotton woven fabrics. *Journal of engineered fibers and fabrics*, 6(4):155892501100600407.

Mozafary, V. and Payvandy, P. (2014). Application of data mining technique in predicting worsted spun yarn quality. *The Journal of The Textile Institute*, 105(1):100–108.

Shishoo, R. (2012). *The global textile and clothing industry: technological advances and future challenges*. Elsevier.

Teli, M., Khare, A., and Chakrabarti, R. (2008). Dependence of yarn and fabric strength on the structural parameters. *AUTEX Research Journal*, 8(3):63–67.

Wirth, R. and Hipp, J. (2000). Crisp-dm: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, pages 29–39. Citeseer.

Witten, I. H., Frank, E., Hall, M. A., and Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Yildirim, P., Birant, D., and Alpyildiz, T. (2018). Data mining and machine learning in textile industry. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(1).

Zeydan, M. (2010). Prediction of fabric tensile strength by modelling the woven fabric. *Woven Fabric Engineering*, page 155.