ACTA · GRAPHICA
ISSN: 0353-4707
e-ISSN: 1848-3828
ZNANSTVENI ČASOPIS ZA TISKARSTVO I GRAFIČKE KOMUNIKACIJE
JOURNAL FOR PRINTING SCIENCE AND GRAPHIC COMMUNICATIONS

Scientific article

Machine learning approach to predict the dot gain of flexographic prints

Soumen Basak, Saritha P.C, Alenrex Maity, Kanai Chandra Paul

Department of Printing Engineering, Jadavpur University, Salt Lake Campus, B-73-80, Plot No. 8, Salt Lake By-Pass, LB Block, Sector-III, Bidhannagar,Kolkata – 700106, West Bengal, India.
Correspondence: soumen.basak@jadavpuruniversity

Abstract:

The computational learning theory and dynamic programming approach of machine learning techniques made it astutely applicable to the various bids such as developing algorithms, understanding patterns, studying observations, forecasting data etc. The algorithms have the features to construct a framework of trained input dataset to deliver sensible and dynamic data driven predictions and judgements. These contrasting peculiarities of the machine learning techniques can be made useful with the various fields of printing technology in a revolutionized manner. Considering these aspects, this work is focussed on to integrating the machine learning possibilities with flexographic print quality evaluation. This paper established a machine learning algorithm in Python to evaluate the flexographic print quality as dot gain as the output response. The data analysis and performance evaluation are done using the supervised machine learning approaches such as linear regression, decision tree and random forest regression algorithms. The results show comparatively a better accuracy with random forest regressor technique with 93.02%. Moreover, the strategy to reduce execution time is put forth that involves using Python to implement the machine learning algorithms in this work.

Keywords:
Flexography, Dot Gain, Machine Learning, Python, Linear Regression, Decision Tree, Random Forest Regressor.

1. Introduction

1.1 Objective of the study

The machine learning is something that is encountered in our everyday lives. This paper examines the application of machine learning technique in the estimation and prediction of flexographic print quality. This aims to implement supervised machine learning techniques such as: linear regression, decision tree and random forest regression algorithms in the most efficient way possible using the Python tool. The experimental background comprises the flexographic printing on 114 tonal gradations on both coated and uncoated paper under three different anilox roller ruling.

1.2 The Flexographic Printing

The direct printing method known as flexographic printing that uses a rotary relief image carrier that is sufficiently flexible to transfer printing ink to a broad range of substrates, including paper, paperboards, plastic films, foils, etc. [1, 2]. A small amount of pressure is sufficient in flexography to move ink from the plate to the substrate at the printing nip [3, 4]. Figure 1 represents the schematic diagram of flexographic printing process.

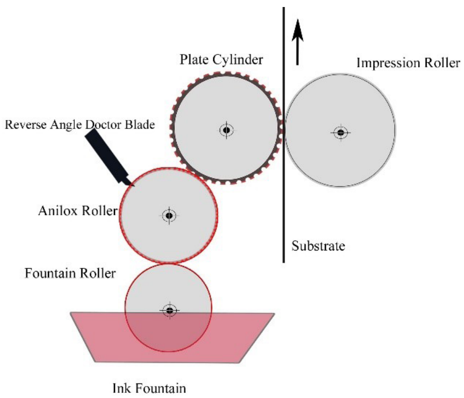

Figure 1: Schematic diagram of a flexographic printing unit

In flexographic halftone printing, it is clear that the anilox frequency at the final print affects both the lower and higher screen rules of the plate image. The finer anilox roll is needed to maintain a cell aperture smaller than the minimum reproducible plate dot. The dot becomes steadier as it gets deeper, which reduces the tone value increase [5, 6].

The amount of ink to be applied to the substrate was greatly influenced by the specifications for the anilox roller cell engraving, which in turn had an impact on the optical density and halftone reproduction of the finished print. The anilox cell capacity was the main factor that determined print quality, but other factors, such the cell's shape, were also discovered to have an impact [7, 8].

A rise in the substrate surface roughness can lead to a sudden decrease in print density while also increasing print density through smoother paper. In addition, the features of the anilox roller cell in flexography are crucial for ensuring accurate print results [9].

The paper's intrinsic tone, roughness, and porous surface all have an impact on the distribution and absorption of ink, which in turn affects print quality. The degree of paper roughness or smoothness has a direct impact on how the ink spreads throughout the print [10, 11].

The assessment of print quality is probably going to make use of quantitative studies such as dot gain, dot area, print contrast, and others. The correctness of the dot replication in the graphic reproduction is monitored using the dot gain and dot area parameters. The degree to which shadow features are accurately depicted on the final print is indicated by the print contrast value [12, 13].

1.3 Machine Learning

The ideas of machine learning have started around the era of 1950's by Alan Turing. Rather than being explicitly coded, machine learning is a collection of algorithms that are taught from data and/or experiences. A distinct set of algorithms is needed for each activity, and these algorithms detect habits to do out specific operations [14, 15].

The algorithms that make up machine learning teach computers to do things that humans do on a regular basis with ease and naturalness [16, 17]. Comparison of conventional programming with machine learning is illustrated in Figure 2.

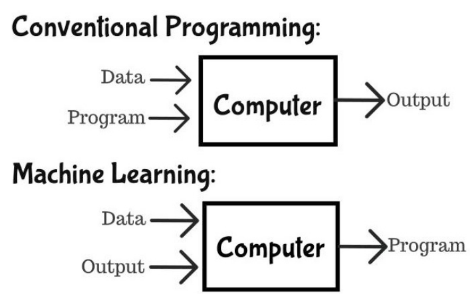

Figure 2: Conventional Programming vs Machine Learning: Machine learning as automating the process of automation, in contrast to traditional programming, which merely reflects automation.

1.3.1 Types of machine learning methods

There are several types of machine learning methods used in different fields. Classification of machine learning techniques is shown in Figure 3.

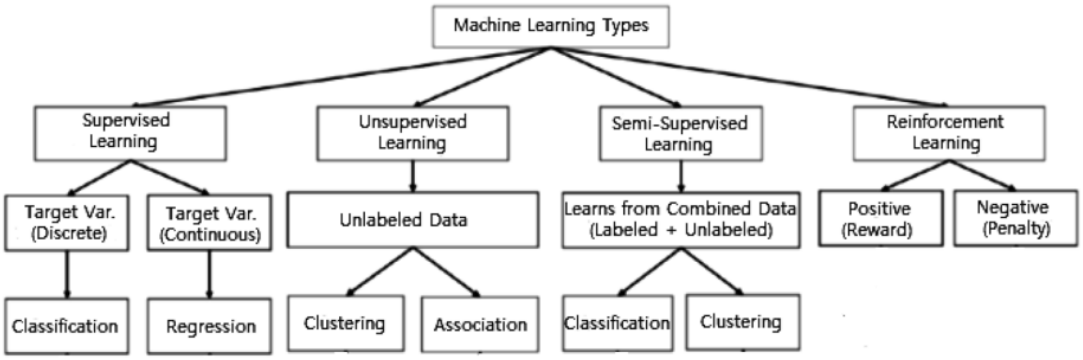

Figure 3: The classification of machine learning techniques

The major categories of machine learning methods are:

Supervised learning: It is a technique where the output is obtained by using the current input data. The machine is trained using labelled data. It takes labelled inputs and maps it to the known outputs. Machines are trained using well labelled training data and on the basis of that data the output is predicted (remember the input data, formulate features, predict) [18]. The output classification and regression supervised learning are the two categories of supervised learning [19].

Classification involves dividing the data into groups based on the features that are unique to the data collection.

Regression is the process of forecasting or drawing conclusions about the data's other features from its available features. Figure 4 shows the block diagram of supervised machine learning methods.

Types of regression:

- Linear regression
- Polynomial regression
- Support vector regression
- Decision tree regression
- Random forest regression
- Ridge Regression
- Lasso Regression
- Logistic Regression

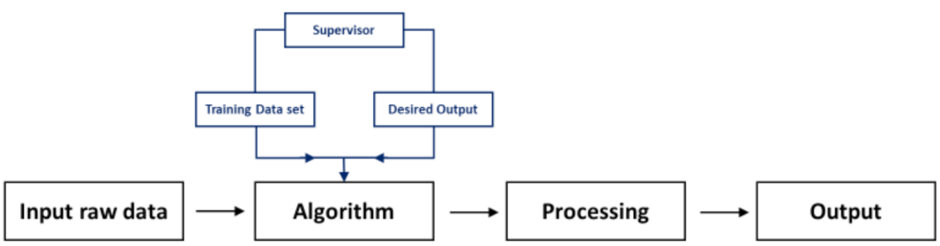

Figure 4: Supervised learning

Unsupervised Learning: This method does not provide output data. The relationships and connections among the data are used to facilitate learning. Also, there are no training data for unsupervised learning. The unlabelled data is used to train the machine. It understands patterns and trends in the data and discovers the output. Figure 5 illustrates the flow chart of unsupervised machine learning techniques. The unsupervised learning comes in two flavours: association and clustering [20].

Clustering is the process of identifying sets

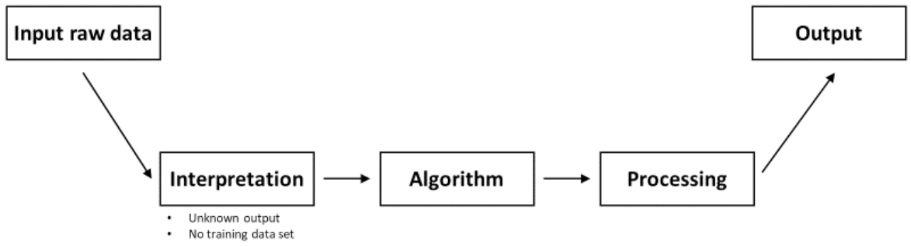

Figure 5: Unsupervised learning

of data that are similar to one another when the data's underlying classifications are unknown.

Association is the process of identifying the relationships and connections between the data within a single data set [20].

Feature extraction: It is the process of choosing a subset of features or creating new features by combining the features pertaining to a group and category of data.

Semi-supervised Learning: When the amount of labelled data is smaller than that of unlabelled data, both supervised and unsupervised learning are insufficient. In these situations, relatively little information about them is inferred from the unlabelled data. Fewer labelled data than anticipated data are present in semi-supervised learning [21, 22].

Reinforcement Learning: In this type of learning, agents pick up new skills through reward systems. While there are start and finish points, the agent's objective is to take the quickest and most accurate route to the destination. Positive reinforcement is given to the agent when they follow the right procedures. However, using the incorrect methods will have unfavourable effects. Learning happens when pursuing the objective. It uses an agent and an environment to produce actions and rewards. There is no pre-defined target variable. It follows trial and error method to arrive at the desired solution [23]. At every step, the agent gives reward. It is applicable to gaming industry, robotics etc. Figure 6 represents reinforcement machine learning techniques.

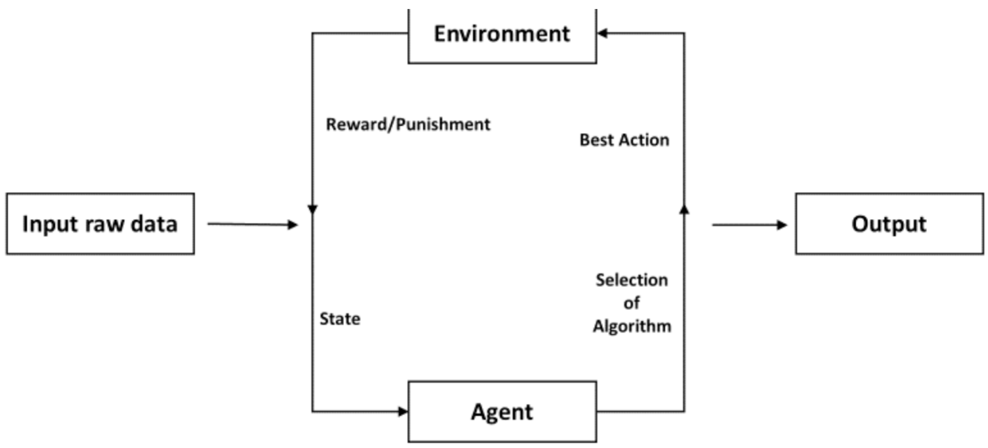

Figure 6: Reinforcement Learning

Python is regarded as a popular language for machine learning for a variety of reasons. Python is a straightforward, elegant language with a clear syntax, very easy text manipulation and processing features, and a vibrant development community that produces a wealth of documentation. Additionally, Python includes an interactive shell that lets programmers view and analyse program parts at the same time as they write code [24].

High-level data types that Python comes with, such as lists, queues, tuples, dictionaries, etc., make it simple to implement abstract ideas and eliminate the need for explicit programming on the part of the programmer. Python is becoming more and more popular across a variety of disciplines, including the scientific and financial communities. It is an open source, concise, clean, and easy to learn language with a huge number of libraries and packages available. Furthermore, even non-programmers might pick up and write Python code with ease; it does not require advanced programming knowledge [25].

The machine learning method that produces a trained model is used in the training phase of workflow. The correctness of the trained model is then repeatedly assessed against test data, which should ideally consist of a distinct collection of data for each test run. The model is then deployed so that the application can use it once it reaches the required accuracy. Figure 7 shows the workflow of machine learning methods.

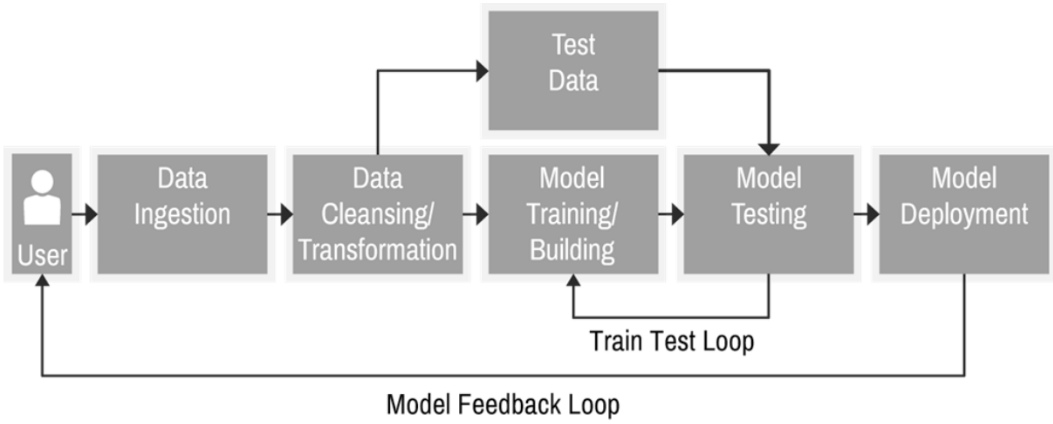

Figure 7: The machine learning workflow: The data input phase of the workflow is followed by data preparation, which converts the clean, valid data into a format that best fits the needs of the machine learning model being employed.

2. Methods

2.1 Experiment

The flexographic experimental trials are executed with three input variables such as: paper substrates with two different roughness levels, anilox rollers with three different line rulings and square shaped AM halftone dots with 114 dot area levels. The anilox roller rulings are selected as: 700 lpi (275.6 lines/cm), 1000 lpi (393.7 lines/cm) and 1300 lpi (lines per inch) (511.8 lines/cm). The paper roughnesses are chosen as: 36 ml/min and 182 ml/min. The engraved anilox cells that features 60º hexagonal in geometry, Volume of cells (BCM 4.1 (700 lpi), BCM 3.9 (1000 lpi), BCM 3.8 (1300 lpi)), Distortion factor (1.3%) with same mounting tape. The Flexographic printing machine such as OMET LAB230 Iflex (Printing speed – 35 m/min, Nip pressure – 3 mm) with Black colour uv ink are employed for the printing trials. The photopolymer plate (Manufacturer – Dupont) with thickness of 1.14 mm magnetic backing is used as the printing image carrier. The parameters such as nip pressure, printing speed, room temperature and ink features etc. are kept constant throughout the printing process.

2.2 Measurements and data collection

The output response was taken as dot gain for the assessment of print quality. So, the quantitative measurement of output parameters was done with Spectro-Densitometer such as X-Rite Spectro Eye instrument. The visual inspection of printed area was observed with a Digital Microscope such as LEICA S8APO. The input datasets for the machine learning algorithms are created from the readings up to 114 print trials. The split ratio of training to test dataset is taken as 7:3. The collected data are inputted to the Python programme to process with machine learning algorithms.

Table 1 *Parameter background at printing*

Factors	Unit	Symbol	Levels	Response
Anilox Ruling	Lines per inch (lpi)	A	700, 1000, 1300	Dot Gain
Paper Roughness	Millilitre per minute (ml/min)	B	36, 182	
Dot Percentage	Percentage (%)	C	5% to 100 % with 5% increments.	
Halftone dot shape	-	D	Square Shape	

Table 2 *Summary of dataset:*

	Paper Roughness	Anilox Roller Ruling	Dot Percentage	Dot Gain
count	114.000000	114.000000	114.000000	114.000000
mean	109.000000	1000.000000	50.000000	10.245614
std	73.322297	246.030432	27.507039	4.103876
min	36.000000	700.000000	5.000000	2.000000
25%	36.000000	700.000000	25.000000	8.000000
50%	109.000000	1000.000000	50.000000	10.000000
75%	182.000000	1300.000000	75.000000	13.000000
max	182.000000	1300.000000	95.000000	19.000000
Shape of data:	(114, 4)			
Shape of split data	The shape of X_train : (85, 3)			
	The shape of X_test : (29, 3)			
	The shape of y_train : (85,)			
	The shape of y_test : (29,)			

Table 3 *Performance accuracy of various techniques*

Sl. No.	Machine Learning Algorithm	Accuracy
1.	Linear Regression	74.85 %
2.	Decision Tree Regressor	89.38 %
3.	Random Forest Regressor	93.02 %

The Table 1 shows different process parameters or factors, their values, corresponding levels and the response variable such as dot gain etc. that are considering in the analysis process.

2.3 Programming to train the model

The programming code to perform the machine learning algorithms such as linear regression, decision tree and random forest regression are done with python programming language. The influence of input parameters over the output response are evaluated with machine learning techniques and the accuracy are compared with each technique.

3. Results

The result obtained at the machine learning algorithms at python programming execution are given at Table 2 & 3. The summary of dataset as Table 2 and the performance accuracy of various algorithms as Table 3 are given.

The accuracy talks about the capability of the developed model to forecast the result from the given input variables.

3.1 Graphical analysis

The graphical plots that demonstrate the nature of distribution of datasets, residuals found at the analysis part and the relationship between various parameters associated with the experimental process.

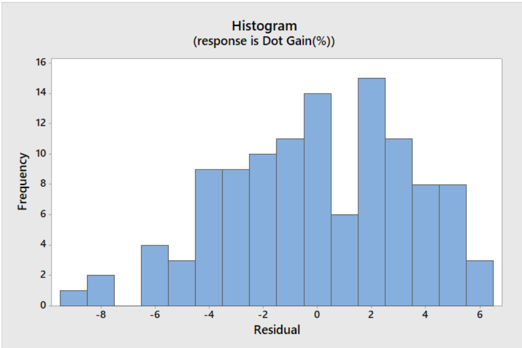

Figure 8: Residual plot for dot gain in histogram

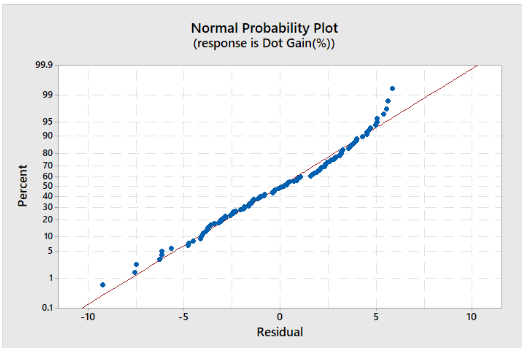

Figure 9: Normal probability plot for dot gain

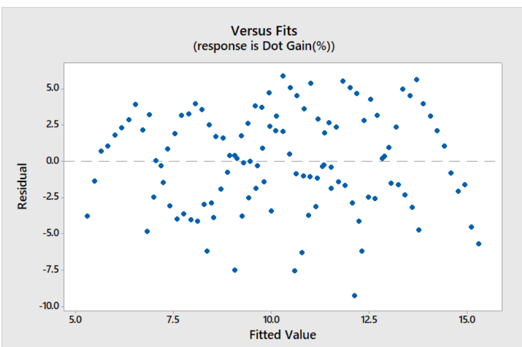

Figure 10: Versus fit plot for dot gain

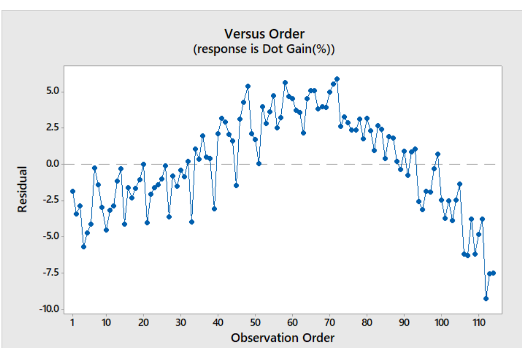

Figure 11: Versus order plot for dot gain

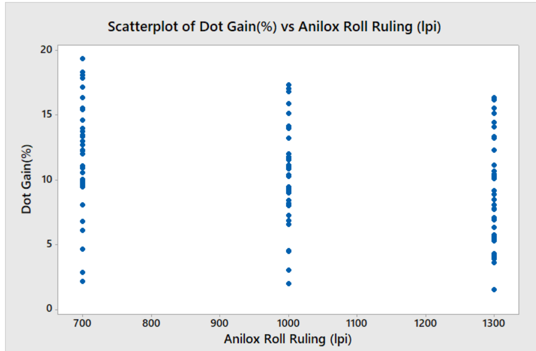

Figure 12: Scatterplot of dot gain vs anilox roller ruling

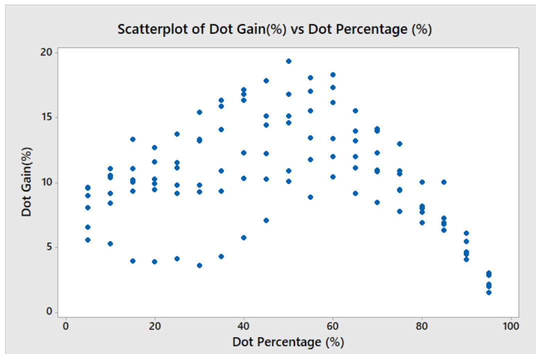

Figure 13: Scatterplot of dot gain vs dot percentage

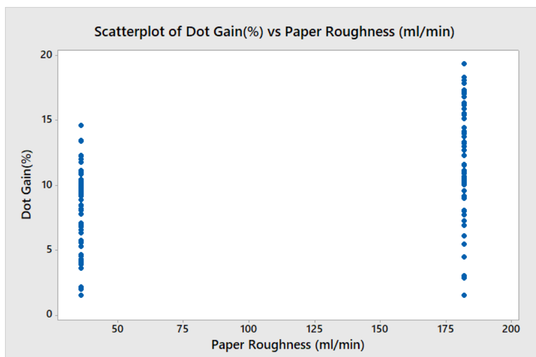

Figure 14: Scatterplot of dot gain vs paper roughness

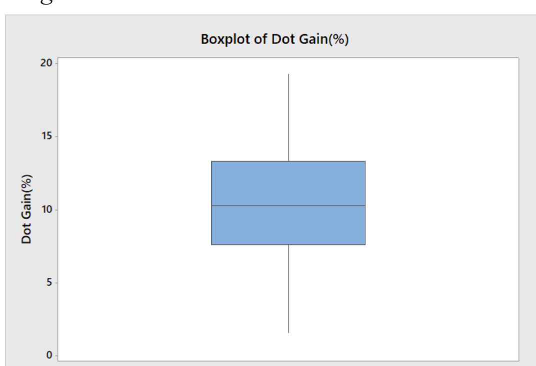

Figure 15: Boxplot of dot gain

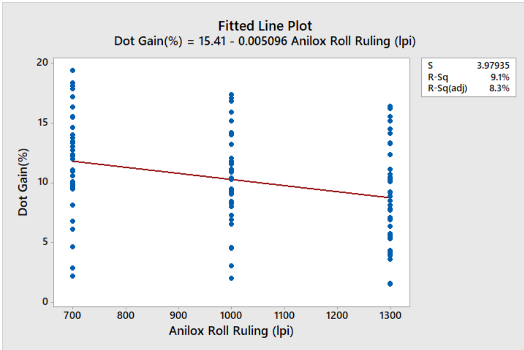

Figure 16: Fitted line plot for dot gain vs anilox roll ruling

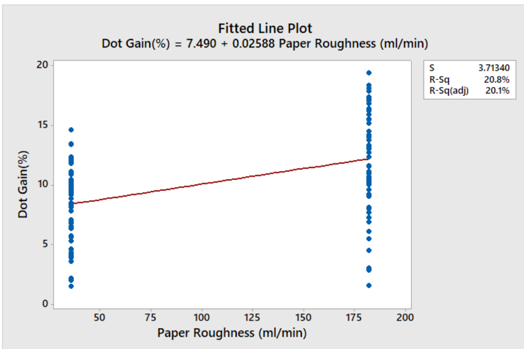

Figure 17: Fitted line plot for dot gain vs paper roughness

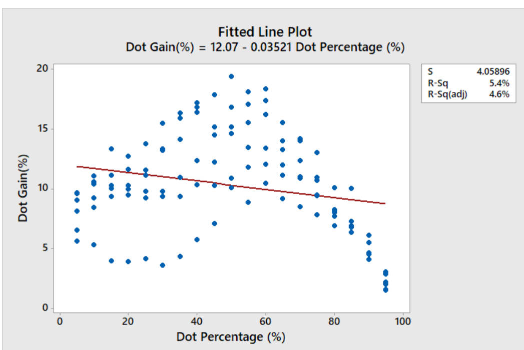

Figure 18: Fitted line plot for dot gain vs dot percentage

4. Discussions

The Table 3 comprises the result that shows the random forest regression algorithm is comparatively better in the prediction of accuracy than linear regression and decision tree algorithms. The random forest regressor is an ensemble learning technique that actually combines the predictions from several decision trees to generate a single, more reliable forecast, so that the accuracy and reliability of this algorithm will be more. When, Linear and decision tree regression utilizes a single model for prediction, the Random Forest uses several decision trees instead. The linear regression will be more accurate if the relationship between variables is linear only so that the linear regression is often less effective than random forest method. For a decision tree algorithm, each decision tree has a significant variance, but when we aggregate them all at once, the variance that results is reduced because each decision tree is perfectly trained on that specific sample data. This method adopted by the random forest make it stronger and more adaptable for regression-related problems. It is simpler to use, less dependent on the training set and is more accurate. It works well with big datasets with plenty of attributes, Outliers, noisy characteristics, and missing data are all handled by it. The main Python machine learning library is called sklearn. It offers a large selection of tools for building, preprocessing, assessing, and applying machine learning models.

The Figure 8 is the histogram plot for the residuals found in the analysis part. The normal distribution of residuals is found somewhat asymmetrical at some points. The Figure 9 represents the normal probability plot that shows the visual data regarding the deviation of response from normal distribution. The Figure 10 shows the versus fit plot that represents the non-linearity, outliers and undesirable errors associated with the dataset found from the experimental process. The Figure 11 indicates the appropriateness of data collection and serial correlation of errors in the experimental sequence.

The Figure 12 is a scatterplot that indicates the possibility of decreasing the dot gain with increasing anilox roller ruling in flexographic printing. The scatterplot of Figure 13 tells the fact that the dot gain will be more if the substrate has high porous texture. The Figure 14 illustrates that, for square dots, the occurrence of dot gain is more at middle tone areas of the print.

The Figure 15 is a boxplot of dot gain showing

5. Conclusion

The various sectors of the work are organized as: background information on flexographic printing and machine learning, delves into supervised techniques using the Python language, outlines the proposed model for data analytics, addresses performance evaluation, examination on result analysis.

The implementation of machine learning algorithms with three different supervised learning algorithms such as linear regression, decision tree and random forest regressor are carried out. The result shows that the random forest regressor having the better efficiency based on the prediction of accuracy of the model.

The graphical plot shows the nature of distribution of datasets and the relationships between output response and input variables associated with the flexographic printing experiment.

References

[1] Abusaq, Z., Zahoor, S., Habib, S., Rehman, M., Mahmood, J., Kanan, M., Mushtaq, T., Ray, R. (2023). Improving energy performance in flexographic printing process through lean and AI techniques: A case study Journal of MDPI. Energies [Internet]. 2023;16:1972. Available from: https://doi.org/10.3390/en16041972

[2] Kipphan, H. (2001). Handbook of Print Media: Technologies and Production Methods. Springer.

[3] Joshi, V.A. (2022). Optimization of flexo process parameters to reduce the overall manufacturing cost. An International Journal of Optimization and Control: Theories & Applications [Internet]. ISSN: 2146-0957, eISSN: 2146-5703 ,2022;12(1):66-78. Available from: https://doi.org/10.11121/ijocta.2022.1137

[4] Flexography: Principles & Practices. Foundation of Flexographic Technical Association, 2013, ISBN-10: 0989437418

[5] Bould, C.D., Hamblyn, M.S., Gethin, T.D., Claypole, C.T. (2010). Effect of impression pressure and anilox specification on solid and halftone density. Journal of Engineering Manufacture [Internet]. 2010. DOI : 10.1177/2041297510394072

[6] R. H. Leach, R. J. Pierce, E. P. Hickman, M. J. Mackenzie, H. G. Smith. (1993). The Printing Ink Manual. Springer, ISBN: 978-0-948905-81-0

[7] Dendge, R.R. (2023). Analysis of dot gain produced by interactions of flexographic plate and anilox roll screen frequencies. Acta Graphica [Internet]. 2023;31(1). Available from: https://actagraphica.hr/index.php/actagraphica/article/view/229 .

[8] Żołek-Tryznowska, Z., Rombel, M., Petriaszwili, G., & Kasikovic, N. (2020). "Influence of Some Flexographic Printing Process Conditions on the Optical Density and Tonal Value Increase of Overprinted Plastic Films." Coatings, 10(9), 816. https://doi.org/10.3390/coatings10090816

[9] Theohari, S., Fraggedakis, E., Tsimis, D., Tsigonias, M., Mandis, D. (2014). Effect of paper properties on print quality by flexographic method. [Internet]. Available from: https://www.researchgate.net/publication/263847022

[10] Tomasegovic, T., Pibernik, J., Poljacek, M.S., Madzar, A. (2021). Optimization of flexographic print properties on ecologically favourable paper substrates [Internet]. 2021;1. Available from: https://doi.org/10.24867/JGED-2021-1-037.

[11] Żołek-Tryznowska, Z.; Rombel, M.; Petriaszwili, G.; Dedijer, S.; Kašiković, N. Influence of Some Flexographic Printing Process Conditions on the Optical Density and Tonal Value Increase of Overprinted Plastic Films. Coatings 2020, 10, 816. https://doi.

org/10.3390/coatings10090816

[12] Dharavath, N.H., Bensen, M.T., Gaddam, B. (2005). Analysis of print attributes of Amplitude Modulated (AM) vs. Frequency Modulated (FM) screening of multicolor offset printing. Journal of Industrial Technology [Internet]. 2005;21(3). Available from: www.nait.org.

[13] Kirwan, M. J. (Ed.). (2008). Paper and Paperboard Packaging Technology. Wiley-Blackwell. ISBN: 978-1-405-16845-8.

[14] Çelik, Ö., Altunaydin, S. (2018). A research on machine learning methods and its applications. Journal of Educational Technology & Online Learning [Internet]. 2018;1(3):25-40. Available from: https://doi.org/10.31681/jetol.457046

[15] Turing, A. M. (1950). "Computing Machinery and Intelligence." Mind, 59(236), 433-460.

[16] Chandel, M., Silakari, S., Pandey, R., Sharma, S. (2022). A study on machine learning and Python's framework. International Journal of Computer Sciences and Engineering [Internet]. Vol. 10, Issue.5, May 2022; E-ISSN: 2347-2693. Available from: https://doi.org/10.26438/ijcse/v10i5.5864

[17] Mitchell, T. M. (1997). Machine Learning. McGraw-Hill

[18] Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.

[19] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.

[20] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

[21] Chapelle, O., Scholkopf, B., & Zien, A. (2006). Semi-Supervised Learning. MIT Press.

[22] Sutton, R. S., & Barto, A. G. (2015). Reinforcement Learning: An Introduction (2nd ed.). MIT Press.

[23] Van Engelen, J. E., & Hoos, H. H. (2020). "A survey on semi-supervised learning." Machine Learning, 109(2), 373-440.

[24] Sodhi, P., Awasth, N., Sharma, V. (2019). Introduction to machine learning and its basic application in Python. SSRN Electronic Journal [Internet]. 2019. Available from: https://doi.org/10.2139/ssrn.3323796

[25] Sisodia, P., Seth, B. (2022). An implementation on Python for data science and machine learning. International Journal of Creative Research Thoughts [Internet]. 2022;10(3). | ISSN: 2320-2882. Available from: www.ijcrt.org