An Application of Predictive Analytics in Manufacturing Sector for Price Prediction and Demand Prediction

Darshan Labhade, Nikhil Lakare, Aniket Mohite, Siddhesh Bhavsar, Sushma Vispute

Abstract*: Predictive analytics is the examination of concerned data so that we can recognize the problem that may arise in the near future. Manufacturers are interested in quality control, and making sure that the whole factory is functioning at the best possible efficiency. Hence, it's feasible to increase manufacturing quality, and expect needs throughout the factory with predictive analytics. Hence, we have proposed an application of predictive analytics in manufacturing sector especially focused on price prediction and demand prediction of various products that get manufactured on regular basis. We have trained and tested different machine learning algorithms that can be used to predict price as well as demand of a particular product using historical data about that product's sales and other transactions. Out of these different tested algorithms, we have selected the regression tree algorithm which gives accuracy of 95.66% for demand prediction and 88.85% for price prediction. Therefore, Regression Tree is best suited for use in manufacturing sector as long as price prediction and demand prediction of a product is concerned. Thus, the proposed application can help the manufacturing sector to improve its overall functioning and efficiency using the price prediction and demand prediction of products.

Keywords:* Analytics, Demand Prediction, Prediction, Manufacturing Sector, Machine Learning Algorithms, Prediction, Price Prediction, Regression Trees.

I. INTRODUCTION

Predictive analytics includes different mathematical and statistical techniques from machine learning and data mining that analyze historical and real-time data to determine the pattern and predict future outcomes such as demand and price of the product. Demand and price forecasting are one of the most important functions of manufacturers. Overestimated and underestimated demand causes excess production and unfulfilled orders respectively. Thus, the accurate demand and price prediction is a real challenge. Traditional prediction methods undergo through serious limitations which affect the forecasting accuracy. Regression tree and Ridge regression have been found to be useful techniques for demand and price prediction due to ability to reduce model complexity and prevent over-fitting.

We know that the manufacturing businesses are growing day by day as the demand for different products is increasing at fast rate. As business size grows, it becomes difficult to manage different parts that are required to manufacture different types of products. Also, many difficulties faced during the job scheduling process due to the inefficient management of resources. This inefficient management of resources leads to the over processing and over-production which leads to the unsatisfactory business outcomes. Our motivation behind this proposed system is to help these growing businesses to take decisions for effective management of the resource, sales and demand of products.

Predictive model has many applications in business. Predictive analytics mostly used in advertising and marketing. In manufacturing sector, it uses historical and real time data and running through algorithms to determine what sales, price of products such that manufacturer might get higher profit as well proper demand management. According to application of predictive model one need to choose different algorithms.

II. LITERATURE SURVEY

Recent study of machine learning in estimation of manufacturing cost of jet engine components, multiple linear regression gives less accuracy on other hand gradient boosted trees gives highest accuracy [1]

Wavelet support vector machine and particle swarm optimization combined and used as new hybrid forecasting model. This hybrid model gives good forecasting results of the product sales and demand in dealing with uncertain data and finite samples [2]

Table- I: Literature Survey

Application	Used Algorithms
Product demand forecasting	WSVM and POS (MSE =26.1667)
The estimation of manufacturing cost of jet engine components	Multiple Linear Regression (R^2 =0.62), Gradient Boosted Trees (R^2 =0.96)

Revised Manuscript Received on June 25, 2020.

***** Correspondence Author

Darshan Labhade, Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India. Email: labhadedarshan@gmail.com

Nikhil Lakare*, Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India. Email: nikhil.lakare4@gmail.com

Aniket Mohite, Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India. Email: aniket91mohite@gmail.com

Siddhesh Bhavsar, Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India. Email: siddhesh099@gmail.com

Sushma Vispute, Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India. Email: visputesushma@gmail.com

III. ALGORITHMIC SURVEY

A. Linear Regression

Linear regression is a systematic, autonomous statistical method to modelling the relationship between a dependent variable and a specified set of independent variables. A linear approach is the simulation of the interaction between a dependent variable and one or more independent variables. It is called basic linear regression because we only have one independent variable. Linear regression should consider the best line to match the database points [3].

B. Polynomial Regression

In this regression, the relationship between dependent and the variable is modelled such that the dependent variable Y is an nth degree function of independent variable Y. The polynomial regression fits into a non-linear relation between X value and Y values. The Polynomial regression is also called as multiple linear regression [4], [5].

C. Ridge Regression

It is a regularization technique used to create a model having a minimum number of parameters when the number of predictor variables in a set exceeds the number of observations, or when a data set has multicollinear. As multicollinearity exists, the least square calculations are impartial, but their variances are broad in order to differentiate themselves from the true value. By adding a degree of bias to the regression estimates, ridge regression reduces the standard errors. Ridge uses L2 norm. This adds regularization terms in the model, which are functions of square of coefficients of parameters [6].

D. Lasso Regression

Lasso stands for Least Absolute Shrinkage and Selection Operator. Through penalizing the model with a penalty word which is the number of absolute coefficients, this strategy shrinks the regression coefficients towards zero. Lasso performs better in situations where some predictors have large coefficients and the remaining have small coefficients. Lasso uses L1 norm. This adds regularization terms in the model, which are functions of absolute value of coefficients [7].

E. Regression Trees

Regression tree allows input variables to be a mixture of continuous and categorical variables. The output of the regression tree is a numerical value. A Regression tree may be considered as a variant of decision trees, designed to approximate real-valued functions, instead of being used for classification methods [8].

IV. DATASETS

To select machine learning algorithm which gives highest accuracy we have tested different algorithms on different datasets. For prediction of demand and price of product we have used two datasets. First dataset used for price prediction and another one dataset used for demand prediction. Dataset of price prediction contains 1200 tuples along with 4 attributes in each tuple. Second dataset is used for prediction of demand of product. Demand of product means future count of product will be sold and calculated using historical sales. Dataset contains 1200 tuples and 2 attributes in each tuple. List of input feature for price prediction is month, sales, manufacturing cost, selling price and for demand prediction is month and sales.

Table- II: Sample Demand Data

Method	Demand
January	21
February	30
March	40
April	42

Table- III: Sample Price Data

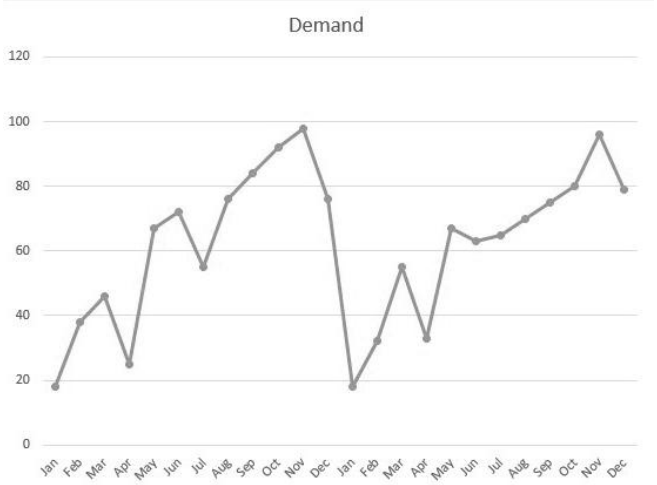

Fig. 1. Trend of Demand Data

Month	Sales	Manufacturing Cost	Selling Price
January	22	3050	5448
February	32	3000	4900
March	43	3594	4950
April	39	2970	4700
May	58	3050	4600

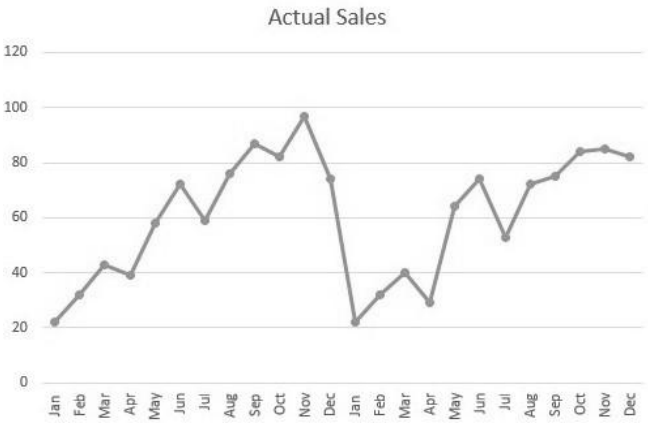

Fig. 1. Trend of Actual Sales Data

Fig. 2. Trend of Price Data

V. PROPOSED MODEL

Our proposed model consists of four major parts viz. Historical Dataset, Data Pre-processing, ML model training unit and trained model unit for serving the prediction for a particular input.

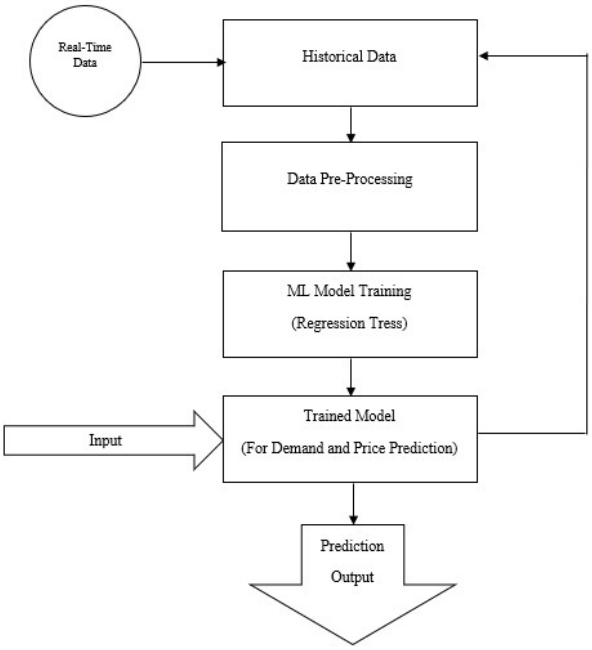

Fig. 2. Proposed Model Diagram

A. Historical Data Unit

This unit holds the whole dataset of historical demands and prices. Also, this unit gets updated with new real time data with the help Real-Time Data updater unit.

B. Data Preprocessing Unit

It consists of different data preprocessing techniques such as data scaling, data normalization and handling missing values.

C. ML Model Training Unit

This unit collects the preprocessed historical data and feeds to the regression tree algorithm for its learning. Regression tree allows input variables to be a mixture of continuous and categorical variables. The output of the regression tree is a numerical value. The trained model is then exported for its use to predict the demand and price of respective product [8]. The training process is repeated when any significant amount of new data is added to the historical data

D. Trained Model Unit

This trained machine learning model accepts the required input from the user and predicts the demand and price of that particular product for a given month

Fig. 4. shows all the above-mentioned module of proposed system

VI. RESULTS OF ALGORITHM

We have tested different algorithms for their error rate and accuracies using historical dataset of demand and price. All results are tabulated as follows.

Table- IV: Results for dataset without Scaling

Algorithm	Dataset	MSE Std Scaler	R2 Std Scaler	Other
Regression Trees	Demand	34.2825	0.9375	MSE
Polynomial Regression	Demand	34.2825	0.9375	Degree=11
Linear Regression	Demand	152.0249	0.7232	
Regression Trees	Price	0.8605	0.8622	MSE
Lasso Regression	Price	14801.3326	0.8093	Normalize= FALSE
Ridge Regression	Price	20490.5519	0.8092	Alpha =18.0

Table- V: Results for dataset with Standard Scaling

Algorithm	Dataset	MSE	R2	Other
Regression Trees	Demand	34.2825	0.9375	MSE
Polynomial Regression	Demand	34.1198	0.9378	Degree=11
Linear Regression	Demand	152.0249	0.7232	
Regression Trees	Price	14986.255	0.8605	MSE
Lasso Regression	Price	20542.91	0.8088	Normalize= FALSE
Ridge Regression	Price	20542.901	0.8088	Alpha =18.0

Table- VI: Results for dataset with MinMax Scaling

Algorithm	Dataset	MSE MinMax	R2 MinMax	Other
Regression Trees	Demand	34.2825	0.9375	MSE
Polynomial Regression	Demand	34.2825	0.9375	Degree=11
Linear Regression	Demand	152.0249	0.7232	
Regression Trees	Price	15007.4087	0.8603	MSE
Lasso Regression	Price	20336.6345	0.8107	Normalize= FALSE
Ridge Regression	Price	20419.0405	0.8099	Alpha =18.0

Table- VII: Results for Accuracies with Cross Validation

Algorithm	Dataset	Accuracy (%)	Folds (CV Values)	Algorithm
Regression Trees	Demand	0.9566	9	Regression Trees
Polynomial Regression	Demand	0.954	4	Polynomial Regression
Linear Regression	Demand	0.7978	7	Linear Regression
Regression Trees	Price	0.8885	7	Regression Trees
Lasso Regression	Price	0.8296	6	Lasso Regression
Ridge Regression	Price	0.8296	6	Ridge Regression

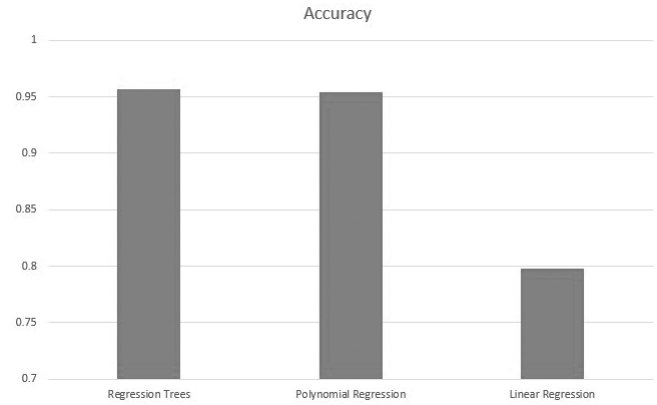

Fig. 5. Accuracy Plot for Demand Data

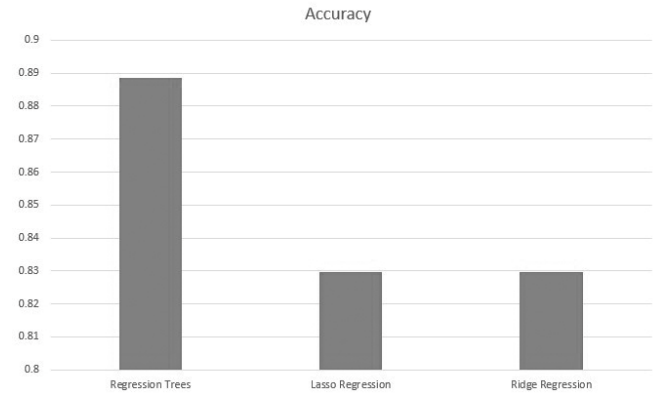

Fig. 6. Accuracy Plot for Price Data

VII. CONCLUSION

Hence, we proposed an application of predictive analytics in manufacturing sector especially focused on price prediction and demand prediction of various products that get manufactured on regular basis. We trained and tested different machine learning algorithms that can be used to predict price as well as demand of a particular product using historical data about that product's sales and other transactions. Out of these different tested algorithms, we selected the regression trees which accuracy of 95.66% for demand prediction and 88.85% for price prediction. Therefore, Regression tree is best suited for use in manufacturing sector as long as price prediction and demand prediction of a product is concerned.

REFERENCES

1. Loyer, J.-L., Henriques, E., Fontul, M., & Wiseall, S. (2016), "Comparison of Machine Learning methods applied to the estimation of manufacturing cost of jet engine components", International Journal of Production Economics, 178, 109–119.
2. Qi Wu. 2010 "Product demand forecasts using wavelet kernel support vector machine and particle swarm optimization in manufacture system", J. Comput. Appl. Math. 233, 10 (March 2010), 2481–2491.
3. George A F Seber and Alan J Lee "Linear Regression Analysis", Wiley, 2003.
4. Nathans, Laura L, Oswald, Frederick L and Nimon, Kim (2012) "Interpreting Multiple Linear Regression: A Guidebook of Variable Importance", Practical Assessment, Research, and Evaluation: Vol. 17, Article 9.
5. Palmer, Phillip B, and Dennis G O'Connell "Regression analysis for prediction: understanding the process", Cardiopulmonary physical therapy journal vol. 20,3 (2009): 23-6.
6. Marquardt, Donald & Snee, Ron. (1975). "Ridge Regression in Practice'', American Statistician - AMER STATIST. 29. 3-20. 10.1080/00031305.1975.10479105.
7. Gauraha, N "Introduction to the LASSO", Reson 23, 439–464 (2018).
8. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interfaces(Translation Journals style)," *IEEE Transl. J. Magn.Jpn.*, vol. 2, Aug. 1987, pp. 740–741 [*Dig. 9th Annu. Conf. Magnetics* Japan, 1982, p. 301 Lingjian Yang, Songsong Liu, Sophia Tsoka, Lazaros G. Papageorgiou "A regression tree approach using mathematical programming, Expert Systems with Applications", Volume 78, (2017).
9. D. Labhade, N. Lakare, A. Mohite, S. Bhavsar, S. Vispute, and G. Mahajan, "An Overview of Machine Learning Techniques and Tools for Predictive Analytics", AJCT, vol. 5, no. 3, Mar. 2020.
10. S. R. Vispute and M. A. Potey, "Automatic text categorization of marathi documents using clustering technique," 2013 15th International Conference on Advanced Computing Technologies (ICACT), Rajampet, 2013, pp. 1-5.
11. Nikita Shahane and Mrs. S. R. Vispute, "Survey of Big Data Analysis Using Predictive Analytics Algorithms and Its Use Cases,", International Journal Of Integrated Computer Applications & Research (IJICAR), vol. 1, issue 1, May 2015.
12. S. R. Vispute, Nilesh Machale, Ashitosh Dhaware, Amruta Khandare, "An overview of Machine Learning Techniques for Student Performance Analysis", International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE), vol. 6, issue 11, Nov. 2018.

AUTHORS PROFILE

Darshan Labhade, Bachelor of Engineering Graduate in Computer engineering.

Nikhil Lakare, Bachelor of Engineering Graduate in Computer Engineering.

Aniket Mohite Bachelor of Engineering Graduate in Computer Engineering.

Siddhesh Bhavsar, Bachelor of Engineering Graduate in Computer Engineering.

Sushma Vispute, Assistant Professor of Computer Engineering Department, Pimpri Chinchwad College of Engineering, Pune.