Application of business intelligence technique to manufacturing system with stochastic process (a case study of “product-based” manufacturing company in Nigeria with make-to-order strategy)

Ayanlowo Babatunde¹, Christian Bolu¹, Festus Oyawale,¹ O.S.I Fayomi

Department of Mechanical Engineering, Covenant University Ota, Ogun State, Nigeria.
*Corresponding Author; ayanlola2002@yahoo.com

Abstract
Manufacturing outfit in Nigeria with make-to-order strategy continually contend with the problem of long waiting time of customers for orders, which arose through inability to predict order quantity amount for customers. This indicates the need for the application of business intelligence technique on large amount of unused data in manufacturing databases, to evaluate and validate probability distribution model that is best suitable for prediction of customers’ order values and subsequently using the best model to predict daily and yearly customers order amount. Many models were evaluated for data extracted from a manufacturing outfit and the best model was selected based on statistical goodness of fit. The normal distribution model was found most appropriate for the prediction of daily and monthly order quantity amount. Based on this, the application of business intelligence technique on extracted data has made the prediction of customers order for a product-based manufacturing outfit with make-to-order strategy possible.

Keywords: Business intelligence, make-to-order, probability distribution models, Maximum log-likelihood, quartile-quartile plots,

1. Introduction
Business intelligence is the use of an organization’s disparate data to provide meaningful information and analysis to employees, customers, suppliers and partners for more efficient and effective decision-making” [1]. Levenson & Boris [2] gave a broad definition of business intelligence as “a set of methodologies, processes, architectures, and technologies that transform raw data into meaningful and useful information used to enable more effective strategic, tactical, and operational insights and decision-making”. In this paper "Business intelligence technique includes three major components as proposed by Scholz et al., [3] in their studies :The data warehouse or data source, BI analytical and statistical tools that analyze and mine data in the data store , a user interface to view the results returned from analysis. This paper is aimed at applying business intelligence technique to extracted data of a product-based manufacturing outfit based in Ota area Ogun state Nigeria, in order to develop a model to forecast daily and monthly demands of customers order quantities of product using five probability distribution models namely Normal, log-Normal, Poisson, Weibull and gamma for the statistical analysis. These models were tested using Kolmogorov-Smirnov test(K), Maximum Log Likelihood (M) and QQ plots, so as to determine the probability distribution model that best fit the product-based manufacturing data extracted and which could be recommended for the prediction of future demands of order
quantities. With the three statistical test analysis, a model that is most suitable for the prediction of future demands of the customer order could be selected. The prediction accuracy of manufacturing outfit models depends primarily on the quality of data collection and how big the data is. It is normally assumed that the largeness of real manufacturing data extracted and properly cleaned will minimized errors, and thus they are widely used in modelling applications. The case study is a product-based manufacturing outfit located in Ota industrial area Ogun state Nigeria. Manufacturing sales data was extracted from two database system (Octopus and SAP-Business One). The data from the two-database system were properly cleaned and merged together to form a full 6 years sales data.

1.1 MODEL SIGNIFICANCE TO PRODUCT-BASED MANUFACTURING SYSTEM WITH MAKE-TO-ORDER STRATEGY

Estimating the future demands of customers order quantities in product-based manufacturing outfit with a make-to order production strategy for certain projected time is very necessary. The significance of customer dissatisfaction due to the long waiting time for various order amounts with stochasticity has always being a known issue in a product-based manufacturing outfit. This necessitates a need for developing models from huge unused sales data stored in these outfits, which can be used for estimating future demands of customers’ orders and that the best distribution model confirmed by more than one goodness of fit tests be adopted for the analysis to predict the most accurate future customers orders [4], [5]. In order to extract meaningful information from outfits data records and to evaluate the most possible nature about the corresponding manufacturing data, hence the application of business intelligence technique for adequate probability distribution function has been found to be the subject of several studies in a manufacturing system [6], [7], [8]. Nevertheless, not much of this technique is applied in a manufacturing outfit that operate make-to-order strategy, which are aimed at obtaining the model that best fit the observed manufacturing data for customer orders. Nevertheless, it is known fact that customer orders vary significantly in every manufacturing system with a make-to-order strategy. Chen et al [9] showed that make-to-order strategy “usually creates additional wait time for the consumer to receive the product demands” which usually brings about customer dissatisfaction in most manufacturing systems. This usually bring about high rate of customers attrition from manufacturing systems.

2. METHODS: DATA EXTRACTION, ANALYSIS AND STATISTICAL TESTS

2.1 Data Extraction

The Manufacturing data of the study were obtained from Kolokorte Nigeria Limited; a product-based manufacturing company with make-to-order production strategy in OTA - Ogun state, Nigeria. Raw sales data samples were extracted using query language from two separate databases, each had a range of three years.

2.2 Data Analysis

The analytical method that was adopted in this paper was the probability distribution function which gives the understanding of the behavioral pattern of customer order quantity and provided a basis for future forecasting. The methods employed included Gamma, Poisson, Weibull, Normal and log-normal distributions. The procedures involved the estimation of parameters such as mean, standard deviation, percentage quantiles, and shape parameter for each of the five selected
distributions. These parameters were used to determine the model equation for the best-fit distribution.

2.3 Statistical Test

The acceptability and reliability of the fitting (probability distribution) models on the extracted data were tested using three statistical tests often known as goodness of fit test. The three GOF tests were necessary because “models are considered dependable and used for applications, if more than one Goodness of fits methods were used to confirm the validity about the distribution chosen”. The statistical tools adopted included: Kolmogorov-Smirnov (D), Maximum-log-likelihood and Quartile-Quartile plots. The Goodness of fit test was carried out in accordance to standard procedure [10], [11]. The three GOFs tests used are described in the following sections.

2.3.1 KOLMOGOROV-SMIRNOV TEST (TEST 1)

Kolmogorov-smirnov statistic is based on the largest vertical difference between the EDF and distribution function of any given data sample. It is computed as follows:

\[ D = \max \left\{ x | F_n(x) - \frac{1}{\theta} \sum_{i=1}^{\theta} y_i \right\} \]  

(1)

where \( F_n(x) \) is the largest vertical distance between the empirical distribution function and the distribution function when the EDF > PDF. \( F(x) \) is the largest vertical distance when EDF is < than the distribution function.

The kolmogorov-smirnov test was adopted in this paper to test the Goodness of Fit for the five selected distributions under this research, this was done by comparing the kolmogorov-smirnov test statistic (D) generated from analysis with the critical value (alpha= .645) obtained from z-tables at 5 % confidence level. Any of the five distributional forms selected will be rejected at 0.05 confidence level, If their respective test statistic (D) is greater than the critical value (alpha=0.645) obtained from the Z-tables. However, if the p-value of obtaining the test statistic for a selected distribution is less than .05 (standard rule), we could conclude that the selected distribution fits the manufacturing sample data collected well [12].

2.3.2 MAXIMUM LOG-LIKELIHOOD TESTS (TEST 2)

The second Goodness of fit test in this paper, to select best fit out of the five selected distributions is the maximum log-likelihood test. The maximum log-likelihood test is the best for generalized linear model i.e. Gamma, Normal, log-normal, Weibull, Poisson. The maximum log-likelihood-ratio statistic (D) uses two statistics known as the Scaled deviance and the Scaled Pearson’s Chi-square. The maximum log-likelihood is expressed in terms of the mean parameter, \( \mu \) and the log-likelihood ratio which is the Scaled deviance and is expressed as follows:

\[ D^*(\mathbf{y}; \mathbf{\theta}) = -2 \sum \left( \hat{\mathbf{\theta}}^{max} \mathbf{y} - \hat{\mathbf{\theta}}^{max} \right) \]  

(2)

where:

\( I(\hat{\mathbf{\theta}}; \mathbf{y}) \): is the log-likelihood under the studied model?

\( I(\hat{\mathbf{\theta}}^{max}; \mathbf{y}) \): is the log-likelihood under the maximum achievable (saturated) model?

For generalized linear models the Scaled deviance is expressed as:

\[ D^*(\mathbf{y}; \mathbf{\theta}) = \frac{1}{\hat{\mathbf{\theta}}} D^* (\mathbf{y}; \mathbf{\theta}) \]  

(3)

where:

\( D^*(\mathbf{y}; \mathbf{\theta}) \): is the residual deviance for the model and it’s the sum of the individual deviance contributions.
\( \phi \): is the dispersion parameter. Also, the Scaled Pearson’s chi-square follows the same approach as the Scaled deviance above. The Pearson’s chi-square statistic is defined as:

\[
X^2 = \sum_{i=1}^{n} \frac{w_i(y_i - \mu_i)^2}{V(\mu_i)}
\]  

(4)

\( y_i \): Is the response variable.
\( w_i \): Weight which is set to one for all observation.
\( \mu_i \): Is the predicted mean values.
\( V(\mu_i) \): Is the variance function for the distribution under study.

Both the Scaled deviance and Scaled Pearson’s chi-square could be used as approximate guides to determine the Goodness of Fit for a given distributional model” [13]. For the given manufacturing order quantity amount sample data collected, if the ratio of the Scaled deviance and the degree of freedom is far greater than 1 i.e. \( D^* (\gamma; \hat{\theta})/DF > 1 \), then over-dispersion is considered to be present in the data and if the ratio of Scaled deviance and degree of freedom is far lesser than 1 i.e \( D^* (\gamma; \hat{\theta})/DF < 1 \) then under-dispersion is considered to be present in data. However if the value of \( D^* (\gamma; \hat{\theta})/DF \) is equal to one, it is a strong evidence against the statistical significance of the distributional form in which the model is built.

3. Results and Discussion

The results of the three goodness of fit tests analysis method employed were determined and the results obtained for each of the five probability distribution models are presented as follows in Table 1, Table2, Fig 1, Fig 2, Fig 3, Fig 4:

Table 1: Summary of parameter estimates for kolmogorov-Smirnov test

| Distribution | Parameters | Goodness of fits |
|--------------|------------|------------------|
|              | Threshold  | Scale | shape | mean | Std.dev | KS-test  | P-value |
| Normal       | 0          | -     | -     | 1609 | 1173    | .09      | <.075   |
| Lognormal    | 0          | 6.92  | 1.20  | 2046 | 3608    | .165     | <.01    |
| Weibull      | 0          | 1703  | 1.2   | 1599 | 1328    | .292     | <.01    |
| Gamma        | 0          | 1328  | 1.2   | 1609 | 1462    | .12      | <.001   |

Table 2: Summary of Scaled Deviance and Scale Pearson Chi-Square

| Distribution | Scale Deviance | Scale Pearson Chi-Square |
|--------------|----------------|--------------------------|
| Normal       | 1.0            | 1.0                      |
| Lognormal    | 1.999          | 1.4567                   |
| Weibull      | 648            | 768                      |
| Gamma        | 1.1302         | .6435                    |
| Poisson      | 946            | 854                      |
The goodness of fit test was performed for the 5 probability distribution models selected for the order quantity amount extracted. The result for the best fit model for the order quantity amount of the product-based manufacturing outfit is presented in Table 3.

Table 3: Summary of best fit methods

| TESTS                   | BEST FIT DISTRIBUTION | CRITERIA                                              |
|-------------------------|------------------------|------------------------------------------------------|
| Kolmogorov-Smirnov      | Normal                 | Statistic value<critical value 0.09<1.645, p-value >0.05 |
| Maximum Log-Likelihood  | Normal                 | Scaled deviance=1 Scaled Pearson chi-square =1         |
| QQ plots                | Normal                 | Linear for normal QQ plots                           |
From Table 3, the best fit model for the quantity ordered amount of the product based manufacturing outfit is the Normal Distribution model with equation given as:

\[ \ln f(x) = 0.8241 + x/2751858. \]  \hspace{1cm} (5)

From table 1: The p-value of normal distribution is greater than 0.05, as compared to the p-value of log-normal, weibull, gamma which is less than .05 respectively. Also, the value of the K-statistic of the normal distribution gives a value lesser than the critical value obtained from tables at 90% confidence level i.e 0.09 < 1.645 (Z-score) when compared with other selected models. And which implies that the model is the strongest.

Additionally, from table 2, Scaled deviance and Scaled Pearson’s chi-square for the normal distribution stays at one, when compared to other Scaled deviance and Scaled pearson’s Chi-square statistic of other selected models and which all implies that the normal distribution model is the strongest. Also, checking the QQ graphs of the normal distribution of fig.2 and comparing it with other models QQ graphs, the normal distribution model trails the QQ linear line the most. Therefore, the model is absolutely satisfactory.

Since the objective of this paper is to evaluate and validate the best probability distribution model using statistical test and also the prediction of daily customer order quantity for next coming years based on the selected best fit model for the product based manufacturing outfit.

The result of the three statistical test was adopted in the selection of the best fit probability distribution model used for the prediction of next 31 daily customer orders and is presented in Table 4. Also, the equation of the best fit probability model which forms the basis of deriving the iterations is presented in (5) above.

4.0 Conclusion

A business intelligence technique approach which includes the following components: data extraction processes, analysis of extracted data for meaningful insights and reporting of various analysis is applied in this paper to sales data of a product-based manufacturing outfit with make-to-order strategy to evaluate and validate the probability models that is best suitable for the prediction of customer order quantity values and subsequently using the best model to predict daily order quantity amount of customers.

Five models were tested on extracted data for order quantity amount and the best model was selected based on the statistical goodness of fit test.

The Normal distribution model was found to be most appropriate for the prediction of the customer’s order quantity amount. The establishment of the best fit probability distribution model would be a useful guide in the prediction of various customers ordered quantity for the product based manufacturing outfit. Based on this, the application of business intelligence technique on extracted data would be useful in the prediction of customers order for a product-based manufacturing outfit with make-to-order strategy possible.

5.0 Recommendation

Extensive research can still be done for this paper in the area of risks analysis to ascertain the rate of customer churn with respect to specific time periods within year. This will actually improve the correctness of our forecast in reality. Also, research can be done to predict overall monthly and yearly demands of customer order of data using other statistical tools and machine learning tools like ARIMA model and random forests model respectively.
References

[1] Bocknor, M.P.A. (2012). Business intelligence applied to Homogeneous Diagnostic Groups. Information Systems and Technologies (CISTI), 5th Iberian Conference, (1 – 5).

[2] Levenson, R. & Boris, Quinn (2008). Worst Practices in Business Intelligence, MAIA Intelligence (2009, July 4), (1-2)(5-10).

[3] Scholz, & Patrick (2005). BENEFITS AND CHALLENGES OF BUSINESS INTELLIGENCE ADOPTION IN SMALL AND MEDIUM-SIZED ENTERPRISE Journal of Information Science, 34(2), (120-156)

[4] Boos, D. (1992). “On Generalized Score Tests,” The American Statistician, 46, 327–333. Dobson, A. (1990)

[5] Bob, D (2009), Inventory management: combining business intelligence & material handling hardware. A journal of Inventory Management in a Manufacturing System (3-7)

[6] Bergeron, B. (2000). Regional business intelligence: The view from Canada. Journal of Information Science, 26(3), (153–160).

[7] Bragg, D., and C. Hahn (spring 1989): “Material Requirements Planning and Purchasing.” Journal of Purchasing and Materials Management 25(1), (41–46).

[8] Chen, H., Chiang, R. H. L., & Storey, V. C. (2009). MIS quarterly special issue on business intelligence research. In MIS central (Vol. 2009), (webpage Handle http://www.misq.org).

[9] Chen, H., Chiang, R. H. L., & Storey, V. C. (2009). MIS quarterly special issue on business intelligence research. In MIS central (Vol. 2009), (webpage Handle http://www.misq.org).

[10] Ralph B. D’Agostino, Stephens, M.A (1986). "Tests for the Normal Distribution". In d Stephens, M.A. Goodness-of-Fit Techniques. New York: Marcel Dekker. ISBN 0-8247-7487-6.

[11] McCullagh, P. & Nelder, J. A. Generalized Linear Models:”Probability Distribution Fitting”, Chapman & Hall. Stokes (Second Edition), London.

[12] Hinkley D.V, Reid N, Snell E.J, (1988) Statistical Theory and Modeling: “Generalized Linear Models,”,(pp.167-500) , Chapman & Hall, London.
