International Journal of Computer Information Systems and Industrial Management Applications.
ISSN 2150-7988 Volume 14 (2022) pp. 048-059

Received: 28 October 2021; Accepted: 2 February, 2022; Published: 9 Apri, 2022

Data Analytics Model for Manufacturing Industry

Adriyendi[1], Ondra Eka Putra[2] and Sarjon Defit[3]

[1] Informatics Group, UIN Batusangkar 27211, Indonesia
elektronikpos@gmail.com

[2] Faculty of Computer Science, UPI YPTK Padang 25211, Indonesia
ondraekaputra17@gmail.com

[3] Faculty of Computer Science, UPI YPTK Padang 25211, Indonesia
sarjond@yahoo.co.uk

***Abstract*: Manufacturing Industry (MI) has problems with Value of Gross Output (VGO), Input Cost (IC), and Value Added (VA) in productivity, investment, trendline, and estimation. To overcome this problem, we carry out data analytic using descriptive model (K Means Clustering/KMC) for productivity, diagnostic model (Naïve Bayes Classifier/NBC) for investment, predictive model (Linear Regression/LR) for product trendlines, and prescriptive model (Monte Carlo Simulation/MCS) for input cost estimation. The results of KMC are 3 clusters. The results of NBC are VGO, VA, and IC influenced by number of establishments, workers engaged, and labor cost. The results of LR shows a trendline model. The results of MCS are 3 IC scenarios. We summarize that high productivity will open up new investment opportunities supported by a linear trend of value of gross output and value added with low input costs.**
Keyword

***Keywords*:** data analytics, clustering, classification, regression, simulation

I. Introduction

The Data Analytics Model (DAM) is very important for all industrial sectors including Manufacturing Industry (MI). DAM is useful for making decisions based on data about productivity, investment, trendline, and estimation on MI. With DAM, industry players can perform data analysis using historical data and real-time data to find trends and patterns [1]. These trends and patterns can help in predicting what is happening (currently) and what is likely to happen (future). DAM model development includes Descriptive Analytics (DsA), Diagnostic Analytics (DcA), Predictive Analytics (PdA), and Prescriptive Analytics (PcA) [2].

DsA is important because it helps understand how MI performs in the context of helping stakeholders interpret MI information. DcA is important to do to find out the causal factors or root cause analysis in exploring data and making correlations on MI. PdA plays a major role in detecting competitors, preventing fraud, optimizing marketing corporation strategies, predicting consumer behavior, product promotion, operations, and organizing the availability of raw materials and business development (investment opportunities) in MI. PdA can increase MI's competitive value. MI plays a broad role in efforts to increase the value of investment and exports which are the mainstay for accelerating economic growth. The PcA aims to determine what actions should be taken to address potential future problems (options and recommendations) or take full advantage of current promising trends (preferences) on MI. The important role of MI is to contribute to Gross Domestic Growth (GDP), taxation, and exports. MI's current mainstay sector is the pharmaceuticals, medicinal chemical and botanical products, food, beverage, printing, and reproduction of recorded media, computer, electronic, and optical products, furniture, wearing apparel, and textiles industries. MI's activities consistently provide broad effects, including increasing the value-added of domestic raw materials (number of establishments), absorption of local workers (workers engaged), and foreign exchange earnings (labor cost). This is related to the role of value of gross output, value-added, input cost on productivity, investment, trendline, estimation in the MI sector [3]. DAM in the MI business process needs to be carried out for future MI development strategies about how the value of gross output and value-added affects productivity, what is the correlation of input cost (IC), the value of gross output (VGO), and value-added (VA) with investment, how is the trend of value of gross output and value-added to productivity, and how to estimate input cost opportunities for investment in MI. Therefore, DsA by using clustering is needed to see the distribution and grouping of productivity. It is important to carry out a Diagnostic Analytics (DcA) using classification to see opportunities for MI investment development. It is also necessary to use Predictive Analytics (PdA) using regression to build a trendline model (calculating trends). Prescriptive Analytics (PcA) is necessary and important to provide options and recommendations (future possibilities, predicting potential outcomes, and making recommendations).

DAM has been proposed by many researchers. In paper [4], it discusses the combination of DAM with supervised and unsupervised learning (clustering and classification) techniques for electrical energy consumption.

In paper [5], it discusses descriptive analytics (data summarization, visualization, and dimension reduction) and predictive analytics (feature selection, feature extraction, clustering) for optimizing motorcycle routes in road traffic safety. In paper [6], it discusses predictive analytics (supervised and unsupervised learning) and prescriptive analytics (multi-objective optimization) in health care. In paper [7], it discusses descriptive analytics (Exploratory Data Analysis/EDA) techniques, predictive analytics (forecast and estimation), and prescriptive analytics (optimization) for Learner Management System (LMS) through web-based applications. In paper [8], it discusses descriptive analytics (visualization) and predictive analytics (trends and future possibilities) for business intelligence. In paper [9], it discusses analytical models ranging from descriptive analytics (network performance), diagnostic analytics (network anomalies), predictive analytics (network congestion), to prescriptive analytics (network expansion) in big data models. In paper [10], it discusses data analytics in state financial management using descriptive analytics (summarization), diagnostic analytics (detection), predictive analytics (identification), and prescriptive analytics (recommendation). In paper [11] it discusses the application of machine learning to descriptive analytics (unsupervised learning) and predictive analytics (supervised learning) for data processing. In paper [12] it discusses the use of DAM (form, structure, tool, sequence, usage, example) with descriptive analytics (report), diagnostic analytics (option), predictive analytics (scenario), and prescriptive analytics (preference) in big data analytics. The main contribution in this paper is summarized follows:

- We introduce data analytic by using the framework model for the manufacturing industry.
- Our model by using a combination of collaborative model in one simple simulation.
- This paper provides a comprehensive analysis of the aspects of descriptive, diagnostic, predictive, and prescriptive for manufacturing industry.

II. Materials and Methods

The Manufacturing Industry (MI) is a specific process (chemical or mechanical) that converts the basic good into finished or semi-finished goods. This process is also carried out to process low-value products into high-value products in economic activity. In this activity, the raw materials are provided by the supplier, while the workers only perform the processing in exchange for compensation. MI are grouped into 4 groups based on the number of workers, namely: large industries (100 workers or more), medium/medium industries (20-99 workers), small industries (5-19 workers), and micro industries (1-4 workers) [3]. Input Costs (IC) are costs for raw materials and supporting materials, fuel, and services (building rent and non-industrial services). Value of Gross Output (VGO) is value of the results of production, electricity sold, buying and selling profits, change in stocks, and other incomes. Value Added (VA) is defined a subtraction from output to input. Labor cost (LC) is defined as compensation for workers (money and goods). Labor cost included salary, overtime pay, wage, a bonus in cash and goods, pension funds, allowance (social and accident), and others [3].

Data Analysis Model (DAM) is the process of analyzing, interpreting, and visualizing data using various tools, techniques, and specific methods. The processed data consists of quantitative data (numbers, ratios, percentages) and qualitative data (features, characteristics, attributes). Techniques, methods, tools are used for structured quantitative data in rows and columns of a spreadsheet. Techniques, methods, tools used for unstructured qualitative data because it does not have a pre-configured format using intelligent computing (artificial intelligence).

The method for DAM used is Statistical Analysis (SA). SA is divided into two, namely descriptive analysis and inferential analysis. Descriptive Analytics (DsA) is an analytical process that describes the distribution of data. The data distribution in question is the measurement of the central tendency and the measurement of shape. DsA deploys visualization to a great extent and convert the data into useful information for analyzing business decisions and outcomes [13].

DsA with numerical measures for centralization measure and spread size. The measure of centering uses the mean, median, and mode. The size of the spread uses the range, variance, and standard deviation. There are four main measures in DsA, namely frequency, tendency, centrality, distribution, and position [14].

Diagnostic Analytics (DcA) is the process of analyzing to find the cause of the emergence of data. DcA is very essential since it gives detailed information about why certain things happened [15]. This can be done after collecting data and aggregating information through DsA. DcA can assist in identifying deviations from the mean, separating patterns, and finding data relationships. DcA also helps in understanding why something happened to the past data [16].

Predictive Analytics (PdA) is a process that uses predictions to find out future events (potential and opportunities that will occur or may occur) [17]. PdA uses techniques such as data mining, statistics, machine learning to analyze current data and make forecasts for the future. This helps find patterns in past data for the identification of risks and opportunities [18].

Prescriptive Analytics (PcA) aims to determine what actions should be taken to address potential future problems or take full advantage of current promising trends. PcA uses advanced tools and technologies, such as machine learning, business rules, and specific algorithms. Machine learning algorithms are capable of capturing the potential correlations amongst information [19]. PcA is done to predict future events and estimate the potential that can be used by suggesting options [20].

We do descriptive analytics by using K Means Clustering (KMC). KMC is a technique to group the similar data points into the same cluster. It’s simple and easy to be solved [21]. We do diagnostic analytics using Naïve Bayes Classifier (NBC). NBC is simple to implement and in-sensitive to irrelevant data [22]. We do predictive analytics using Linear Regression (LR). LR is the simplest type of regression and plays a central role in statistics [23]. We do prescriptive analytics using Monte Carlo Simulation (MCS). MCS is used to approximate the hypervolume. Its performance highly depends on the distribution of data points in the high dimensional space [24].

Dimension/ Characteristic	Dimension 1	Dimension 2	Dimension 3	Dimension 4
Analytics	Descriptive	Diagnostic	Predictive	Prescriptive
Model	Clustering	Classification	Regression	Simulation
Technique	K-Means Clustering	Naïve Bayes Classification	Linear Regression	Monte Carlo Simulation
Input	Value of Gross Output (VGO), Value Added (VA)	Value of Gross Output (VGO), Value Added (VA), Input Cost (IC)	Value of Gross Output (VGO), Value Added (VA)	Input Cost (IC)
Process	Dataset, Centroid, Iteration, Ratio, Cluster	Dataset, Prior Probability, Posterior Probability, Maximum Posterior, Class	Dataset, Independent Variable, Dependent Variable, Estimation, Trendline	Dataset, Distribution Probability, Distribution Cumulative, Random Number, Simulation
Target	Productivity	Investment	Trendline	Estimation
Output	(High, ~~Middle~~, ~~Low~~)	Investment (Yes, ~~No~~)	Trendline (Increase, ~~Decrease~~)	Estimation (Feasible, ~~Infeasible~~)
Research Question	How to clustering productivity in manufacturing industry	How to classify investment in manufacturing industry	How to predict trendline in manufacturing industry	How to estimate input cost in manufacturing industry

Table 1. Methods..

K-Means Clustering [25]

Steps 1: Clusters the data into groups from dataset.
Steps 2: Select k points at random as centroid (cluster center).
Steps 3: Assign all points to the nearest centroid according to the Euclidean Distance (ED) with the formula:

$$d(p, q) = \sqrt{+ (q_1 - p_1)^2 + (q_2 - p_2)^2} \tag{1}$$

Let point p have coordinate (p_1, p_2) and let point q have coordinates (q_1, q_2), then the distance between p and q is d (p, q).
Steps 4: Calculate the centroid (iteration) in each cluster.
Steps 5: Repeat step 2, 3, and 4 until data points convergence.

Naïve Bayes Classifier [26]

Steps 1: Separate by class.
Steps 2: Determine dataset (data training and data testing).
Steps 3: Summarize data by class.
Steps 4: Probability density function with the formula:

$$P\,(A \mid B) = \frac{P\,(B \mid A)\,.\;P\,(A)}{P\,(B)} \tag{2}$$

P (A|B) is posterior probability, P (B|A) is likelihood, P (A) is class prior probability, and P (B) is predictor prior probability.
Steps 5: Class probability based on maximum value for probability.

Linear Regression [27]

Steps 1: Preparing dataset.
Steps 2: Relationship between two variables (dependent and independent) with the formula:

$$Y = a + b(X) \tag{3}$$

Where Y is the dependent variable, X is the independent variable, b is the slope of the line, and a is the y-intercept.
Steps 3: use the formula (3) to find a and b.
Steps 4: estimating the model.
Steps 5: fitting the trendline.

Monte Carlo Simulation [28]

Steps 1: Preparing dataset.
Steps 2: Determine distribution of probability.
Steps 3: Determine distribution of cumulative.
Steps 4: Generate random number with the formula:

$$X_{n+1} = (a\ X_n + b)\ mod\ m \tag{4}$$

Steps 5: Aggregate the result of simulation.

Table 2. Techniques

Based on the facts of the past, data modeling supplies some information and obtain knowledge from its [29]. The adaptive modelling with considers optimal solution to assists decision-maker in the manufacturing industry [30].

III. Result and Discussion

A. Data Analytics Model

Data Analytic Model (DAM) is the process of analyzing, interpreting, and visualizing data using various tools, techniques, and specific methods. The processed DAM by using descriptive, diagnostic, predictive, prescriptive analytic based on framework model is shown in Figure 1.

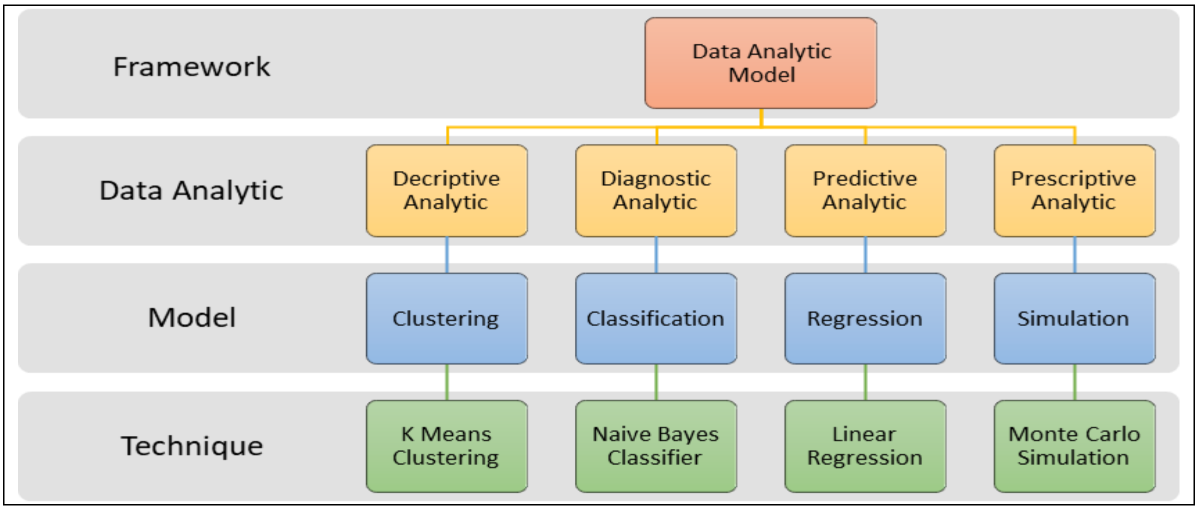

Figure 1. Framework model

A. Descriptive Analytic

Descriptive Analytic (DsA) by using K Means Clustering (KMC). Dataset can be seen in Table 3.

Manufacturing Industry Group (MIG)	Value of Gross Output (VGO)	Value Added (VA)
A (printing and reproduction of recorded media)	1	3
B (food)	3	3
C (beverages)	4	3
D (pharmaceuticals, medicinal, chemical and botanical products)	5	3
E (computer, electronic, and optical products)	1	2
F (wearing apparel)	4	2
G (furniture)	1	1

Table 3. KMC dataset.

Centroid is determined by random (B, E, F) based on *k* value = 3. Centroid is shown in Table 4.

Centroid	Value of Gross Output (VGO)	Value Added (VA)
B	3	3
E	1	2
F	4	2

Table 4. Centroid.

Iteration stopped when the current ratio ≤ prior ratio (iteration 3) shown in Table 5.

Ratio 1 (iteration 1)					Ratio n (iteration n)					
Centroid		Distance	Current Ratio	Prior Ratio	Centroid		Distance	Current Ratio		Prior Ratio
C1	C2	2,2360	WCV/BCV	-	C1	C2	3,1622	WC/BCV	-	C1
C1	C3	1,4242	6,6502/9	-	C1	C3	1,5811	7,2929/5	-	C1
C2	C3	3	*0,7379*	*0*	C2	C3	2,5495	*1,4585*	*1,4585*	C2
BCV	sum	6,6502	WCV=SMD	-	BCV	sum	7,2929	WCV=SMD	-	BCV

Table 5. Ratio.

Iteration based on centroid (C1, C2, C3), member cluster (MC), and distance (minimum distance/MD, square of minimum distance/SMD). Clustering with minimum distance based on Euclidean Distance (ED). Iteration can be seen in Table 6.

Centroid	**Cluster**			**MC**	**ED**		**New Cluster**					
Centroid	C1	C2	C3	MC	MD	SMD	C1		C2		C3	
MIG	B	E	F	MC	MD	SMD	VGO	VA	VGO	VA	VGO	VA
A	2	1	3,1622	C2	1	1	-	-	1	3	-	-
B	0	2,2360	1,4142	C1	0	0	3	3	-	-	-	-
C	1	3,1622	1	C1	1	1	4	3	-	-	-	-
D	2	4,1231	2,2360	C1	2	4	5	3	-	-	-	-
E	2,2360	0	3	C2	0	0	-	-	1	2	-	-
F	1,4142	3	0	C3	0	0	-	-	-	-	4	2
G	2,8284	1	3,1622	C2	1	1	-	-	1	1	-	-
H	2	2,2360	1,4142	C3	1,4142	2	-	-	-	-	3	1
ED = SQRT((C6-$C17) ^2+($D$6-$D17) ^2) ED = SQRT ((1-3) ^2 + (3-3) ^2) = 2					WCV	9 (sum)	4 (avg)	3 (avg)	1 (avg)	2 (avg)	3,5 (avg)	1,5 (avg)

Table 6. Iteration.

The new centroid shown in Figure 2.

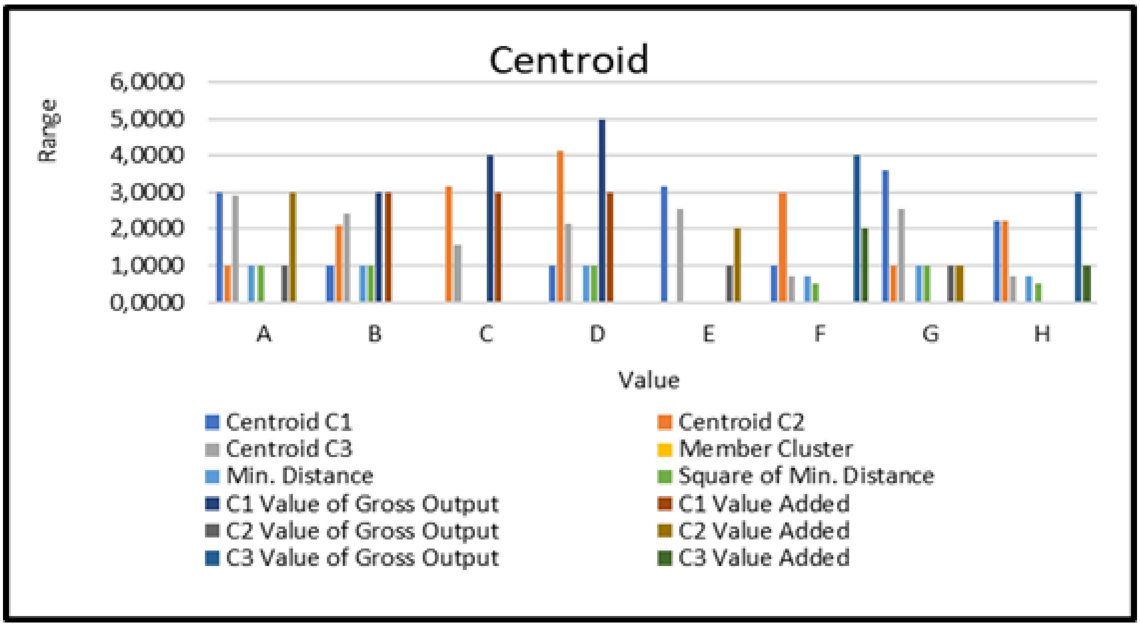

Figure 2. Centroid

The results of the descriptive analysis showed that there was a grouping of productivity into 3 clusters (cluster 1 = high, cluster 2 = middle, cluster 3 = low). Cluster 1 is high productivity which consists of the food, beverages, and pharmacy (medicinal, chemical, and botanical products) industry groups. Cluster 2 is a high productivity group consisting of the printing industry (printing and reproduction of recorded media), computer (computer, electronic, and optical products), and furniture. Cluster 3 is high productivity which consists of the wearing apparel and textiles industry group. This is due to the difference in the distribution of values from the value of gross output and value added to productivity. The result of clustering can be seen in Figure 3.

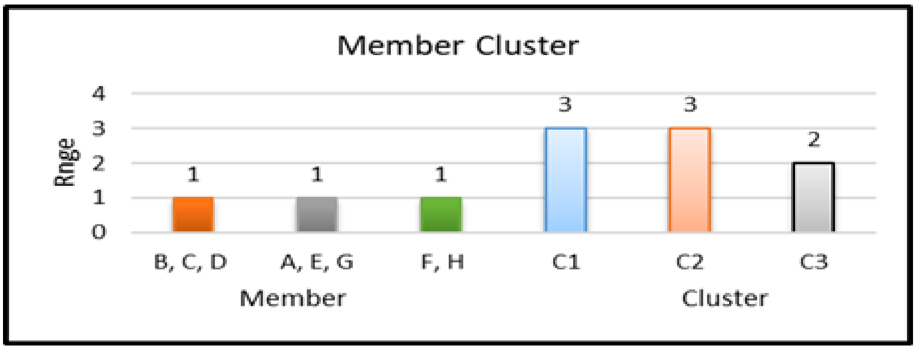

Figure 3. Clustering

C. Diagnostic Analytic

Diagnostic Analytic (DcA) by using Naïve Bayes Classification (NBC). Dataset with attributes: Number of Establishments (NE), Workers Engaged (WE), Labor Cost (LC), Manufacturing Industry Category (MIC), Value of Gross Output (VGA), Input Cost (IC), Value Added (VA), and Probability of Investment (PI). Data training display in Table 7.

No	**MIC**		**Input**		**Target**	No	**MIC**	**Input**	**Target**		
Id	MIC	VGA	IC	VA	PI	Id	MIC	VGA	IC	VA	PI
1	NE	Decrease	Middle	Big	No	11	WE	Increase	Middle	Small	Yes
2	NE	Decrease	Low	Big	No	12	WE	Increase	Low	Big	Yes
3	NE	Decrease	Low	Small	Yes	13	LC	Decrease	High	Big	No
4	NE	Increase	Middle	Big	Yes	14	LC	Decrease	High	Small	Yes
5	NE	Increase	Low	Big	No	15	LC	Decrease	Low	Small	No
6	WE	Decrease	High	Big	No	16	LC	Increase	High	Big	Yes
7	WE	Decrease	Middle	Big	No	17	LC	Increase	Middle	Big	No
8	WE	Decrease	Low	Small	Yes	18	LC	Increase	Middle	Small	Yes
9	WE	Increase	High	Big	Yes	19	LC	Increase	Low	Big	No
10	WE	Increase	Middle	Big	Yes	-	-	-	-	-	

Table 7. Data training.

Data Testing can be seen in Table 8.

No	**MIC**		**Input**		**Target**
Id	MIC	Value of Gross Output	Input Cost	Value Added	Probability of Investment
20	Workers Engaged	Increase	Low	Small	?

Table 8. Data testing.

The result of calculation for Prior Probability, Posterior Probability, Maximum Posterior, and Class displays in Table 9.

Attribute	**MIC**	**PI**	**WE (Yes, No)**	**PI (Yes, No)**	**WE/PI (Prior Probability)**
MI Group	WE	Yes	5 (Yes)	10 (Yes	0,5000
	WE	No	2 (No)	9 (No)	0,2222
VGO	Increase	Yes	7 (Yes)	10 (Yes)	0,7000
	Increase	No	3 (No)	9 (No)	0.3333
IC	Low	Yes	3 (Yes)	10 (Yes)	0.3333
	Low	No	4 (No)	9 (No)	0.4444
VA	Small	Yes	5 (Yes)	10 (Yes)	0,5000
	Small	No	1 (No)	9 (No)	0,1111
Attribute	Value	PI (Yes, No)	Count (Yes, No)	Posterior Probability	-
PI	Yes	0,0525 (Yes)	10 (Yes)	0,5250	-
PI	No	0,0036 (No)	9 (No)	0,0329	-

Table 9. NBC calculation.

If PI Yes > PI No, then maximum posterior is PI = Yes. This result became class based on the classifier. A classifier is shown in Figure 4.

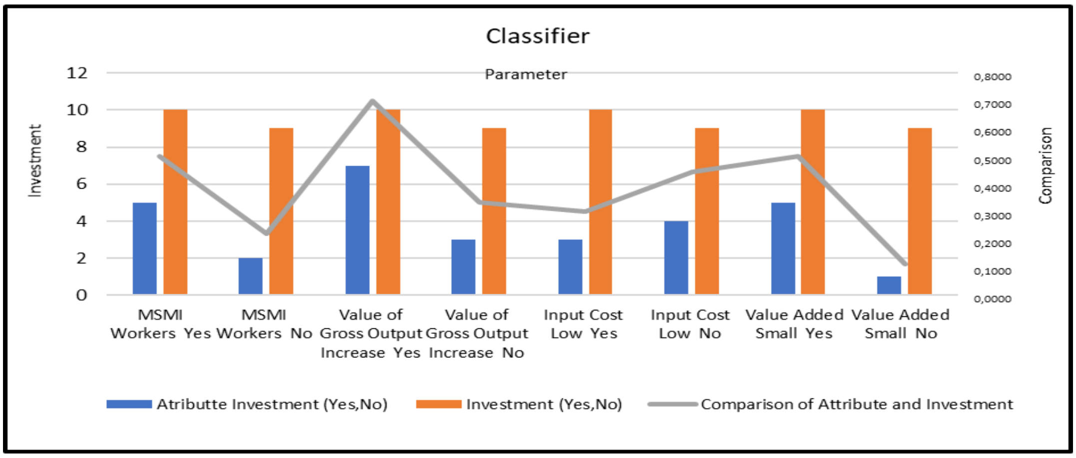

Figure 4. Classifier

Classifier shows parameter classifier, investment, attribute investment, and comparison of attribute and investment.

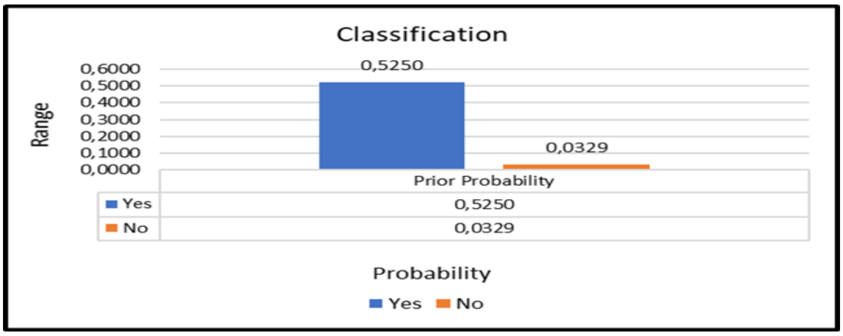

Figure 5. Classification

The results of the diagnostic analysis show that the Number of Establishment (NE), Workers Engaged (WE), and Labor Cost (LC) greatly determine the value of the value of gross output, value added, and input cost. This change in value greatly influences decisions about investment probability (PI). It can be proven that WE with a value of VGO = Increase, IC = Low, and VA = Small will produce PI = Yes with an opportunity value of 0.5250 or 52.50%.

D. Predictive Analytic

Predictive Analytic (PdA) by using Linear Regression (LR). Estimation can be seen in Table 10.

Period	Independent Variable (x)	Dependent Variable (y)	Estimate (y')	Estimation Error (y-y')	Square of Error (y-y')
1	4	6	6,07	0,07	0,00
2	4	5	6,07	1,07	1,14
2	4	7	6,07	0,93	0,87
4	5	6	6,97	0,97	0,93
5	5	8	6,97	1,03	1,07
6	5	7	6,97	0,03	0,00
7	6	7	7,87	0,87	0,75
8	6	8	7,87	0,13	0,02
9	6	9	7,87	1,13	1,28
10	7	9	8,77	0,23	0,05
11	7	8	8,77	0,77	0,59
12	7	9	8,77	0,23	0,05
Linear Regression y = a + bx a = 2,47 b = 0,90			Estimator if x = 7 then y (x) = a + bx y (x) = 2,47 + 0,90 (7) = 8,77		

Table 10. Estimation

In Table 16, the standard error of estimation (estimation error) is a measure of the distribution/scatter of the observed values around the regression line (a measure of the accuracy of the estimation). The standard error of the regression coefficient (square of error) is to measure the magnitude of the deviation from each regression coefficient. The lower the standard error, the more the variable plays a role in the model and vice versa. Data visualization for estimation showed in Figure 6.

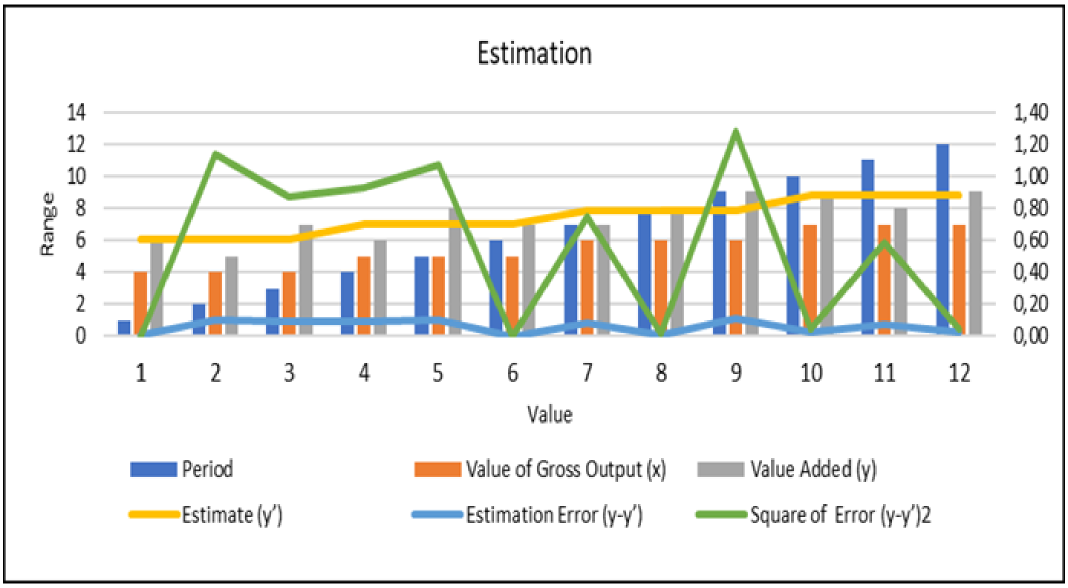

Figure 6. Estimation

The resulting linear model is y(x) = 2.47 + 0.90x, while the prediction of productivity growth for output value 7 is y (7) = 2.47 + 0.90*7 = 8.77. If the Value of Gross Output or VGO (x) = 7 (e. g. 7, the highest of VGO is taken from Table 10), with an intercept constant (a) is 2,47 and a regression coefficient (b) is 0,90, the Estimate (y') value will be increased by 8,77. It means that the VGO if you want to increase it by 7, the Estimate (y') value for productivity will increase by 8,77shown in Figure 7.

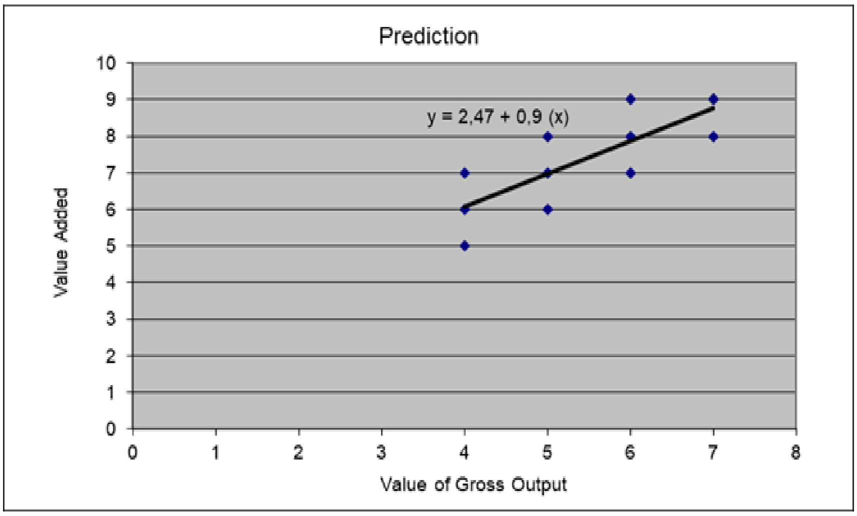

Figure 7. Prediction

The results of the predictive analysis show a trendline with a linear equation model, namely (y(x) = 2.47 + 0.90x) for the value of gross output and value added. This means that the value of gross output and value added play a significant role in the linear model.

E. Prescriptive Analytic

Prescriptive Analytic (PcA). PcA by using Monte Carlo Simulation (MCS). Distribution shows in Table 11.

Input Cost	Distribution (Frequency/Total)	Distribution of Probability	Distribution of Cumulative	Interval (Random Number)
1	6/80	0,08	0,08	01 … 08
2	7/80	0,09	0,16	09 … 17
3	5/80	0,06	0,23	18 … 25
4	8/80	0,10	0,33	26 … 33
5	5/80	0,06	0,39	34 … 42
6	6/80	0,08	0,46	43 … 50
7	5/80	0,06	0,53	51 … 58
8	9/80	0,11	0,64	59 … 67
9	6/80	0,08	0,71	68 … 75
10	8/80	0,10	0,81	76 … 83
11	7/80	0,09	0,90	84 … 91
12	8/80	0,10	1,00	92 … 100

Table 11. Distribution.

Random Number based on interval (01…100) by using Excel formula = RANDBETWEEN (1;100) display in Table 12.

No	Random Number 1	Random Number 2	Random Number 3	Random Number 4	Random Number 5
1	36	63	17	68	88
2	52	69	88	94	74
3	31	33	5	97	3
4	53	56	49	41	79
5	27	83	91	53	71
6	90	72	32	52	77
7	14	54	65	12	91
8	40	34	39	77	66
9	41	16	91	99	75
10	61	30	4	58	96
11	86	90	21	23	9
12	23	79	20	33	73

Table 12. Random Number.

Simulation by Random Number (RN1, RN2, RN3) with scenario can be seen in Table 13.

Input Cost	RN 1	RN 2	RN 3	Scenario 1	Scenario 2	Scenario 3
1	36	63	17	5	8	2
2	52	69	88	7	9	11
3	31	33	5	4	4	1
4	53	56	49	7	7	6
5	27	83	91	4	10	11
6	90	72	32	11	9	4
7	14	54	65	2	7	8
8	40	34	39	5	5	5
9	41	16	91	5	2	11
10	61	30	4	8	4	1
11	86	90	21	11	11	3
12	23	79	20	3	10	3
Total				72	86	66
Average				6,00	7,17	5,50

Table 13. Simulation.

Based on Table 13, there are 3 scenarios from simulation. Scenario 1 with average of input cost is 6,00. Scenario 2 with average of input cost is 7,17. Scenario 3 with average of input cost is 5,50. The best value of input cost is 5,50 based on minimum value from the simulation.

The results of the prescriptive analysis show a simulation that produces 3 scenarios from the input cost. The three scenarios are input costs with an average value of 6.00 for scenario 1, an average value of 7.17 for scenario 2, and an average value of 5.50 for scenario 3. The best value is input cost with the lowest average of 5.50.

Estimate input cost based on distribution, Random Number (RN) 3, and Scenario 3 shown in Figure 8. Data visualization for simulation based on Table 13 is displayed in Figure 9. Comparison for data analytic model in display in Table 14.

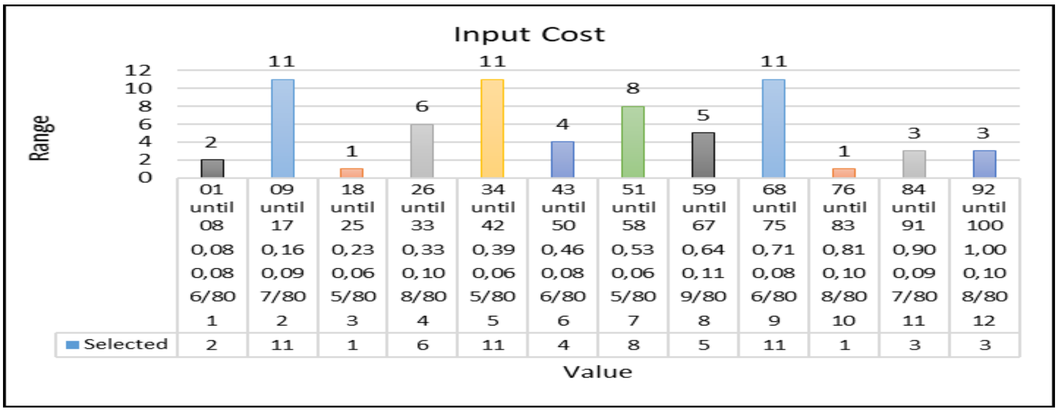

Figure 8. Input cost

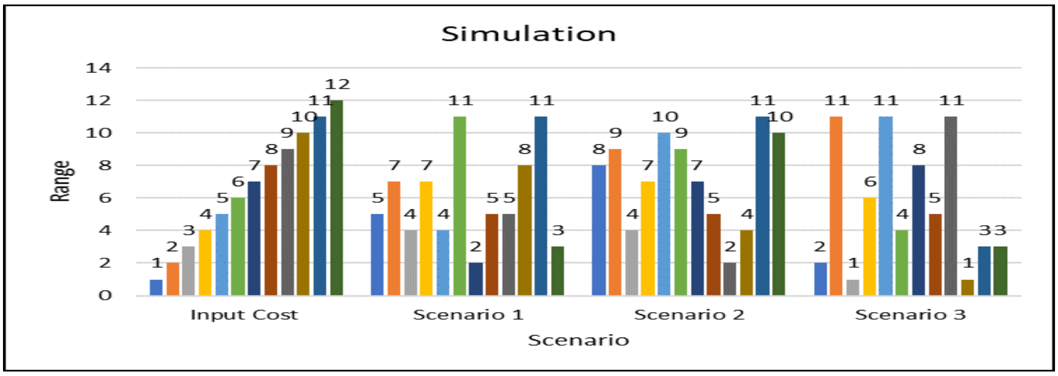

Figure 9. Simulation

Model/ Analytic	**Descriptive**	**Diagnostic**	**Predictive**	**Prescriptive**
Form	What	Why	If	How
Description	What happened about productivity in manufacturing industry?	Why did it happen about investment in manufacturing industry?	What will happen about trendline product in manufacturing industry?	How can we make it happen about estimation cost in manufacturing industry?
Time	Past	Past	Future	Future
Shows	Visualization past data	Identify correlation	Forecast future	Suggest actions
Analysis	Analyzes historical data *to learn* about what is happening in past and present	Uses data analysis *to answer* why a problem is occurring	Uses past and present data *to forecast* and create models to make prediction about the future	Uses data model forecasting *to test* the likely outcome of different actions based on data
Results	The results of the descriptive analysis showed that there was a grouping of productivity into 3 clusters (high, middle, low). This is due to the difference in the distribution of values from the value of gross output and value added to productivity.	The results of the diagnostic analysis show that the value of gross output, value added, and input cost influenced by the Number of Establishment (NE), Workers Engaged (WE), and Labor Cost (LC) toward investment.	The results of the predictive analysis show a trendline model: y(x) = 2.47 + 0 .90x). This means that the value of gross output and value added play a significant role in the linear model.	The results of the prescriptive analysis show a simulation that produces 3 scenarios from the input cost. The best value is input cost with the lowest average of 5.50.
Preference	value of gross output and value added (normative)	value of gross output, value added, and input cost (corrective)	value of gross output and value added (innovative)	input cost (adaptive)

Table 14. Comparison.

IV. Conclusion

We have discussed the data analytics model for manufacturing industry. We do a simulation, create a framework, and data processing by using data analytics models. This work provides a comprehensive analysis of the aspects of descriptive, diagnostic, predictive, and prescriptive for manufacturing industry. We conclude that high productivity will open up new investment opportunities supported by a linear trend of value of gross output and value added with low input costs. The interpretation of this is if productivity is high, investment is yes, trendline is increase, estimation is feasible, then framework model is prospective. The implication is that the data analysis model can be a source or supporting material for making transformative and futuristic manufacturing industry policies. The next research opportunity is the development of a prescriptive model with preferences and recommendations using deep learning in the manufacturing industry.

References

[1] L. Tang, Y. Meng. "Data Analytics and Optimization for Smart Industry", *Frontiers of Engineering Management*, VIII (2), pp. 157-171, 2021.

[2] E.T. Tosida, F.D. Wihartiko, I. Hermadi, Y. Nurhadriyani, Feriadi. "Big Data Management Model for Rice Commodity towards National Food Policy", Journal of RESTI, IV (3), pp. 142-154, 2020.

[3] BPS Indonesia, "The Annual Large and Medium Manufacturing Establishment Survey", in the *BPS Annual Report,* Jakarta, 2020.

[4] R. Kanagaraj, N. Rajkumar, K. Srinivasan. "Multiclass Normalized Clustering and Classification Model for Electricity Consumption Data Analysis in Machine Learning Techniques", *Journal of Ambient Intelligence and Humanized Computing,* XII (5), pp. 5093-5103, 2021.

[5] A. Mehdizadeh, A.M. Cai, Q. Hui, M.A.A. Yazdi, N.M. Kalejahi, A. Vinel, S.R. Rigdon, K.C. Davis, FM. Megahed. "A Review Data Analytic Applications in Road Traffic Safety", *Sensor,* XX (4), pp. 1-24, 2020.

[6] J. Lopez, T. Guimaraes, M.F. Santos. "Predictive and Prescriptive Analytics in Healthcare A Survey", *Procedia Computer Science,* (170), pp. 1029-1034, 2020.

[7] S. Islam, H. Mahmud. "An Intelligence Learner Management System using Learning Analytics and Machine Learning", in *International Conference on Education Technology and Computers,* pp. 1200-125, 2020.

[8] J. Hlavac, J. Stefanovic. "Machine Learning and Business Intelligence of From Descriptive Analytics to Predictive Analytics", *Cybernetics & Informatics,* pp. 1-4, 2020.

[9] S. Rizal. "Development of Big Data Analytics Model", *Information Technology Engineering Journals,* IV (1), pp. 14-25, 2019.

[10] B.A.A. Uliansyah, A. Darono, Febrian. "Data Analytic in State Finance Management", *Teknika,* X (2), pp. 160-168, 2020.

[11] P. Santos, H. Abijono, N.L. Anggreini. "Supervised Learning and Unsupervised Learning in Data Processing", *Saintek,* IV (2), pp. 315-318, 2021.

[12] A. Manshur. "One Data, Big Data, and Data Analytics", *Bappenas Working Papers,* IV (1), pp. 30-46, 2021.

[13] W. Raghupathi, V. Raghupathi. "Contemporary Business Analytic: An Overview", *Data,* VI (86), pp. 1-11, 2021.

[14] D. Bertsimasa, N. Kallus. "From Predictive to Prescriptive Analytics", *Management Science,* 66 (3), pp. 1025-1044, 2020.

[15] W.T. Lai, R.B. Chen. "A Review of Bayesian Group Selection Approach for Linear Regression", *WIREs Computational Statistics,* XIII (4), pp. 1-22, 2020.

[16] D.H. Maulud, A.M. Abdulazeez. "A Review on Linear Regression Comprehensive in Machine Learning", *Journal of Applied Science and Technology Trends,* I (4), pp. 140-147, 2020.

[17] A. Alghamdi, A. Baz, T.Alsubait, H. Alhakami. "Healthcare Analytic: A Comprehensive Review", *Engineering Technology & Applied Science Research,* XI (1), pp. 6650-6655, 2021.

[18] J. Tinoco, M. Parente, A.G. Corella, P. Cortez, D.Toll. "Predictive and Prescriptive Analytics in Transportation Geotechnics", *Transportation Engineering,* V (100074), pp. 1-13, 2021.

[19] A. Adriyendi, Y. Melia. "Classification using Naïve Bayes and K-Nearest Neighbor on Information Technology", *Journal of Technology and Business Information Systems,* II (1), pp. 99-107, 2020.

[20] M.O. Callaghan, D.K. Balasubramanian, F. Behgounia, N. Shukla, P. Jayaswal, B. Zohuri. "Prescriptive Analytic for Airline Price Optimization Driven by Artificial Intelligence, Machine Learning, and Excel Spreadsheet Algorithm", *Aditum Journal of Clinical and Biomedical Research,* II (3), pp. 1-8, 2021.

[21] Z.Z. Long, G. Xu, J. Du, H. Zu, T. Yan, Y.F. Yu. "Flexible Subspace Clustering: A Joint Feature Selection and K-Means Clustering Framework", *Big Data Research,* XXIII (100170), pp. 1-9, 2021.

[22] H. Hubert, P. Phoenix, R. Sudaryono, D. Suhartono. "Classification Promotion Images using Optical Character Recognition and Naïve Bayes Classifier", *Procedia Computer Science,* (179), 498-506, 2021.

[23] W.H. Laverty, I.W. Kelly. "Exploring the Effect Assumption Violation on Simple Linear Regression and Correlation using Excel", *American Journal of Theoretical and Applied Statistics,* (10), pp. 194-201, 2021.

[24] J.D. Deng, Q. Zhang. "Combining Simple and Adaptive Monte Carlo Methods for Approximating Hypervolume", *IEEE Transaction on Evolutionary Computation,* XXIV (5), pp. 896-907, 2020.

[25] K. Golalipur, E. Akbari, S. Hamidi, M. Lee, R. Enayatifar. "From Clustering to Clustering Ensemble Selection", *Engineering Applications of Artificial Intelligence,* 104 (104388), pp. 1-21, 2021.

[26] S. Sendari, I.A.E. Zaeni, D.C. Lestari, H.P. Hariyadi, "Opinion Analysis for Emotional Classification on Emoji Tweets using Naïve Bayes Algorithm", *Knowledge Engineering and Data Science,* III (1), pp. 50-59, 2021.

[27] A.K. Subramanian, A. Samanta, S. Manickam, A. Kumar, S. Shiaeles, A. Mahendran. "Linear Regression Trust Management System for IoT Systems", *Cybernetics and Information Technologies,* XXI (4), pp. 15-27, 2021.

[28] R. Prawita, S. Sumijan, G.W. Nurachyo. "Monte Carlo Simulation for Stock Control in Office Stationery", *Journal of Economic Business and Informatics,* III (2), pp. 72-77, 2021.

[29] T. Sangeetha, G. Mary A. "A Study on Different Methods of Outlier Detection Algorithms in Data Mining", *International Journal of Computer Information Systems and Industrial Management Applications,* (12), pp. 398-411, 2020.

[30] M. Asrol, D. Delfitriani. "An Optimize and Adaptive Modeling for Sugarcane Harvesting and Transportation", *International Journal of Computer Information Systems and Industrial Management Applications,* (13), pp. 233-241, 2021.

Author Biographies

Adriyendi. He is a lecturer (assistant professor) at UIN Batusangkar, Indonesia. His research interests include artificial intelligence, machine learning, decision support system, social computing, food computing, data mining, data science, and information engineering. He can be reached at *elektronikpos@gmail.com*.

Ondra Eka Putra. He is a lecturer (assistant professor) at UPI YPTK Padang, Indonesia. His research interests include artificial intelligence, machine learning, decision support system, software engineering, computer system, and computer vision. He can be reached at *ondraekaputra17@gmail.com*.

Sarjon Defit. He is a lecturer (professor) at UPI YPTK Padang, Indonesia. His research interests include artificial intelligence, decision support systems, machine learning, data mining, and database. He can be reached at *sarjond@yahoo.co.uk*.