IT Journal Research and Development (ITJRD)
Vol.8, No.1, August 2023, E-ISSN : 2528-4053 | P-ISSN : 2528-4061
DOI : 10.25299/itjrd.2022.12202

Forecasting Simcard Demand Using Linear Regression Method

Monica Sitompul, Mhd. Arief Hasan, Mariza Devega
Informatics Engineering Study Program, Faculty of Computer Science, Lancang Kuning University
monicasitompul14@gmail.com, m.arif@unilak.ac.id, marizadevega@unilak.ac.id

Article Info

Article history:

Received Jul 27, 2023
Revised Aug 08, 2023
Accepted Okt 02, 2023

Keyword:

Forecasting
Demand
Linear regression
Python

ABSTRACT

This research aims to analyze and forecast the monthly demand for SIM cards at Sinar X-Sis Cell Pekanbaru for the period of 25 to 48 months ahead, specifically from January 2022 to December 2023. The analysis is based on 24 months of sales data spanning from January 2020 to December 2021. The methodology employed is Linear Regression, processed through both manual calculations and testing using Google Colaboratory. The findings indicate an increase in demand for the KP03 package card, whereas the KP01, KP02, KP04, and KP05 SIM cards are projected to experience a decrease in demand in the future. The implementation of the Linear Regression method involved testing and forecasting calculations using Google Colaboratory tools, resulting in graphical representations that illustrate the anticipated increases and decreases in sales for each SIM card from January 2022 to December 2023. With these improvements, the abstract provides a clearer and more detailed overview of the research's objectives, methods, and findings.

Corresponding Author:

Mhd Arief Hasan
Informatics Engineering Study Program, Faculty of Computer Science, Lancang Kuning University
Jln. Yos Sudarso, KM. 8. Rumbai, Pekanbaru, Riau
Email: m.arif@unilak.ac.id

1. INTRODUCTION

Currently the development of information technology is growing very rapidly in various aspects of life. The development of information technology can be felt both in the aspects of communication, economy, education, government and others. With the development of information technology, it makes it easier to carry out life activities. Needs that continue to grow or increase create innovation or a growing mindset in developing something new to support these needs. In line with the growing needs, the development of information technology is also growing.

The business of selling simcards can be said to be quite a profitable business but has a very high risk because many internet service providers compete in providing good internet services in accessing information and carrying out cheap package rates and supported by good marketing which has an effect on the number of users in a certain area. provider. The number of sales that occur in each type of package card varies and tends to fluctuate in sales, making it difficult for Sinar E-Xis Cell owners to determine the number of reorders for the internet package because there is no forecast to determine how much to buy. So a forecasting system is needed that can make it easier for the owner to determine how much to provide so as not to disappoint customers[1]. Production forecasting is a form of decision making that is used as the basis for many manufacturing and service industries. Therefore, companies that are able to produce products on time and in the right amount are companies that are able to survive in competition. This demand forecasting is used to predict the demand for products that are independent (independent), such as forecasting finished forms. The linear regression method is a statistical tool that is used to determine between one or several variables for one variable. The benefits of linear regression include regression analysis which is more accurate in conducting correlation analysis, because the analysis is difficult in making changes to one variable with another variable. (slop) can be done. In regression analysis, forecasting or forecasting the dependent variable on the value of the independent variable is more accurate. In addition, this analysis can determine which dependent variable is positive and which is negative and to predict the value of the dependent variable if the value of the independent variable experiences increase or decrease and the independent variable. The data used is interval or ratio scale data[2], [3].

Previous research has endeavored to forecast SIM card demand using a variety of methods. The strengths of prior work include the development of reasonably accurate demand prediction models that have provided valuable insights to the industry. These methods have laid a solid foundation for understanding demand trends. However, shortcomings in some studies are evident, as they may have limited scope, failing to encompass various factors influencing SIM card demand, such as demographic aspects and evolving consumer behavior[4]–[6]. Additionally, some studies may not adequately consider changing technology trends and recent developments in the telecommunications industry, potentially affecting the relevance of their findings. In the proposed work, the choice of the algorithm, specifically linear regression, is justified due to its capability to model the relationship between variables, making it suitable for predicting SIM card demand based on sales data. The use of linear regression allows us to capture and analyze the impact of key variables on demand, providing a transparent and interpretable approach for forecasting[7]–[9].

2. METHODS USED

2.1 Research Stages

In this research, we adopt a linear regression approach as the primary algorithm to forecast SIM card demand based on sales data. The methodological steps we followed are as follows:

1. **Data Collection**: We collected sales data of SIM cards over the past two years from SINAR E-XIX CELL, including variables such as the number of SIM cards sold, the type of card package, package prices, and other relevant factors affecting demand.
2. **Data Cleaning**: The collected data was analyzed and cleaned to eliminate incomplete or outlier data that might affect the accuracy of the model. Cleaned data was used in subsequent analyses[10].
3. **Variable Selection**: We selected the number of SIM cards sold as the dependent variable, while independent variables included the type of card package, package prices, and other relevant factors that might influence demand.
4. **Linear Regression Model**: We employed a linear regression model to establish the relationship between independent and dependent variables. This model is described by a mathematical equation that allows us to forecast SIM card demand based on independent variables.
5. **Model Evaluation**: To measure the accuracy of the developed linear regression model, we conducted evaluations using standard evaluation metrics such as Root Mean Squared Error (RMSE) and R-squared (R^2).
6. **Comparison with Existing Methods**: To assess the effectiveness of our proposed system, we compared its forecasting performance with existing methods commonly used in the industry for SIM card demand prediction. We considered alternative approaches such as time series forecasting and machine learning algorithms to evaluate the accuracy and suitability of our linear regression-based model.
7. **Python Implementation**: All analyses and computations were carried out using the Python programming language, along with statistical libraries such as NumPy, pandas, and scikit-learn.

By incorporating this comparison with existing methods, we provide a comprehensive evaluation of the proposed system's performance and its advantages over conventional approaches to SIM card demand forecasting. [11]–[13]By following these methodological steps, we were able to design and implement a linear regression algorithm for effectively forecasting SIM card demand without including screenshots in the algorithm section.

2.2. Collection Data

Data analysis in this study uses the Linear Regression method to forecast the growth in demand for internet package cards at Sinar X-Sis Cell Pekanbaru for the next 2 years based on the data obtained, namely sales of the best-selling package cards at Sinar X-Sis Cell Pekanbaru. This forecast uses data on the sale of goods as much as 5 data on sales of package cards as material for calculating the method. Based on the data provided, namely from January 2020 to December 2021, the intended forecast is the forecast for January 2022 to December 2023. The sales data for the 5 types of package cards in units are as follows:

Table 1. Simcard Sales Data

2020						2021					
Month	AXIS	TRI	XL	TEL	SMARTF	Month	AXIS	TRI	XL	TEL	SMARF
JAN	277	374	244	533	82	JANI	204	310	247	455	70
FEB	248	245	213	376	76	FEB	225	234	221	230	64
MAR	279	223	190	297	83	MAR	197	227	193	321	49
APR	263	298	234	234	79	APR	189	301	233	105	58
MEY	239	276	186	278	69	MEI	167	244	212	234	67
JUN	254	302	105	390	101	JUN	210	267	143	203	87
JUL	283	332	207	452	87	JUL	199	190	186	190	82
AUG	217	294	178	249	60	AGU	238	189	229	239	59
SEP	213	278	119	478	59	SEP	221	208	301	268	70
OCT	175	310	254	369	87	OKT	234	288	200	341	72
NOV	202	286	260	475	83	NOV	185	167	180	302	80
DEC	103	321	276	412	72	DES	188	154	292	404	89
TOTAL	2753	3539	2466	4543	938	TOTAL	2457	2779	2637	3292	847

This data is used as a manual calculation dataset with Ms. Excel as well as testing using the Google Colaboratory Tools with the Python programming language. However, this data needs to go through the data pre-processing stage first.

2.3. Data processing

In pre-processing this data is done by taking raw data from Table 1 and then recapitulating it in the form of data stored in Excel format (.xls), which will later be used as new data consisting of variable x, namely the nth Month Data (x), then each item code KP01 to KP05 is used as the nth (y) y variable. The data that has been transformed can be seen in the following table:

Table 2. Data Sets

Monnth/year	Month to-n	AXIS	TRI	XL	TELKOMSEL	SMARTFREN
	x	y1	y2	y3	y4	y5

Jan-20	1	277	374	244	533	82
Feb-20	2	248	245	213	376	76
Mar-20	3	279	223	190	297	83
Apr-20	4	263	298	234	234	79
May-20	5	239	276	186	278	69
Jun-20	6	254	302	105	390	101
Jul-20	7	283	332	207	452	87
Aug-20	8	217	294	178	249	60
Sep-20	9	213	278	119	478	59
Oct-20	10	175	310	254	369	87
Nov-20	11	202	286	260	475	83
Dec-20	12	103	321	276	412	72
Jan-21	13	204	310	247	455	70
Feb-21	14	225	234	221	230	64
Mar-21	15	197	227	193	321	49
Apr-21	16	189	301	233	105	58
May-21	17	167	244	212	234	67
Jun-21	18	210	267	143	203	87
Jul-21	19	199	190	186	190	82
Aug-21	20	238	189	229	239	59
Sep-21	21	221	208	301	268	70
Oct-21	22	234	288	200	341	72
Nov-21	23	185	167	180	302	80
Dec-21	24	188	154	292	404	89

In the table above, x is the actual period of the monthly transaction order, while y1 to y5 are Sim card codes that contain transaction data for 24 months (2 Years 2020-2021).

2.4. Determining the Independent Variable (x)

The causative variable used is the year period (x) which will be predicted and the consequential variable is the demand for card packages (y).

2.5. Determine Value Constant a and b

After getting the values of X, Y, XY, and XX, calculations are carried out to get the values of a and b. The value of adan b is the coefficient used in forming the regression equation model which will then be used to carry out the prediction stages. The values of a and b can be calculated using the formula below:

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2} \qquad (1)$$

$$b = \frac{n(\sum xy)(\sum x^2) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} \qquad (2)$$

$$y = a + bx \qquad (3)$$

2.5. Data Analysis Using Linear Regression Method

At this stage, the coefficients a and b that have been obtained in the previous stage are then used to obtain a linear regression equation.

2.6. Prediction Results

The linear regression equation model that was obtained in the previous stage is then used to make predictions by applying the predicted time series to it.

The efficiency of the system is evaluated through a set of performance metrics, including Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), R-squared (R^2), Mean Absolute Percentage Error (MAPE), forecast bias, and residual analysis. RMSE and MAE measure the accuracy of predictions, with lower values indicating better accuracy, while R^2 assesses how well the model explains variance in SIM card demand. MAPE quantifies errors as a percentage of actual values, and forecast bias checks for systematic overestimation or underestimation. Residual analysis identifies patterns or outliers in prediction errors. Additionally, computational efficiency and adherence to specific business-defined accuracy and precision metrics are considered to assess the overall system's effectiveness in forecasting SIM card demand[14].

3. RESULTS AND DISCUSSION

3.1. Analysis Calculation Regression Linear

The stages of forecasting analysis using the Linear Regression method are as follows:

1. **Dataset Creation** : In this case the Linear Regression method requires 2 variables, namely, the sales period (month) as the x variable and monthly product sales data as the y variable.
2. **Formation of a Linear Regression Model :** The dataset that has been created is then used to determine the total value of x, y, xy, and x2. The sales period is formed in numerical form in column x. Column y is the quantity of product sales. xy is the product of x and y and x2 is the squared result of the contents of column x. The following are the values from the formation of the Linear Regression Model as an example for KP01 sales data:

Table 3. Formation of the KP01 Linear Regression Model

Month-Year	Month to- (x)	Stock Sold (y)	x^2	x.y
Jan-20	**1**	277	1	277
Feb-20	**2**	248	4	496
Mar-20	**3**	279	9	837
Apr-20	**4**	263	16	1052
May-20	**5**	239	25	1195
Jun-20	**6**	254	36	1524
Jul-20	**7**	283	49	1981
Aug-20	**8**	217	64	1736
Sep-20	**9**	213	81	1917
Oct-20	**10**	175	100	1750
Nov-20	**11**	202	121	2222
Dec-20	**12**	103	144	1236
Jan-21	**13**	204	169	2652
Feb-21	**14**	225	196	3150
Mar-21	**15**	197	225	2955

Apr-21	**16**	189	256	3024
May-21	**17**	167	289	2839
Jun-21	**18**	210	324	3780
Jul-21	**19**	199	361	3781
Aug-21	**20**	238	400	4760
Sep-21	**21**	221	441	4641
Oct-21	**22**	234	484	5148
Nov-21	**23**	185	529	4255
Dec-21	**24**	188	576	4512
Total :	**300**	5210	4900	61720

In table 3 above there are the total values of x, y, xy, and x2 that have been obtained which are then used to calculate the values of a and b as in the following equation:

$$a = \frac{(\sum y)(\sum x^2)-(\sum x)(\sum xy)}{n(\sum x^2)-(\sum x)^2}$$

$$b = \frac{n(\sum xy)(\sum x^2)-(\sum x)(\sum y)}{n(\sum x^2)-(\sum x)^2}$$

$$y = \mathrm{a} + \mathrm{b}x \qquad (4)$$

Information:

a : constant value a
b : constant value b
y : linear regression

The values of a and b are the coefficients used in forming the regression equation model which will then be used to carry out the prediction stages. From the results of calculations using the formula above, the constant value b and the constant value a are obtained from the calculation results using. After obtaining the constant values b and a, calculations are performed to obtain the value of the regression equation or y.

For example, Forecasting on KP01 in the 25 month:

n = 24
a = ((sum y)(sum x^2))-((sum x)(sum x.y))/sum n(sum x^2)-(sum x)^2
= 254,0942

b = sum n*(sum x.y)-((sum x)(sum y))/sum n*(sum x^2)-(sum x)^2
= -2,9609
after obtaining the values of a and b, then enter the values of a and b into the following equation (forcesting monht to-25) :

Y = a+b.x
= 254,094202898551+-2,96086956521739.25
= 180,0724638

Based on the above calculations, look for the forecasting value of each sales data from January 2020 to December 2021 (1st to 24th months) and also sales data that will be forecast from January 2022 to December 2023 (25th to months 48).

3.2 Implementation And Analysis Results

After preparing the data, then testing with tools *Google Colaboratory* . The stages are as follows:

1. Preparing Data : The data prepared is the Dataset contained in Table 2.2 which is stored in xlsx format. This data will later be used as material for analysis calculations which are uploaded to the *Google Colaboratory* page.
2. Implementation with tools *Google Colaboratory* :Forecasting package card sales by implementing Linear Regression on *Google Colaboratory*. Here is the *Google Colaboratory* main page.

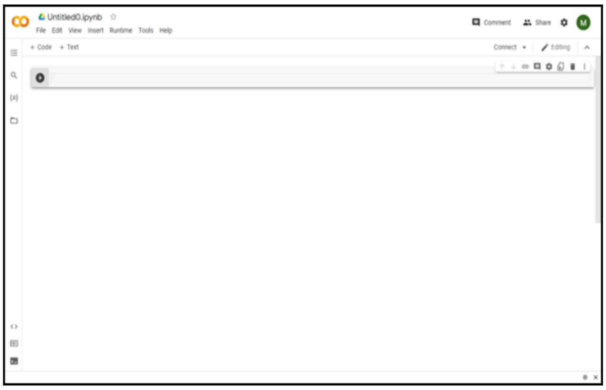

Figure 1. Tools *Google Colaboratory*

In this section, we will discuss the stages of testing the dataset. The stages of forecasting with Linear Regression using *Google Colaboratory* are as follows:

a. *Upload* Data On *Google Drive* : *Upload* data is done to enter data to be tested in *Excel* format. Next, enter data into *Google Drive* so that data can be retrieved by coding on *Google Colaboratory*.
b. Mengimport Library : In the *Google Colaboratory* implementation, it has a class, namely the process class. In addition, this software uses the python library which is shown in Figure 2

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
```

Figure 2. Import Library

Based on the image above, the following is an explanation of each *Python library* that is entered in the *Google Colaboratory* cell code:

Tabel 4. Import Library Description

No	Code	Description
1	Import numpy as np	Provide object array, analyze, numeric ,do matrices and, math operations
2	Import pandas as pd	To analyze, clean, explore
3	Import matplotlip.pyplot as plt	To make plot

4	From skleam.linear_model import Linear Regression	To count LinearRegression

c. Calling Data From *Google Drive* : At this stage, retrieve data that has been uploaded to Google Drive before, with the code in the cell code as shown in the following image:

```
from google.colab import drive
drive.mount ('/content/drive/')
# Mengimpor dataset
dataset = pd.read_excel ('/content/drive/MyDrive/kartupaket.xlsx')
x = dataset.iloc[:24, [0]].values
y1 = dataset.iloc[:24, [1]].values
y2 = dataset.iloc[:24, [2]].values
y3 = dataset.iloc[:24, [3]].values
y4 = dataset.iloc[:24, [4]].values
y5 = dataset.iloc[:24, [5]].values

dataset.head()
```

Figure 3. Data Calling

Then run the cell code in Figure 3 above and generate the same data call as the data we uploaded to *Google Drive* before.

d. Define model Regression Linear : Based on the previously called dataset, then a Linear Regression model is defined for each item whose sales will be forecasted. The following is the display of the cell code and the output of the code being executed:

```
modelKP01 = LinearRegression(fit_intercept=True)
modelKP02 = LinearRegression(fit_intercept=True)
modelKP03 = LinearRegression(fit_intercept=True)
modelKP04 = LinearRegression(fit_intercept=True)
modelKP05 = LinearRegression(fit_intercept=True)

modelKP01.fit(x,y1)
modelKP02.fit(x,y2)
modelKP03.fit(x,y3)
modelKP04.fit(x,y4)
modelKP05.fit(x,y5)
```

Figure 4. Models Regression Linear (1)

```
arraya = np.array([
            modelKP01.intercept_,
            modelKP02.intercept_,
            modelKP03.intercept_,
            modelKP04.intercept_,
            modelKP05.intercept_,
            ]);
arraycoef = np.array([
            *modelKP01.coef_,
            *modelKP02.coef_,
            *modelKP03.coef_,
            *modelKP04.coef_,
            *modelKP05.coef_,
            ]);
```

Figure 5. Regression Linear (2)

Figures 4 and 5 are *cell codes* that contain coding to create *a Linear Regression* model where the output will produce *intercep*t (a) and *coef* (b) values which can be seen in the following figure:

```
Model Regresi KP01
Nilai a :  254.09420289855072
Nilai b :  -2.960869565217390 4

Model Regresi KP02
Nilai a :  324.8804347826087
Nilai b :  -4.930434782608695

Model Regresi KP03
Nilai a :  194.21739130434784
Nilai b :  1.4726086956521736

Model Regresi KP04
Nilai a :  395.52898550724 64
Nilai b :  -5.525652173913044

Model Regresi KP05
Nilai a :  78.21739130434783
Nilai b :  -0.3073913043478261 3
```

Figure 6. value *Intercept (a)* and *Coef (b)*

e. Doing Forecasting : Forecasting is done by using the values a and b in each item code used as material for calculations following the Linear Regression formula as follows:

$$y = a + b.x \qquad (5)$$

a : value konstanta a
b : value konstanta b
x : variabel aktual x
y : regesi linear

Ting at moon January 2021 until January 2022(Monthto-13) for KP01 based on the formula above, then apply the *Linear Regression* formula to the cell code in *Google Colaboratory*, as an example of sales forecasting.

```
# Peramalan bulan ke - n
# X adalah bulan
X = range(1,49)
a = arraya
b = arraycoef

Y1 = [*a[0]+b[0]*X[0],*a[1]+b[1]*X[0],*a[2]+b[2]*X[0],*a[3]+b[3]*X[0],*a[4
]+b[4]*X[0]]
Y2 = [*a[0]+b[0]*X[1],*a[1]+b[1]*X[1],*a[2]+b[2]*X[1],*a[3]+b[3]*X[1],*a[4
]+b[4]*X[1]]
Y3 = [*a[0]+b[0]*X[2],*a[1]+b[1]*X[2],*a[2]+b[2]*X[2],*a[3]+b[3]*X[2],*a[4
]+b[4]*X[2]]
Y4 = [*a[0]+b[0]*X[3],*a[1]+b[1]*X[3],*a[2]+b[2]*X[3],*a[3]+b[3]*X[3],*a[4
]+b[4]*X[3]]
Y5 = [*a[0]+b[0]*X[4],*a[1]+b[1]*X[4],*a[2]+b[2]*X[4],*a[3]+b[3]*X[4],*a[4
]+b[4]*X[4]]
.......
Y46 = [*a[0]+b[0]*X[45],*a[1]+b[1]*X[45],*a[2]+b[2]*X[45],*a[3]+b[3]*X[45]
,*a[4]+b[4]*X[45]]
Y47 = [*a[0]+b[0]*X[46],*a[1]+b[1]*X[46],*a[2]+b[2]*X[46],*a[3]+b[3]*X[46]
,*a[4]+b[4]*X[46]]
Y48 = [*a[0]+b[0]*X[47],*a[1]+b[1]*X[47],*a[2]+b[2]*X[47],*a[3]+b[3]*X[47]
,*a[4]+b[4]*X[47]]

Y = np.array([Y1,Y2,Y3,Y4,Y5,Y6,Y7,Y8,Y9,Y10,
              Y11,Y12,Y13,Y14,Y15,Y16,Y17,Y18,Y19,Y20,
              Y21,Y22,Y23,Y24,Y25,Y26,Y27,Y28,Y29,Y30,
              Y31,Y32,Y33,Y34,Y35,Y36,Y37,Y38,Y39,Y40,
              Y41,Y42,Y43,Y44,Y45,Y46,Y47,Y48])
```

Figure 7. *Cell Code* formula *Regression Linear*

Then run the *cell code* in Figure 7 above and produce *output* as shown in Figure 8 The following is the result of the *call code* in the *Google colaboratory* applying the *linear regression* formula. The following figure explains that in 2 years sales results can be predicted the average number of sales for the next 2 years as shown in the image below:

	x	y1	y2	y3	y4	y5
0	1	251.133333	319.950000	195.690000	390.003333	77.910000
1	2	248.172464	315.019565	197.162609	384.477681	77.602609
2	3	245.211594	310.089130	198.635217	378.952029	77.295217
3	4	242.250725	305.158696	200.107826	373.426377	76.987826
4	5	239.289855	300.228261	201.580435	367.900725	76.680435
5	6	236.328986	295.297826	203.053043	362.375072	76.373043
6	7	233.368116	290.367391	204.525652	356.849420	76.065652
7	8	230.407246	285.436957	205.998261	351.323768	75.758261
8	9	227.446377	280.506522	207.470870	345.798116	75.450870
9	10	224.485507	275.576087	208.943478	340.272464	75.143478
10	11	221.524638	270.645652	210.416087	334.746812	74.836087
11	12	218.563768	265.715217	211.888696	329.221159	74.528696
12	13	215.602899	260.784783	213.361304	323.695507	74.221304
13	14	212.642029	255.854348	214.833913	318.169855	73.913913
14	15	209.681159	250.923913	216.306522	312.644203	73.606522
15	16	206.720290	245.993478	217.779130	307.118551	73.299130
16	17	203.759420	241.063043	219.251739	301.592899	72.991739
17	18	200.798551	236.132609	220.724348	296.067246	72.684348
18	19	197.837681	231.202174	222.196957	290.541594	72.376957
19	20	194.876812	226.271739	223.669565	285.015942	72.069565
20	21	191.915942	221.341304	225.142174	279.490290	71.762174
21	22	188.955072	216.410870	226.614783	273.964638	71.454783
22	23	185.994203	211.480435	228.087391	268.438986	71.147391
23	24	183.033333	206.550000	229.560000	262.913333	70.840000
24	25	180.072464	201.619565	231.032609	257.387681	70.532609
25	26	177.111594	196.689130	232.505217	251.862029	70.225217

Figure 8. Result Forecasting (1)

26	27	174.150725	191.758696	233.977826	246.336377	69.917826
27	28	171.189855	186.828261	235.450435	240.810725	69.610435
28	29	168.228986	181.897826	236.923043	235.285072	69.303043
29	30	165.268116	176.967391	238.395652	229.759420	68.995652
30	31	162.307246	172.036957	239.868261	224.233768	68.688261
31	32	159.346377	167.106522	241.340870	218.708116	68.380870
32	33	156.385507	162.176087	242.813478	213.182464	68.073478
33	34	153.424638	157.245652	244.286087	207.656812	67.766087
34	35	150.463768	152.315217	245.758696	202.131159	67.458696
35	36	147.502899	147.384783	247.231304	196.605507	67.151304
36	37	144.542029	142.454348	248.703913	191.079855	66.843913
37	38	141.581159	137.523913	250.176522	185.554203	66.536522
38	39	138.620290	132.593478	251.649130	180.028551	66.229130
39	40	135.659420	127.663043	253.121739	174.502899	65.921739
40	41	132.698551	122.732609	254.594348	168.977246	65.614348
41	42	129.737681	117.802174	256.066957	163.451594	65.306957
42	43	126.776812	112.871739	257.539565	157.925942	64.999565
43	44	123.815942	107.941304	259.012174	152.400290	64.692174
44	45	120.855072	103.010870	260.484783	146.874638	64.384783
45	46	117.894203	98.080435	261.957391	141.348986	64.077391
46	47	114.933333	93.150000	263.430000	135.823333	63.770000
47	48	111.972464	88.219565	264.902609	130.297681	63.462609

Figure 9. Result Forecasting (2)

The figure above shows the forecasting output for each month (48 months) from January 2020 to December 2023 for each card package KP01 to KP05.

f. Forecasting Visualization

Visualization Forecasting is made in graphical form which is displayed in the form of actual data and forecasting data for each item, while the graph can be seen in the following figures:

Figure 10. Grafik Forecasting KP01

The figure above shows the results of the KP01 forecasting chart, namely the Axis card, which shows a decrease in demand in the next 2 (two) years.

Figure 11. Grafik Forecasting KP02

The figure above shows the results of the KP01 forecasting chart, namely the TRI card, which shows a decrease in demand in the next 2 (two) years.

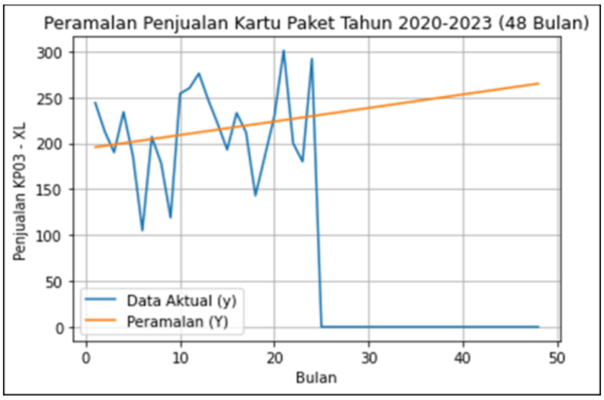

Figure 12. Grafik Forecasting KP03

In the picture above is the result of the KP03 forecasting chart, namely the XL package card which shows an increase in demand in the next 2 (two) years.

(d)

Figure 13. Grafik Forecasting KP04

The figure above is the result of the KP04 forecasting chart, namely the TELKOMSEL package card which shows a decrease in demand in the next two years.

(e)

Figure 14. Forecasting KP05 Graph

The figure above shows the results of the KP05 forecasting chart, namely the SMARTFREN card package, which shows a decrease in demand in the next 2 (two) years.

Based on the graphs in Figure 10 to Figure 14 it can be seen that the graph shows sales data (Actual Data) with a blue line for 48 months from January 2020 to December 2023, there is training data from month 1 to month 24 and month to -25 to 48th is testing data for which there is no sales data so forecasting is necessary. Furthermore, each graph also displays a yellow straight line on each data card sales forecasting package KP01 to KP05. From the five graphs above, it can be seen that there is one forecasting graph showing a line going up which states that the demand for simcards until December 2023 will increase, namely KP03, while KP01, KP02, KP04, and KP05 show a line

4. CONCLUSION

Based on the research that has been done by the author, it can be concluded from the research that has been made as follows:

1. The analysis was carried out by forecasting data on sales of Sinar X-Sis Cell Pekanbaru simcards on demand for monthly simcard in the 25th to 48th months to come or more precisely from January 2022 to December 2023 based on sales data 24 months from from January 2020 to December 2021 using the Linear Regression method which was processed using manual calculations and testing Google Colaboratory too\ls to produce a regression line that shows an increase in demand for the KP03 package card, while the KP01, KP02, KP04 and KP05 simcards show a decrease in demand in the future come.

2. The implementation of the Linear Regression method is carried out by testing forecasting calculations using Google Colaboratory tools so as to produce graphs showing increases and decreases in each simcard sales in the future from January 2022 to December 2023. future packages in January 2022 to December 2023.

REFERENCES

[1] J. Wang and S. Duan, "An online anomaly learning and forecasting model for large-scale service of internet of things," *Proceedings - 2014 International Conference on Identification, Information and Knowledge in the Internet of Things, IIKI 2014*, pp. 152–157, 2014, doi: 10.1109/IIKI.2014.38.

[2] F. Isyarah, A. Hasan, and F. Wiza, "CLUSTERING DAERAH MISKIN DI PROVINSI RIAU MENGGUNAKAN METODE K-MEANS," 2020.

[3] N. Nasution, F. Feldiansyah, A. Zamsuri, and M. A. Hasan, "Synthetic Minority Oversampling Technique for Efforts to Improve Imbalanced Data in Classification of Lettuce Plant Diseases," *JURNAL TEKNOLOGI DAN OPEN SOURCE*, pp. 31–40, Feb. 2023, doi: 10.36378/jtos.v6i1.2883.

[4] E. M. S. Rochman and A. Rachmad, "Forecasting Application for Simpati Telkomsel Card Using Backpropagation (Case Study in Bangkalan Madura-Indonesia)," *Adv Sci Lett*, vol. 23, no. 12, pp. 12340–12343, Mar. 2018, doi: 10.1166/asl.2017.10635.

[5] H. Djanggih, *Urgency Legal Aspects of Growth Information Technology In Indonesia*. 2017. [Online]. Available: https://www.researchgate.net/publication/317267323

[6] W. Segoro and W. Widiawati, "The Influence Factors on the Satisfaction of Telkomsel SIM Card Users in Indonesia," 2019.

[7] W. Segoro and W. Widiawati, "The Influence Factors on the Satisfaction of Telkomsel SIM Card Users in Indonesia," 2019.

[8] A. Brzeziński, T. Dybicz, and Ł. Szymański, "DEMAND MODEL IN THE AGGLOMERATION USING SIM CARDS."

[9] M. Pradana, H. Hafid, M. Diandri, and N. Bunga, "The Influence of Brand Equity toward Indonesian Prepaid SIM Card Customers' Purchase Decision," 2018. [Online]. Available: https://www.researchgate.net/publication/328698576

[10] H. Sunaryanto, M. A. Hasan, and G. Guntoro, "Classification Analysis of Unilak Informatics Engineering Students Using Support Vector Machine (SVM), Iterative Dichotomiser 3 (ID3), Random Forest and K-Nearest Neighbors (KNN)," *IT Journal Research and Development*, vol. 7, no. 1, pp. 36–42, Aug. 2022, doi: 10.25299/itjrd.2022.8912.

[11] J. Ye, R. Janardan, and Q. Li, "Two-Dimensional Linear Discriminant Analysis," in *Advances in Neural Information Processing Systems*, L. Saul, Y. Weiss, and L. Bottou, Eds., MIT Press, 2004. [Online]. Available: https://proceedings.neurips.cc/paper/2004/file/86ecfcbc1e9f1ae5ee2d71910877da36-Paper.pdf

[12] C. H. Park and H. Park, "A comparison of generalized linear discriminant analysis algorithms," *Pattern Recognit*, vol. 41, no. 3, pp. 1083–1097, 2008, doi: https://doi.org/10.1016/j.patcog.2007.07.022.

[13] A. Sharma and K. K. Paliwal, "Linear discriminant analysis for the small sample size problem: an overview," *International Journal of Machine Learning and Cybernetics*, vol. 6, no. 3, pp. 443–454, 2015, doi: 10.1007/s13042-013-0226-9.

[14] J. Wang, M. Xu, H. Wang, and J. Zhang, "Classification of Imbalanced Data by Using the SMOTE Algorithm and Locally Linear Embedding."