Zero : Jurnal Sains, Matematika, dan Terapan
E-ISSN : 2580-5754
P-ISSN : 2580-569X
Vol. 7, No. 1, June 2023
pp. 14-22

Prediction of Tobacco Production With of Tobacco Production With *Singular Spectrum Analysis* (SSA)

Ardaniah Hazrah[1], **Rina Widyasari**[2]
[1,2]Department of Mathematics, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

Article Info
Article history:
Received mm dd, yyyy
Revised mm dd, yyyy
Accepted: July, 23, 2023

Keywords:
Rediction tobacco
Singular Spectrum Analisis

ABSTRACT

The increasing population growth in North Sumatra Province every year has an impact on the increasing need for basic resources for daily life in North Sumatra, so it is necessary to do a forecast to anticipate future food shortages. This study aims to determine the results of tobacco production in North Sumatra Province in 2024 and the accuracy of the methods used. The results of the study obtained the forecast results of rice production in North Sumatra Province in 2024 with the Singular Spectrum Analysis (SSA) method from Quarter I to Quarter III of 665 tons, 725 tons, and 740 tons respectively with the level of forecasting accuracy based on the standard MAPE value obtained of 15.78%. The MAPE value obtained is less than 10% and close to 0%, meaning that the SSA method with windows length 9 and 7 groups is very accurately used to forecast tobacco production in North Sumatra Province.

Corresponding Author:

Ardaniah Hazrah,
Department of Mathematics, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia
Email: ardaniah0703193097@uinsu.ac.id

1. INTRODUCTION

Economic growth and development in Indonesia is driven by contributions to agriculture because in developing countries, agriculture is a potential economic sector and is the backbone in improving the national economy which is able to provide employment opportunities and contribute to the formation of GDP (Setyawan, Subantoro, & Prabowo, 2016).In 2017, agriculture contributed considerable national income. This can be seen in the contribution of the agricultural sector to the Gross Domestic Product which is high at 1,785,880.7 trillion rupiah. While the agricultural sector itself consists of 5 subsectors, namely, food crops, plantations, livestock, forestry, and fisheries, which make a major contribution to regional income (BPS, 2018). Tobacco is a leading commodity in the plantation sector which is very important in the Indonesian economy. Products that can be produced or traded are tobacco leaves and cigarettes. Each region has tobacco products with their own distinctive flavors. Therefore, tobacco in Indonesia is in great demand by local and international communities (Rofiuddin & Widayati, 2018).

The tobacco industry in Indonesia is growing rapidly in line with the increasing number of smokers. This is due to the smoking habit of the Indonesian people. In line with that, the tobacco industry plays a role in the national economy as a contributor to state revenue through excise taxes. The growth of the cigarette industry is certainly followed by the development of tobacco cultivation by local farmers, and has played a role as a provision of employment, as well as a source of income for the community and the regional economy (Rachmat, 2010). SSA is a flexible forecasting method. SSA was introduced by Broomhead and King in 1986 and is still being developed (Golyandina, Nekrutkin, & Zhigljavsky, 2001). SSA is a nonparametric analysis technique that does not require statistical assumptions such as stationarity or normality of residuals (Zhang, Wang, He, Peng, & Ren, 2011). The basic SSA algorithm is divided into two stages, namely decomposition and reconstruction. The forecasting result of the SSA method is a combination of seasonal and trend components from the reconstruction stage (Zhigljavsky, 2010). Research related to the above includes Asnafiyah Asrof et al (2019) which examines the use of the SSA method to forecast red chili production which contains seasonal elements showing a good level of accuracy with a Mean Absolute Percentage Error (MAPE) value of 18.23%.Yogo Aryo Jatmiko et al (2020) examined the SSA method in forecasting shallot production and obtained accurate forecasting results with a MAPE value of 15.63%.Desy Tresnowati Hardi (2019) conducted research and obtained results that the SSA method was accurately used in forecasting East Java GRDP with a MAPE value of 1.59%. There are many forecasting methods that are currently developing. However, forecasting results will be useful if using the right forecasting method.

The seasonal element is influenced by the harvest period which will be high in certain months. Based on this, forecasting can be done using the Singular Spectrum Analysis (SSA) method because the SSA method decomposes time series data into components in the form of trends, seasonality, cyclicality and *noise*. SSA is also a forecasting method that is more flexible than other forecasting methods because it uses a nonparametric approach, namely there is no need for assumption tests such as independence and normality of residuals. and because the data on tobacco production is not stationary so the right method to use is SSA. Based on this, predictions can be made using the *Singular Spectrum Analysis* method.shows that the SSA method is the most accurate to use. Based on the explanation above, the researcher chose the title "Prediction of Tobacco Production with the *Singular Spectrum Analysis* Method" With this research can help identify patterns and trends from historical data on tobacco production, so whether tobacco production can keep up with consumer demand for the needs of the tobacco industry. And allows researchers and decision makers to make more accurate estimates of Tobacco Production in the coming year. Thus this can support better planning and decision making in the Tobacco Industry.

2. RESEARCH METHOD

The study employed a kind of quantitative research. The data collection initiative in the study is to collect a small percentage of the cost of tobacco production in north Sumatra next year from 2020-2022 acquired from the north Sumatra plantation and farm service website. The stages in data analysis in this study include the following:
1. Data description
A. research data
B. plot data
C. data structure
2. An oral analysis with the following two approaches:
A. decomposition, this stage is divided into two steps which is:
Embedding, at this step the embedding matrix Dimensions. Additional autism From the time sequencing data with theselection of Windows surveys
2) lower value (SVD). At this stage, the value of eigentriple of the symmetrical matrix = steaming vapor
B. reconstruction, this stage is divided into two steps which is:
1) grouping, at this stage begins a grouping of eigentriple based on the characteristics of each component
2) averaging diagonal is done with a reconstruction of each matrix found in the inequality matrix Into a new sequence of time data with length of matter.
3. Calculating the forecasting results of the time sequence, averaging diagonally With an r-forecasting method
Calculating accuracy of fortunation using mean absolem perecentage error (mape)

3. RESULT AND ANALYSIS

E-ISSN: 2580-5754P-ISSN : 2580-569X

The data analyzed in this study are North Sumatra tobacco production data compiled quarterly obtained from the North Sumatra Provincial Plantation and Livestock Service from 2020 to 2023 which are presented in Table 4.1 below:

4.1 Quartile Data Table

Year	Quarter 1	Quarter 2	Quarter 3
2020	**450**	**550**	**650**
2021	**505**	**650**	**480**
2022	**650**	**450**	**550**
2023	**600**	**680**	**720**

From the data above, descriptive analysis is carried out which is shown in Table 4.2 below

4.2 Perquartile Descriptive Data Table

Description	Quartile 1	Quartile 2	Quartile 3
Minimum	450	450	480
Maximum	650	680	720
Mean	551.25	582.5	600
Std.Deviation	90.40418501	104.363148	106.1445555

Data pattern

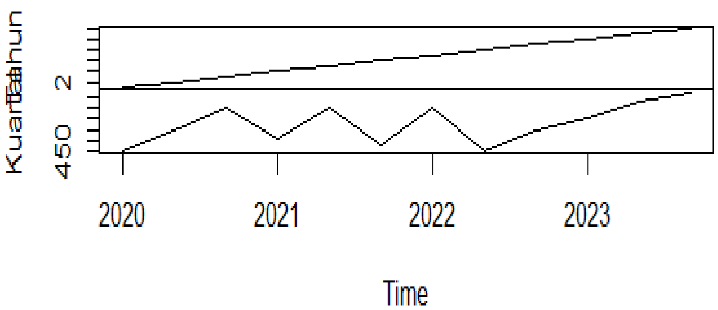

Figure 4.1 Time Series Plot of Tobacco Production data of North Sumatra Province

Figure 4.1 shows that the pattern of tobacco production data has decreased and increased repeatedly so that data fluctuations are identified as containing seasonal elements. Furthermore, to determine whether the data is stationary can be identified by using the ADF (Augmented Dickey-Fuller) test.

4.3 ADF Test

Test Statistics	Value
Dickey-Fuller	-5.4493
p-value	0.01

From Table 4.3, the ADF test statistic value is obtained with the hypothesis:

H0 : Data is not Stationary

H1 : Data Stationary

Decision: Based on the ADF test obtained, the p-value of 0.01 <0.05 so that H0 is rejected, meaning that the data is stationary.

Data Structure

To conduct the analysis, the data is first divided into two, namely in sample data and out sample data which are presented in Table 4.4 and Table 4.5 below, respectively:

Table 4.4 In-sample data of Tobacco Production of North Sumatra Province

Year	Quarter 1	Quarter 2	Quarter 3
2020	450	550	650
2021	505	650	480
2022	650	450	550

Table 4.5 Out sample data of Tobacco Production of North Sumatra Province

Year	Quarter 1	Quarter 2	Quarter 3
2023	600	680	720

The in sample data above will be used to create the trajectory matrix ***X***. While the out sample data is used to validate the accuracy of the forecasting.

Stages of Tobacco Production Forecasting using the Singular Spectrum Analysis (SSA) method

Decomposition

Embedding

At this stage, the data in Table 4.6 is converted into one-dimensional data with a total of 12 data from Quarter I of 2020 to Quarter III of 2023 with the following arrangement:

[Quartile 1 2020]	$[x]_1$	450
[Quartile 2 2020]	$[x]_2$	550
[Quartile 3 2020]	$[x]_3$	650
[Quartile 1 2021]	$[x]_4$	505
[Quartile 2 2021]	$[x]_5$	650
[Quartile 3 2021]	$[x]_6$	480
[Quartile 1 2022]	$[x]_7$	650
:		
[Quartile 3 2023]	$[x]_{12}$	720

Furthermore, the data is converted into multidimensional data referred to as a trajectory matrix ***X*** with dimensions $L \times K$.

Table 4.6 MAPE Results

1	0.1000
2	0.1000
3	0.0846
4	0.1881
5	0.1077
6	0.1667
7	0.1077
8	0.1000
9	0.1000

10	0.2000
11	0.1029
12	0.1389

Singular Value Decomposition

Furthermore, at this stage the calculation is carried out to find the eigentriple value based on the trajectory matrix ***X*** The initial step taken is to form a symmetrical matrix ***S*** = ***XXT*** as follows:

```
      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
[1,] 1182525 1261500 1235150 1225000 1206250 1143750 1200500 1238400 1383100
[2,] 1261500 1402525 1326000 1365150 1290250 1271250 1317750 1350000 1503900
[3,] 1235150 1326000 1330425 1280500 1303400 1196750 1295250 1286650 1448100
[4,] 1225000 1365150 1280500 1330425 1244750 1238400 1274750 1314750 1462150
[5,] 1206250 1290250 1303400 1244750 1277900 1164000 1266000 1252500 1411500
[6,] 1143750 1271250 1196750 1238400 1164000 1157900 1182000 1217500 1356000
[9,] 1383100 1503900 1448100 1462150 1411500 1356000 1433500 1475100 1643300
```

Singular Value

Table 4.7 Eigenvalues and Singular Values

No.	Eigenvalues	Singular values
1	1.173525	3.425675
2	1.319418	3.632380
3	4.281461	2.069169
4	2.989676	1.729068
5	2.698877	1.642826
:	:	:

Eigenvector

Table 4.8 Eigenvectors

No.	U_1	U_2	...	U_9
1	-0.3150154	0.08280501	...	0.371083927
2	-0.3438252	-0.29361990	...	-0.3001997
3	-0.3327782	0.47190051	...	0.702704571
:	:	:	:	:
9	-0.3731083	-0.09104359	...	-0.062305000

Principal Component

After the eigenvalue and eigenvector are obtained, then the principal component value is calculated using equation 4.9, the results of which are presented as follows:

Table 4.9 Principal component values

No.	V_1	V_2	V_3	V_4	$V5$	...
1	-0.4801092	0.5340461	-0.6691902	0.1909825	-8.131236	...
2	-0.4947376	-0.4927051	-0.2315751	-0.6773842	6.401185	...
3	-0.5080318	0.4773567	0.6853733	-0.2104701	8.996261	...
:	:	:	:	:	:	:

Reconstruction Stage

At this stage, the eigentriple grouping of SVD results is carried out based on the characteristics of each component. To determine the members of the group is done by looking at the plot of the eigenvector. The following shows the plot of the eigenvector with the value of i = 1,2, ... ,

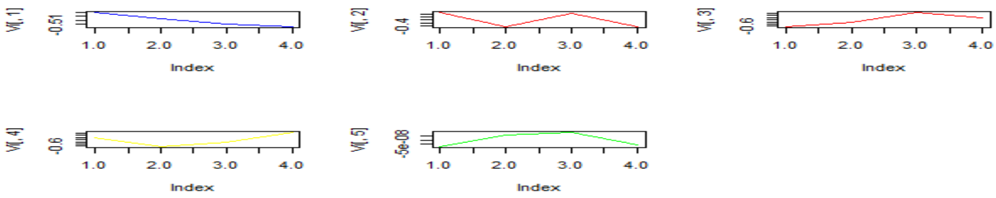

Figure 4.2 Data Eigenvector Plot

The W-correlation plot is shown as follows:

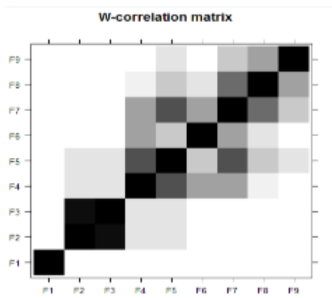

Figure 4.3 W-correlation plot

From figure 4.3 it is seen that out of the 9 components that have it can be grouped into 7 groups based on the low height of correlation of each group. The low height of correlation can be seen from the bright, dark slices between components of f1 to f9. It shows that the three component f2 f3 and the f5 and f7 have strong correlation according to the darkness the component component at f1, f4, f6, f8 and f9 has a weak correlation between the components.

$$X_n \quad \begin{bmatrix} 518.1051 & 533.8912 & 548.2375 & 557.2311 \\ 565.4886 & 582.7184 & 598.3767 & 608.1928 \\ 547.3195 & 563.9957 & 579.1510 & 588.6517 \\ 549.0181 & 565.7461 & 580.9483 & 590.4785 \\ 533.9439 & 550.2126 & 564.9975 & 574.2660 \\ 511.1469 & 526.7210 & 540.8747 & 549.7474 \\ 539.6138 & 556.0553 & 570.9971 & 580.3641 \end{bmatrix}$$

549.2942	566.0305	581.2405	590.7754
613.6504	632.3476	649.3396	659.9917

X_{12}

-52.67827	-38.60765	84.761555	2.576502
-15.39043	66.93354	-93.484877	42.155849
108.23047	-78.68059	64.732745	-88.934026
-43.28350	81.64850	-101.757872	62.131259
122.25934	-92.21466	78.166245	-102.227353
-17.74799	75.75505	-105.640816	47.855598
89.25119	-31.09290	2.294425	-55.451258
-108.55433	16.81379	28.964599	56.325537
-72.05723	-2.52990	39.925099	30.140946

X_{13}

-15.42682612	54.7164568	17.0009533	-54.8075644
-0.09813716	0.3480766	0.1081509	-0.3486562
-5.54997369	19.6848589	6.1162836	-19.7176359
-0.73457934	2.6054341	0.8095346	-2.6097724
-6.20327877	22.0020263	6.8362508	-22.0386616
-13.39895103	47.5239119	14.7661573	-47.6030432
21.13498422	-74.9623702	-23.2915622	75.0871889
9.26017374	-32.8443383	-10.2050662	32.8990269
8.40684888	-29.8177331	-9.2646695	29.8673821

Calculating Forecasting Results from Time Series Data of Diagonal Averaging Results with R-forecasting Meth od

The next step is to do forecasting on out sample data using R-forecasting method. The following shows the results of the out sample data forecast:

Table 4.10 Forecasting results on out sample data

Year	Quarter	Actual Data	Forecast Data	Diffee nce
	Quarter 1	600	665	65
2023	**Quarter 2**	680	725	45
	Quarter 3	720	740	20

Figure 4.4 Plot of Actual Data and Forecast Data

In Figure 4.4, it can be seen that the plot results of actual and forecast data have almost the same looping pattern, where the forecast data is close to the actual data.

Forecasting Accuracy

The forecasting accuracy of the SSA method in this study uses the Mean Absolute Percentage error (MAPE) method. Based on the equation, the MAPE value for the out sample data is presented in Table 4.11 below:

Table 4.11 MAPE Value on Out Sample Data Forecasting Results

Year	Quarter	Actual Data	Forecast Data	Differen ce	Error
	Quarter 1	600	665	65	0.000000
2023	**Quarter 2**	680	725	45	0.000000
	Quarter 3	720	740	20	4.736952

Total Error	4.736952
Mean Absolute Percentage Error (MAPE)	1.578984

From Table 4.11, it can be seen that the error value between actual data and forecast data is quite small. Judging from the MAPE value obtained of 1.578984 or 15.78%. With windows length, forecasting is carried out on tobacco production in 2024. The forecasting results are as follows:

Table 4.12 Forecast results of tobacco production in 2024

Quarter	Forecast Data
Quarter 1	665
Quarter 2	725
Quarter 3	740

From Table 4.12, it can be seen that the forecast of tobacco production in 2024 in the first quarter is 665 tons, the forecast in the second quarter is 725 tons, and the forecast in the third quarter is 740 tons.

Discussion

The data analyzed in this study is the tobacco production data of North Sumatra province from 2020 to 2023 using the Singular Spectrum Analysis (SSA) method. Based on the results of the analysis, the forecast results in the first quarter of 2023 tend to be 665 tons, a difference of 65 tons with actual data of 600 tons and has an error of 0.000. And in Quarter II - 2023, the actual data has an amount of 680 tons and has a forecast forecast that will increase by 45 tons to 725 tons with an error of 0.000. And in the third quarter of 2023, the forecast data is 740 tons and the actual data at the beginning of 720 has a difference of 20 tons and an error of 4.736952. There is a not too significant difference in the forecast results with the actual data in 2023 from Quarter I to Quarter III as seen from the standard Mean Absolute Percentage Error (MAPE) value. The MAPE value is 15.78%. This shows that the forecasting results are accurate enough to forecast rice production in 2023, where the forecast results obtained tend to be higher than in 2023. According to data from the kemenperin.go.id the purchase of tobacco in north Sumatra is now 1780 tons. This represents the approaching target of 2000 tons of tonase. That's why the tobacco production buyer for 2023 would have been sufficient.

4. CONCLUSIOON

The results of forecasting tobacco production in North Sumatra province in 2024 with the Singular Spectrum Analysis (SSA) method from Quarter I to Quarter III are 665 tons, 725 tons, and 740 tons, respectively. The level of accuracy of rice production forecasting with the SSA method based on the standard MAPE value is obtained at 15.78%. the purchase of tobacco in north Sumatra is now 1780 tons. the purchase of tobacco in north Sumatra is now 1780 tons

REFERENCES

[1] Ali Jamil. 2022. Statistik Unggulan Perkebunan Nasional, Jakarta : Sekretariat Derekorat Jendral Perkebunan.

[2] Andika, Gilang Bimasakti. (2020). "Peramalan Nilai Tukar Petani menggunakan Metode Singular Spectrum Analysis". E-Jurnal Matematika.Vol 9(3), pp. 171-176.

[3] BPS. (2022). Statistik Tembakau Indonesia. Jakarta: Badan Pusat Statistik.

[4] George E.P, Box (2008). *Time Series Analysis Forecasting and Control.* (Canada: John Wiley &Sons)

[5] Golyandina dan Zhigljavsky. (2013). Singular Spectrum Analysis for Time Series. STAT ME.

[6] Hardi, Desy Tresnowati dkk, (2019) "Peramalan Produk Domestik Bruto (PDB) Sektor Pertanian, Kehutanan, dan Perikanan Menggunakan Singular Spectrum Analysis (SSA)". Jurnal Gaussian, Volume 8, Nomor 1.

[7] Hidayat,Kukuh Wahyu.(2020)"*Pemodelan Jumlah Titik Panas di Provinsi Kalimantan Timur dengan Metode Singular Spectrum Analysis*". Jambura Jurnal of Probability and Statistics. Vol 1 no 2.

[8] Idrus, Rezky Amalia. (2022)."*Penerapan Metode Singular Spectrum dalam Peramalan Jumlah Produksi Beras di Kabupaten Gowa*". Journal of Statistic and its Aplication on Teaching and Research. Vol 4, No 2.

[9] Indi,Yohanes,Gema.(2022)."*Analisis Faktor-Faktor Yang Mempengaruhi Usaha Tani Tembakau di Desa Buduan Kecamatan Suboh Kabupaten Situbondo*" Prosiding Nasional,

[10] Luluk,M Burhanis,Imam (2021)"Optimasi K-Nearest Neighbor Dengan Partical Swarm Optimazation Untuk Memprediksi Harga Tembakau",Jurna Teknoologi Informasi Management.Vol 1 No 2.