

Predicting Scholarship Grants Using Data Mining Techniques

Allemar Jhone P. Delima

Abstract—Almost all major colleges and state universities assess students' comprehensive quality and set different rewards and regulations for the various level in order to stimulate students' interest to study and participate in extracurricular activities. The main reward system that is used is the providing of financial incentives such as scholarship grants. In this paper, several data mining techniques such as clustering and forecasting was integrated to discover and assess future outcomes and matters concerning The Student Financial Assistance Unit (SFAU) of Surigao State College of Technology (SSCT) holds all the records of scholarship grants and its grantees from June 2014. The study visualizes the increase of the grantees in every scholarship grants in the next five years to prepare the budget that needs to be allocated by the sponsoring agents. ARIMA(1,0,0) model for time series analysis was used in the study and found out to be very effective as it produced results as shown in Fig. 11-23.

Index Terms—Scholarship prediction, scholarship grants, kmeans, arima, trend analysis.

I. INTRODUCTION

Every university has its own methods to realize scholarship assessment. Scholarships are established by schools as a motivation for students that have outstanding and exceptional achievements. It is a form of incentive and benefit to inspire and embolden students and improve schools' learning principles [1]. From that point of view, an optimal way is considered to assess and distribute scholarship grants.

Data Mining is the process of extracting information from large data sets through the use of algorithms and methods drawn from the field of statistics and Database Management Systems [2]. Clustering, decision trees, genetic algorithms, Bayes classifiers, association rules, neural networks, and support vector machines to name some, are the algorithms and methods that can be used in data mining analysis that allow getting important information from the database [3].

One of the most popular data mining techniques is clustering. It represents an unsupervised learning method whose objective is to divide the data set so that the distance among the clusters should be minimal, whereas the inter-cluster distance should be maximal. One of the methods frequently used in data partition is the k-means data. The k-means algorithm is the easiest and the most common algorithm based on the squared error criterion [3].

The study clustered the indexed data of the different scholarship programs offered in Surigao State College of

Technology, the only State College in the province of Surigao del Norte, Philippines as to the number of scholarship grantees using K-Means. To forecast, this study used Autoregressive Integrated Moving Average (ARIMA). From the obtained results, this will also provide a predicted increase and decrease of the grantees in the next five years. This is to monitor and to identify which scholarship grant will exhibit an increase in number of grantees within targeted time as a preparation for the budget allocation by the sponsoring agents.

II. RELATED LITERATURE

Data mining also termed as Knowledge Discovery in Databases (KDD), is a medium of discovering novel and potentially valuable information from large amounts of data [4]. The field of Educational Data Mining is fresh, new, and developing in the field of education sector which can also be applied in areas such as government, accounts, sports, transportation and a lot more [5]. This new emerging field concerns with developing methods that discover knowledge from data originating from educational environments. Data mining techniques such as decision trees, Naïve Bayes, K-Nearest neighbor, neural networks, K-means clustering and many others are instrumental in extracting data from the datasets [6].

Cluster analysis is considered as one of the most important techniques in data mining. It has attracted more and more attention in this big data era [7]. Data clustering is used to recognize patterns or clusters from a set of objects. In general, data clustering divides set of objects into groups or clusters where the objects in the same cluster or group are identical to each other than to objects from the other clusters [8].

A study analyzed crimes such as theft, homicide, and various drug offenses along with suspicious activities, noise complaints, and burglar alarm by using qualitative and quantitative approach [9]. Using K-means clustering data mining approach on a crime dataset from the New South Wales region of Australia, crime rates of each type of crimes and cities with high crime rates have been found.

Furthermore, the data mining techniques were implemented to understand specific trends and pattern of terrorist attacks in India. K-means clustering was used to determine the year wherein the terrorist groups were most active and also which terrorist group has affected the most [10].

On the other note, the literature on representations and distance measures for time-series, clustering, and classification is extensive [11]-[13]. Time series analysis method is a kind of data mining method, which is a sequence

Manuscript received December 13, 2018; revised June 5, 2019.

A. J. Delima is with the College of Engineering and Information Technology, Surigao State College of Technology, Surigao City, Philippines (e-mail: allemarjpd@ssct.edu.ph).

of data points, typically consisting of successive measurements made over a time interval. It is a method for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values [14].

Prediction can be made utilizing autoregressive integrated moving average (ARIMA) algorithm that used historical data in predicting cases such as in education, society, climate, health, and others. In other countries, the use of ARIMA algorithm in forecasting incidence of hemorrhagic fever with renal syndrome in China was observed [15]. Meanwhile, the ARIMA model was also used in forecasting dengue hemorrhagic fever cases in Southern Thailand [16]. Further, the potential and effectiveness of using ARIMA modeling in the prediction of travel time to the urban roadway was also proven [17].

III. OPERATIONAL FRAMEWORK

Fig. 1 depicts the architectural framework of the study. The data gathered in the office of the Student Financial Assistance Unit was reprocessed and selected attributes were used in clustering. The clustered scholarship grants is then forecasted as to the number of grantees for the future years.

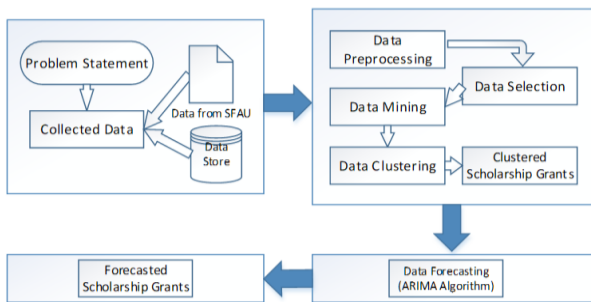


Fig. 1. The architectural design of the study.

A. Data

The data that were used are the following but are not limited as to wit: different scholarship grants that can be availed in SSCT, the gender of the grantees, and the number of grantees per scholarship grants per year.

B. Clustering

TABLE I: INDEXED SCHOLARSHIP GRANTS IN SSCT

	SCHOLARSHIP PROGRAM	MALE	FEMALE
1	Academic Scholarship	4	6
2	Athletic Scholarship	5	6
3	Barangay Scholarship	14	13
4	Choir	13	17
5	City Scholarship	147	247
6	CSSGP	7	15
7	Dance Troupe	5	5
8	ESGP-PA	61	167
9	LGU Basilisa	11	15
10	LGU Claver	11	19
11	PGMC (Platinum Group Metal Corporation)	7	7
12	Provincial Eskolaran	177	302
13	StuFAPS	9	21
14	Taganito Mining Corporation	27	32
15	Tulong Dunong 01	4	16
16	Tulong Dunong 02 (Cong. Bag-ao, Barbers/Romarate)	467	828
17	Tulong Dunong 02 (Cong. Matugas)	200	314

Clustering algorithms divide the group of objects into clusters, where objects in each cluster are similar to each other [18]. Indexed scholarship grants were grouped according to sex. Table I presents the datasets of indexed scholarship grants and its grantees according to sex from 2015 to 2017.

This algorithm is iterative and repeats for each object. It converges until the objects are stable. K-Means clustering is simple, and the necessary steps it follows are 1. Some clusters, K, is determined. 2. Assume a centroid or center of the K clusters. Any object can be randomly chosen and initialized as an initial centroid, or the first K objects can serve as the initial centroids. 3. The calculation of the distance of each object from each of the centroids. 4. Group the objects based on minimum distance (find the closest centroid for each object). Fig. 2 presents the flowchart of the k-means algorithm.

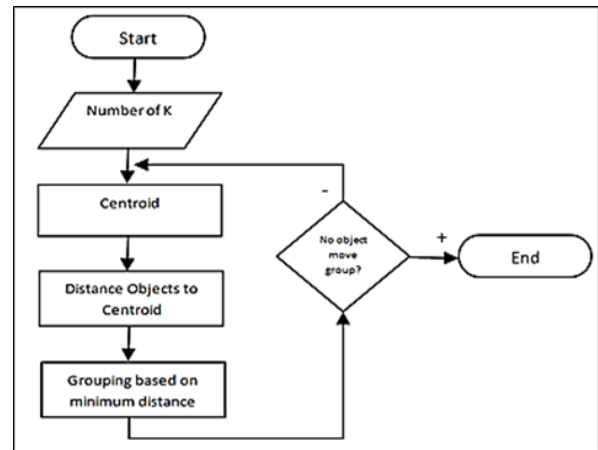


Fig. 2. K-Means algorithm flowchart.

Step 1: Initial value of centroids: Barangay Scholarship and ESGP-PA.

Let $C1$ and $C2$ denote the coordinate of the centroids, then $C1 = (14,13)$ and $C2 = (61,167)$

Step 2: Objects-Centroids distances denoted as OCD^0 : The Euclidean distance is used to obtain the distance. The distance matrix at iteration 0 is showed in Table II.

TABLE II: OBJECTS-CENTROID DISTANCES: ITERATION 0 DENOTED AS $OCDI^0$

	SCHOLARSHIP PROGRAM	MALE	FEMALE
1	Academic Scholarship	12.2065556	170.7922715
2	Athletic Scholarship	11.4017543	170.4611393
3	Barangay Scholarship	0.0000000	161.0124219
4	Choir	4.1231056	157.4928570
5	City Scholarship	269.1560885	117.4563749
6	CSSGP	7.2801099	161.3071604
7	Dance Troupe	12.0415946	171.4059509
8	ESGP-PA	161.0124219	0.0000000
9	LGU Basilisa	3.6055513	160.0124995
10	LGU Claver	6.7082039	156.2177967
11	PGMC (Platinum Group Metal Corporation)	9.2195445	168.8668114
12	Provincial Eskolaran	331.7981314	177.9915728
13	StuFAPS	9.4339811	154.9838701
14	Taganito Mining Corporation	23.0217289	139.2156600
15	Tulong Dunong 01	10.4403065	161.4001239
16	Tulong Dunong 02 (Cong. Bag-ao, Barbers/Romarate)	932.4344481	775.7299788
17	Tulong Dunong 02 (Cong. Matugas)	353.8318810	202.3116408

Table II represents the indexed data. The third column refers to the distance of each scholarship grants with corresponding value to the first centroid. On the other hand,

the last column is the distance of each object to the second centroid.

```
> results <- kmeans(mydatakmeans.features, 2)
> results
K-means clustering with 2 clusters of sizes 13, 4

Cluster means:
  MALE  FEMALE
1 13.69231 26.07692
2 247.75000 422.75000

Clustering vector:
[1] 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 2 2

within cluster sum of squares by cluster:
[1] 25065.69 287029.50
(between_SS / total_SS = 67.5 %)
```

Fig. 3. Clustering results using K-Means algorithm in r studio.

```
> results$cluster
[1] 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 2 2
```

Fig. 4. Result of the cluster function.

TABLE III: OBJECT CLUSTERING: ITERATION 0 DENOTED AS OCLI⁰

SCHOLARSHIP PROGRAM	Group 1	Group 2
Academic Scholarship	1	0
Athletic Scholarship	1	0
Barangay Scholarship	1	0
Choir	1	0
City Scholarship	0	1
CSSGP	1	0
Dance Troupe	1	0
ESGP-PA	1	0
LGU Basilisa	1	0
LGU Claver	1	0
PGMC (Platinum Group Metal Corporation)	1	0
Provincial Eskolaran	0	1
StuFAPS	1	0
Taganito Mining Corporation	1	0
Tulong Dunong 01	1	0
Tulong Dunong 02 (Cong. Bag-ao, Barbers/Romarate)	0	1
Tulong Dunong 02 (Cong. Matugas)	0	1

TABLE IV: OBJECTS-CENTROID DISTANCES: ITERATION 1 DENOTED AS OCLI¹

	SCHOLARSHIP PROGRAM	MALE	FEMALE
1	Academic Scholarship	6.7268120	280.2320467
2	Athletic Scholarship	6.0207973	279.7230773
3	Barangay Scholarship	5.5901699	269.1560885
4	Choir	7.6321688	266.1879036
5	City Scholarship	273.8179140	0.0000000
6	CSSGP	4.9244289	270.9686329
7	Dance Troupe	6.8007353	280.5851030
8	ESGP-PA	164.9128558	117.4563749
9	LGU Basilisa	4.9244289	268.9237810
10	LGU Claver	8.7321246	265.4806961
11	PGMC (Platinum Group Metal Corporation)	4.0311289	277.8488798
12	Provincial Eskolaran	336.4465039	62.6498204
13	StuFAPS	10.5000000	264.8018127
14	Taganito Mining Corporation	28.0401498	246.2214450
15	Tulong Dunong 01	7.4330344	271.6799588
16	Tulong Dunong 02 (Cong. Bag-ao, Barbers/Romarate)	937.0540273	663.2955601
17	Tulong Dunong 02 (Cong. Matugas)	358.5990100	85.4283325

Step 3: Objects clustering denoted as OCLI⁰: Based on the result generated by K-Means function in R where each indexed scholarship grants was assigned based on the minimum distance, thus, Academic Scholarship, Athletic Scholarship, Barangay Scholarship, Choir, CSSGP, Dance Troupe, ESGP-PA, LGU Basilisa, LGU Claver, PGMC, StuFAPS, Taganito Mining Corporation, Tulong Dunong 01 were assigned to group 1. On the other hand, City Scholarship, Provincial Eskolaran, Tulong Dunong 02 (Cong. Bag-ao, Cong. Barbers/Romarate), Tulong Dunong 02 (Cong. Matugas) were assigned to group 2.

Step 4: Iteration – 1, determine centroids: Compute the new centroid of each group based on the new memberships. Group 1 has twelve members. Thus the centroid is the average coordinated among the twelve members: C1 = (117/12, 172/12). On the other hand, group 2 has five members. Hence the centroid is the average coordinated among the three members: C2 = (1052/5, 1858/5).

Step 5: Iteration 1, Objects-Centroid distances denoted as OCDI¹: Repeat the process in step number 2 to obtain the new distance matrix based on the new groupings. The results of the process are shown in Table IV.

Step 6: Iteration-1, Objects Clustering denoted as OCLI¹. Assign each object based on the minimum distance. Based on the new distance matrix, the new group pattern presented in Table V.

TABLE V: OBJECT CLUSTERING: ITERATION 1 DENOTED AS OCLI¹

SCHOLARSHIP PROGRAM	Group 1	Group 2
Academic Scholarship	1	0
Athletic Scholarship	1	0
Barangay Scholarship	1	0
Choir	1	0
City Scholarship	0	1
CSSGP	1	0
Dance Troupe	1	0
ESGP-PA	1	0
LGU Basilisa	1	0
LGU Claver	1	0
PGMC (Platinum Group Metal Corporation)	1	0
Provincial Eskolaran	0	1
StuFAPS	1	0
Taganito Mining Corporation	1	0
Tulong Dunong 01	1	0
Tulong Dunong 02 (Cong. Bag-ao, Barbers/Romarate)	0	1
Tulong Dunong 02 (Cong. Matugas)	0	1

The results showed that OCLI¹ is equal to OCLI⁰. Comparing the groups of the last iteration reveals that the objects are stable. Thus, the computation process of the k-mean clustering has reached its stability, and no more repetition is needed. The final grouping is presented in Fig. 5.

```
Console -Rmeans/
> results$cluster
[1] 1 1 1 1 2 1 1 1 1 1 1 2 1 1 1 2 2
> table(mydatakmeans$SCHOLARSHIP.PROGRAM, results$cluster)

Academic scholarship      1 2
Athletic scholarship      1 0
Barangay scholarship      1 0
Choir                      1 0
City scholarship          0 1
CSSGP                     1 0
Dance Troupe              1 0
ESGP-PA                   1 0
LGU Basilisa              1 0
LGU Claver                1 0
PGMC (Platinum Group Metal corporation) 1 0
Provincial Eskolaran     0 1
StuFAPS (HS, GAD)        1 0
Taganito Mining Corporation 1 0
Tulong Dunong 01         1 0
Tulong dunong 02 (cong. Bag-ao, Cong. Barbers/Romarate) 0 1
Tulong Dunong 02 (Cong. Matugas) 0 1
```

Fig. 5. Clustered data.

C. Data Forecasting

The study utilized the ARIMA(1,0,0) model to forecast the increase for the next five years. The main concept of the algorithm is that a corresponding mathematical model describes a group of orderly time series data formed over time, and then future increase is forecasted. Fig. 6 shows the algorithm flowchart of ARIMA.

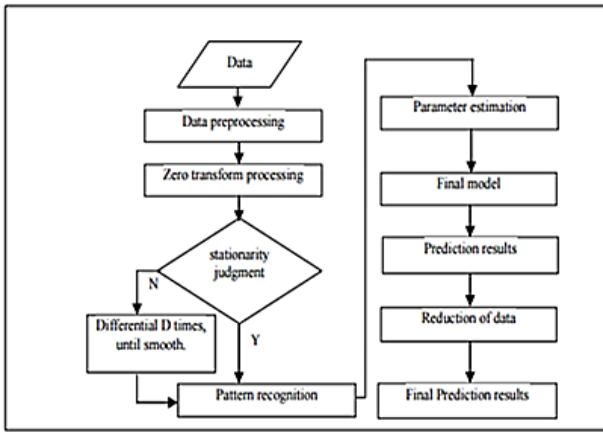


Fig. 6. ARIMA algorithm flowchart.

The study used R programming language, which runs in RStudio software for coding and plotting the graphs and analysis of the data set.

Date	Academic_Scholarship	Athletic_Scholarship	Barangay_Scholarship	Choir	CSSGP	Dance_Troupe	ESGP_PA	LGU_Basilisa	LGU_Claver	PGMC	StuFAPS	Taganito_Mining_Corporation	Tulong_Dunong_02_BBR	Tulong_Dunong_02_M
1 2015-01-01	5	4	8	14	13	7	126	5	0	1	3	3	3	1
2 2016-01-01	2	3	8	7	6	1	49	5	4	5	14	19	19	6
3 2017-01-01	3	4	11	9	3	2	53	16	18	8	13	37	37	13

Fig. 6. Indexed scholarship programs of cluster 1.

Date	City_Scholarship	Provincial_Eskolaran	Tulong_Dunong_02_BBR	Tulong_Dunong_02_M
1 2015-01-01	35	44	968	158
2 2016-01-01	151	205	181	229
3 2017-01-01	208	230	146	127

Fig. 7. The indexed scholarship program of cluster 2.

D. Trend Analysis

The R Language, which runs in RStudio software was utilized for the trend analysis of indexed data of the different scholarship grants offered in Surigao State College of Technology. Initially, the time series graph obtained in every cluster shows the behavior of indexed scholarship data. Fig. 9 and 10 show the time series graph of group one and two, respectively.

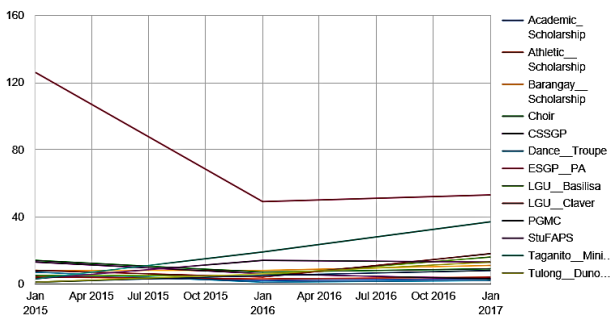


Fig. 9. Time series plot for cluster 1.

In group 1, ESGP-PA showed the highest number of grantees among the scholarship programs in Cluster 1. Moreover, it showed a very high mark during the year 2015 and showed a decreasing pattern starting from the year 2016 but has slightly increased in the year 2017.

On the other hand, Taganito Mining Corporation scholarship program showed an increasing pattern from the year 2015 to 2017 as shown in Fig. 9 that makes it next to

ESGP-PA as to the scholarship program with the most number of grantees in the cluster 1.

Furthermore, there is a declining pattern of many scholarship programs from 2015 to 2016 but then showed an increasing pattern from the year 2016 to 2017.

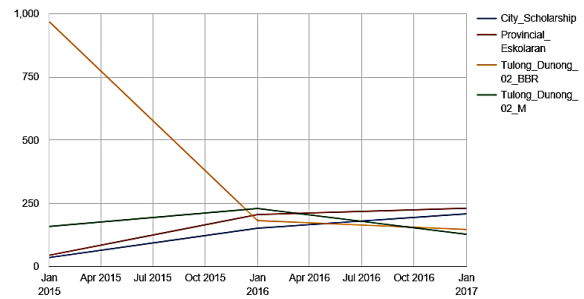


Fig. 10. Time series plot for cluster 2.

In group 2, Tulong Dunong 02 scholarship program by Cong. Bag-ao, Cong. Barbers and Cong. Romarate has the most number of grantees in the group in the year 2015. Later, it showed a decreasing pattern from 2015 to 2017.

On the contrary, City Scholarship and Provincial Eskolaran showed an increasing pattern from the year 2015 to the year 2017, which means an increase of grantees during that year was observed.

Moreover, Tulong Dunong 02 scholarship program under Cong. Matugas had an increasing pattern from the year 2015 to 2016 and a decreasing pattern from the year 2016 to 2017 as shown in Fig. 10.

E. Forecasting

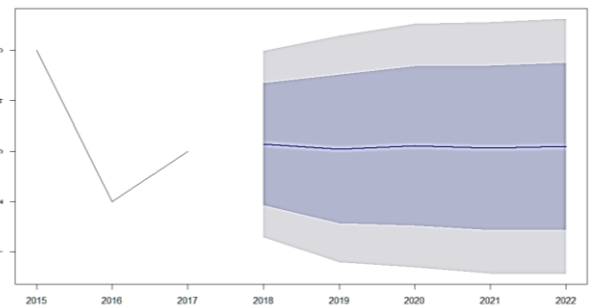


Fig. 11. Forecasted grantee for academic scholarship grant from 2017-2022.

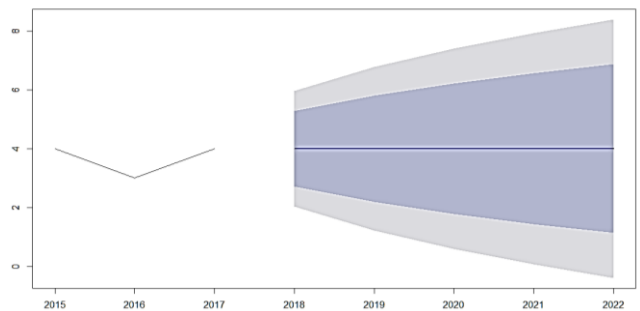


Fig. 12. Forecasted grantees for athletic scholarship grant from 2017-2022.

ARIMA(1,0,0) model was used to forecast the number of grantees for the next five years. Fig. 11-23 showed the graph of the predicted grantees for each scholarship grants from 2017 to 2022 having a high-95 % prediction.

A steady pattern is shown for the forecasted grantees for the

Academic Scholarship grant from the year 2017 to 2022 having been forecasted an average of 3 grantees per year with the highest number of 5 on the year 2022 is evident in the Fig. 11.

The predicted grantees for the Athletic Scholarship grant showed data stability from 2017 to 2022. The prediction signifies that the forecasted data reach the same number of the grantee in SSCT during the year 2015 to 2017, as shown in Fig. 12.

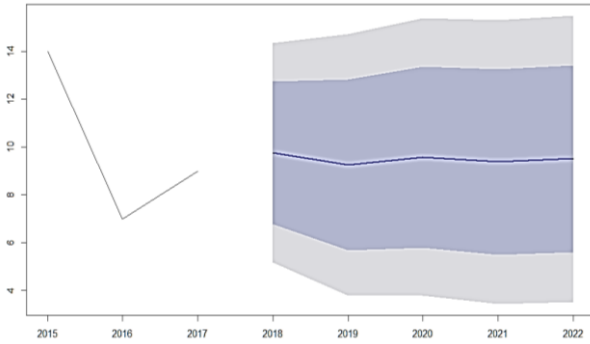


Fig. 13. Forecasted grantee for choir scholarship grant from 2017-2022.

The forecasted grantees for the Choir Scholarship grant from 2017 to 2022 in SSCT showed a slightly increased pattern from 2017 to 2018, as shown in Fig. 13 graph of the forecast.

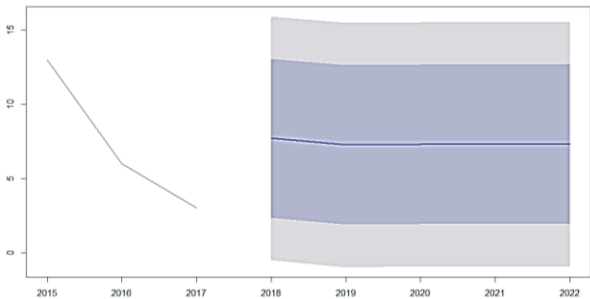


Fig. 14. Forecasted grantee for CSSGP scholarship grant from 2017-2022.

The forecasted grantees for the CSSGP Scholarship grant from 2017 to 2022 showed an increased forecast pattern from 2017 to 2018 and a steady trend from 2018 to 2022 based on Fig. 14.

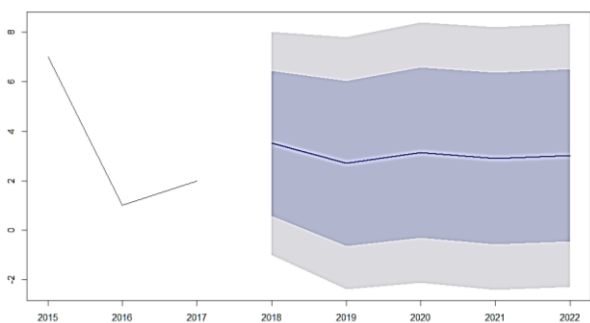


Fig. 15. Forecasted grantee for dance troupe and ESGP-PA scholarship grants from 2017-2022.

The forecasted grantees for Dance Troupe and ESGP-PA scholarship grants from 2017 to 2022 showed an increased pattern from the year 2017 to 2018 and a slightly reduced

pattern from 2018 to 2019 as evident in Fig. 15. Moreover, a prediction of nine with high-11 grantees on the year 2019 for the Barangay Scholarship grant is depicted with data stability from the year 2020 to 2022

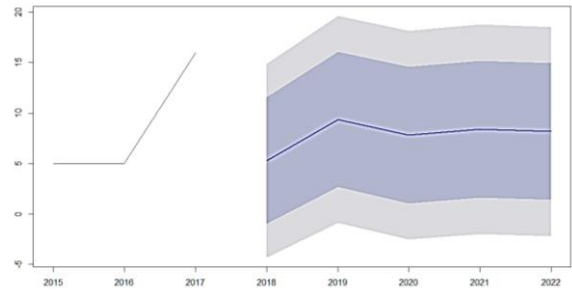


Fig. 16. Forecasted grantee for LGU Basilisa and Barangay scholarship grant from 2017-2022.

Forecasted grantees for the LGU Basilisa and Barangay Scholarship grants from 2017 to 2022 showed a rapidly decreased pattern from 2017 to 2018. Moreover, there is an increased design from 2018 to 2019, as evident in the graph. Further, there is a slightly decreased pattern from the year 2019 to 2020, as shown in Fig. 16.

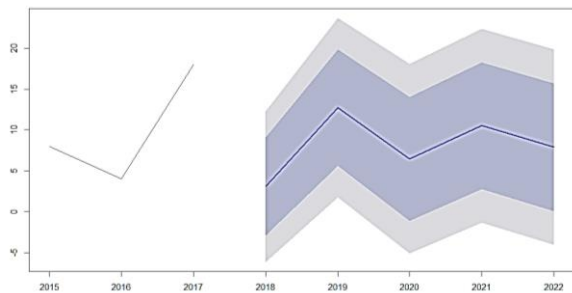


Fig. 17. Forecasted grantee for LGU claver scholarship grant from 2017 to 2022.

Forecasted grantees for the LGU Claver Scholarship grant from 2017 to 2022 showed a rapidly decreased pattern from 2017 to 2018. Moreover, there is an increased design of the grant from 2018 to 2019, as evident in the graph. From the year 2019 to 2020, it is evident that there is a slightly decreased pattern of the grantees. Further, there is a slightly increased pattern from the year 2020 to 2021 and a decrease after, as shown in Fig. 17.

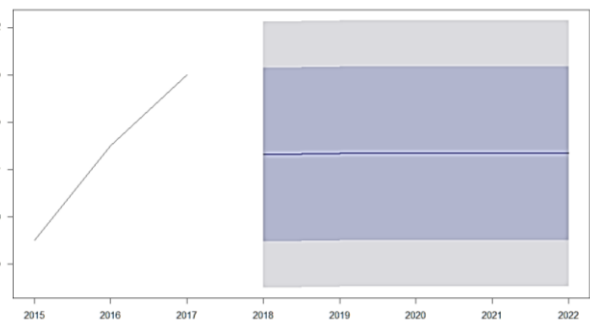


Fig. 18. Forecasted grantee for PGMC and City Scholarship grants from 2017 to 2022.

Forecasted grantees for PGMC and City Scholarship grants showed a stability pattern from 2018 to 2022 but a rapid decrease from the year before, as shown in Fig. 18.

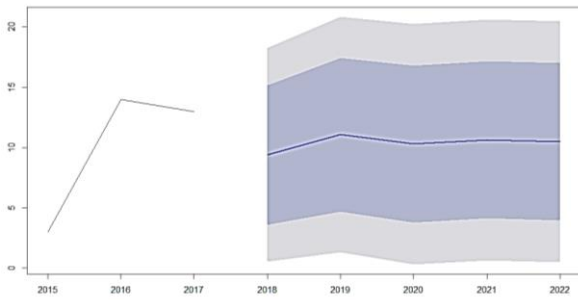


Fig. 19. Forecasted grantee for StuFAPS and TMC scholarship grants from 2017 to 2022.

Forecasted grantees for the StuFAPS Scholarship grant from 2017 to 2022 showed a rapidly decreased pattern from 2017 to 2018. Moreover, there is an increased design from 2018 to 2019, as evident in the graph. Further, there is a slightly decreased pattern from the year 2019 to 2020, and a stable graph in the following years is shown in Fig. 19.

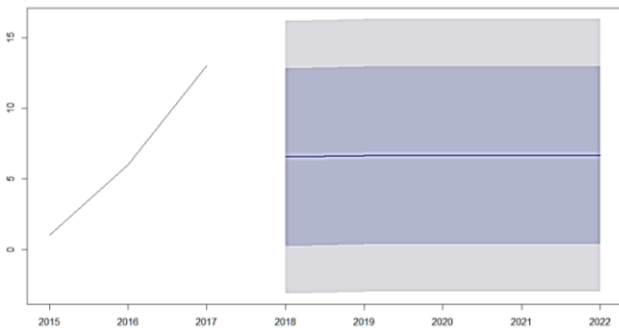


Fig. 20. Forecasted grantee for TD-01 Scholarship grant from 2017 to 2022.

Forecasted grantees for Tulong Dunong-01 showed a decrease of the grantees from 2017 to 2018 and a stability pattern from 2018 to 2022, as shown in Fig. 20.

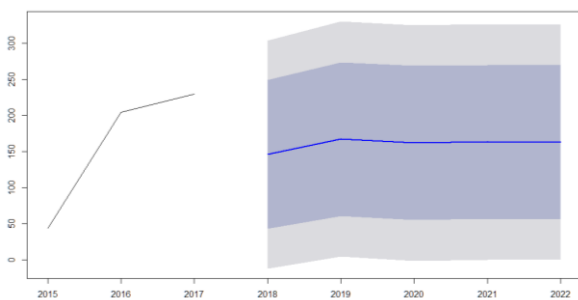


Fig. 21. Forecasted grantee for provincial Eskolaran grant from 2017 to 2022.

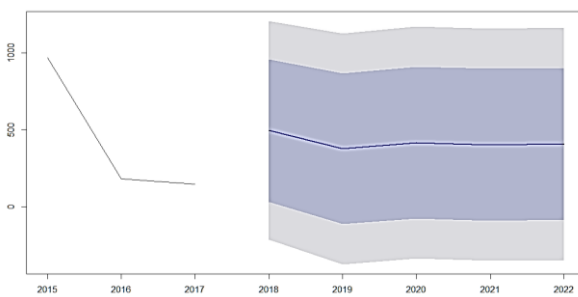


Fig. 22. Forecasted grantee for TD-022 scholarship grant from 2017 to 2022.

The forecasted grantees for Provincial Eskolaran from 2017 to 2022 showed a decreasing pattern from the year 2017

to 2018. Moreover, there is a slightly increased pattern from 2018 to 2019 and a stable pattern onwards based in Fig. 21.

The forecasted grantees for Tulong Dunong-02 under Cong. Bag-ao, Cong. Barbers and Romarate Scholarship grant from 2017 to 2022 showed an increased pattern from the year 2017 to 2018. Moreover, there is a slightly reduced pattern from 2018 to 2019 and a steady pattern from 2020 to 2022, as shown in Fig. 22.

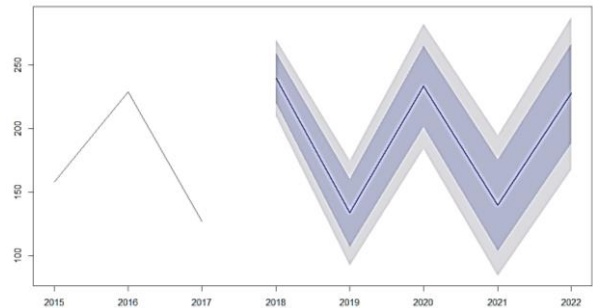


Fig. 23. Forecasted grantee for TD-0222 scholarship grant from 2017 to 2022.

Forecasted grantees for the Tulong Dunong-02 under Congressman Matugas Scholarship grant from 2017 to 2022 showed a rapidly increased pattern from 2017 to 2018. Moreover, there is a decreased design of the grant from 2018 to 2019, as evident in the graph. From the year 2019 to 2020, it is evident that there is a rapidly increased pattern of the grantees. Further, there is a slightly decreased pattern from the year 2020 to 2021 and an increase after, as shown in Fig. 23.

IV. CONCLUSION

Predicting grantees for each scholarship grants can be beneficial to the sponsoring agents since it will give them insight as to the number of their future grantees. This allows them to better prepare for budget allocation. In the study, the use of k-means clustering algorithm proved to be a useful tool that grouped every scholarship grants offered in SSCT. The algorithm was able to group scholarship grants that share common features according to their centroid distances. Furthermore, it is evident that the ARIMA algorithm with ARIMA(1,0,0) model helped in predicting the possible grantees in each scholarship grants.

In cluster 1 of the indexed scholarship grants, scholarship grant under LGU Claver showed the highest number of increase in grantees, specifically during the year 2019 with the highest number. On the other hand, Tulong Dunong-02 scholarship grant under Congressman Matugas has the highest predicted grantees for the year 2018, 2020, and 2022, which is higher than those from the year 2015 to 2017. Further, data stability is depicted among other identified scholarship grants from the two clusters.

REFERENCES

- [1] W. Wei, J. Han, J. Kong, and H. Xia, "Prediction of the Scholarship Using Comprehensive Development," in *Proc. 4th Int. Conf. Enterp. Syst. Adv. Enterp. Syst. ES, 2016*, pp. 183-188.
- [2] I. A. Khan and J. T. Choi, "An application of educational data mining (EDM) technique for scholarship prediction," *Int. J. Softw. Eng. Its Appl.*, vol. 8, no. 12, pp. 31-42, 2014.

- [3] E. Şuşnea, "Using data mining techniques in higher education," *High Educ.*, vol. 1, no. 1, pp. 68-72, 1996.
- [4] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 2012.
- [5] P. Kaur, M. Singh, and G. S. Josan, "Classification and prediction based data mining algorithms to predict slow learners in education sector," *Procedia Comput. Sci.*, vol. 57, pp. 500-508, 2015.
- [6] S. K. Yadav and S. Pal, "Data mining : A prediction for performance improvement of engineering students using classification," *World Comput. Sci. Inf. Technol. J. WCSIT*, vol. 2, no. 2, pp. 51-56, 2012.
- [7] C. Jinyin, L. Xiang, Z. Haibing, and B. Xintong, "A novel cluster center fast determination clustering algorithm," *Appl. Soft Comput. J.*, vol. 57, pp. 539-555, 2017.
- [8] G. Gan and M. K. P. Ng, "K-Means clustering with outlier removal," *Pattern Recognit. Lett.*, vol. 90, pp. 8-14, 2017.
- [9] J. Agarwal, "Crime analysis using k-means clustering," *Int. J. Comput. Appl.*, vol. 83, no. 4, pp. 975-8887, 2013.
- [10] P. Gupta, A. S. Sabitha, and T. Choudhury, "Terrorist attacks analysis using clustering algorithm," © *Springer Nat. Singapore Pte Ltd.*, pp. 317-328, 2018.
- [11] E. Keogh and S. Kasetty, "On the need for time series data mining benchmarks," in *Proc. 8th ACM SIGKDD Int. Conf. Knowl. Discov. data Min. - KDD '02*, 2002, p. 102.
- [12] X. Wang, A. Mueen, H. Ding, G. Trajcevski, P. Scheuermann, and E. Keogh, "Experimental comparison of representation methods and distance measures for time series data," *Data Min. Knowl. Discov.*, vol. 26, no. 2, pp. 275-309, 2013.
- [13] T. Warren Liao, "Clustering of time series data — A survey," *Pattern Recognit.*, vol. 38, no. 11, pp. 1857-1874, 2005.
- [14] W. Hongbin, D. Shuangyu, D. Jianzhuo, R. Ming, and D. Ming, *Study on Condition Pre-warning Method of Power Transformer Based on Load Time Series Model*, pp. 223-227, 2015.
- [15] Q. Liu, X. Liu, B. Jiang, and W. Yang, *Forecasting Incidence of Hemorrhagic Fever with Renal Syndrome in China Using ARIMA Model*, 2011.
- [16] S. Promprou, M. Jaroensutasinee, and K. Jaroensutasinee, "Forecasting dengue haemorrhagic fever cases in Southern Thailand using ARIMA Models," *Dengue Bull.*, vol. 30, pp. 99-106, 2006.
- [17] D. Billings and Y. Jiann-Shiou, "Application of the ARIMA models to urban roadway travel time prediction — A case study," in *Proc. IEEE Int. Conf. on Systems, Man and Cybernetics*, 2006, pp. 2529-2534.
- [18] D. Hand, D. Hand, H. Mannila, H. Mannila, P. Smyth, and P. Smyth, *Principles of Data Mining*, 2001.



Allemar Jhone P. Delima was born at Surigao City, province of Surigao del Norte, Philippines. He completed his master degree in information technology at Surigao State College of Technology, Surigao City, Philippines in last 2016. With his quest in improving his professional career, he took a doctoral degree and is now a student of the Doctor in information technology at the Technological Institute of the Philippines – Quezon City, Philippines. He is currently connected in Surigao State College of Technology as a faculty in the College of Engineering and Information Technology. His research interest is in data mining and data analytics.