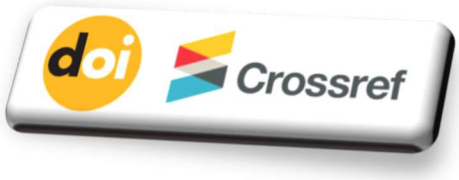
Journal, Advances in Mathematical & Computational Sciences
Vol. 11 No. 3, September, 2023
www.isteams.net/mathematics-computationaljournal



Journal of Advances in Mathematical & Computational Sciences
An International Pan-African Multidisciplinary Journal of the SMART Research Group
International Centre for IT & Development (ICITD) USA

Available online at https://www.isteams.net/ mathematics-computationaljournal.info
CrossREF Member Listing - https://www.crossref.org/06members/50go-live.html

OYSCATECH Students' Admission: Exploratory Data Analysis and Forecast Model Formulation

[1]Oyewo Damilola T., [2]Awobode Gbolagade O. & [2]Ojo Victoria O.
[1]Computer Science Department, Oyo State College of Agriculture and Technology, Igboora, Nigeria
[2]Statistics Department, Oyo State College of Agriculture and Technology, Igboora, Nigeria
***Corresponding author**: dammyoyewo@gmail.com
Phone: +2347033278662

ABSTRACT

The quest for institutional advancement and improved profitability has made managers to take a second look at data, and data has proven to have a voice if it is queried in the right manner. Admission record in an educational institution is capable of dictating the tone for financial stability and educational standard improvement. OYSCATECH Igboora is earnestly in dire need of scientific tools to assist in the analytics of the existing data and forecasting of future statistics towards future admission processes. A five year admission record was obtained from the registry of OYSCATECH between 2017/2018 and 2021/2022 for both Higher National Diploma (HND) and National Diploma(ND) admissions. Naïve approach, exponential smoothening, Linear regression and 2 years moving average models were used to forecast the records. When the models were evaluated for performance metrics using accuracy, mean absolute error, mean square error and mean absolute percent error , results for HND record showed that Linear regression had an accuracy of 77%, which was highest from among the tested models, and MSE of 21,031 which was also the least score obtained. For the ND record too, linear regression also showed a sterling performance of 98% and 812 in both the accuracy and MSE evaluations respectively. Exploratory data analysis was performed using Tableau, and it revealed interesting insights which can guide the management officers in steering the wheel of the college aright.

Keywords: Data analysis, forecasting models, EDA, Student admission

Oyewo Damilola T., Awobode Gbolagade O. & Ojo Victoria O. (2023): OYSCATECH Students' Admission: Exploratory Data Analysis and Forecast Model Formulation. Journal of Advances in Mathematical & Computational Science. Vol. 11, No. 3. Pp 21-36 dx.doi.org/10.22624/AIMS/MATHS/V11N3P1 Available online at www.isteams.net/mathematics-computationaljournal.

1. INTRODUCTION

In this present world dispensation, tertiary institutions are inundated with the responsibility of self-sustenance for salary payment, award of scholarships to both academic and non-academic staff members, infrastructural development and for general day-to-day running of the institutions.

This task requires that the institutions management come to their wit to stay afloat. It is observed that maximizing yearly admission potentials of the institution is one of the easiest ways to generate substantial revenue. According to a survey of more than 1,000 senior executives conducted by PwC, highly data-driven organizations are three times more likely to report significant improvements in decision-making compared to those who rely less on data (Tim, 2021).

A scientific method of analyzing past and present admission records for the purpose of smart decision making is known as forecasting. Tertiary Institution admission process can vary widely depending on the level of education and the institution or program being applied to. Forecasting helps to control influx of students to specific departments based on certain projections. Josephine (2017) confirmed that the influx of students is such that there could be insufficient seat in many lecture rooms, shortage of hostel accommodation and traffic congestions.

Forecasting is the process of making predictions or estimates about future events based on past and present data. Forecasting plays a crucial role in decision-making in various fields, including business, finance, economics, weather, and sports. It enables organizations to anticipate future trends, identify potential opportunities and risks, and make informed decisions to achieve their goals (Hyndman and Athanasopoulous, 2018). It involves using historical data to predict future trends and patterns. The process of forecasting involves collecting and analyzing data, identifying patterns, and using statistical models to make predictions.

There are various methods and tools used in forecasting, including statistical models, time series analysis, and machine learning algorithms (Rameshwar, 2022). In recent years, the development of advanced technologies such as big data analytics and artificial intelligence has significantly improved the accuracy and speed of forecasting (Makridakis et al.,2020). According to Bernardita (2023), out of the colossal volume of data created every day, a mere 0.5% is actually analyzed and used for data discovery, improvement, and intelligence. With the increasing availability of data and advancements in technology, forecasting has become more accurate and efficient. Tools such as Excel, Python and Tableau have made it easier for businesses and organizations to analyze and interpret data to make better decisions. These tools have the capability to allow users to create visualizations and interactive dashboards to track key performance indicators and identify trends in real-time.

Student`s admission data analytics and forecast help each institution management to plan ahead annually towards the expected students population. This is lacking in OYSCATECH. Hence, this study. In this study, we explored the development of admission forecasting models using Microsoft Excel and Tableau for Exploratory Data Analysis (EDA) in Oyo State College of Agriculture and Technology (OYSCATECH), Igboora.

The objectives of the Study are as follows:

i. to obtain 5 years admission record.
ii. to formulate a forecast in Microsoft excel using 2 years moving average, naïve approach, exponential smoothening and Linear regression.
iii. to evaluate the models by obtaining the accuracy, mean absolute error, mean square error and mean absolute percent error

2. MATERIALS AND METHODS

Data analysis is the process of collecting, modeling, and analyzing data using various statistical and logical methods and techniques. All these were carried out in the course of this study. We obtained 5 years National Diploma (ND) and Higher National Dipolma (HND) between 2017/2018 and 2021/2022 academic session admission record from the college's admissions unit of the Registry. The data collected includes Registration number, student name, gender, department, state and local government as shown in Figure 2.1.

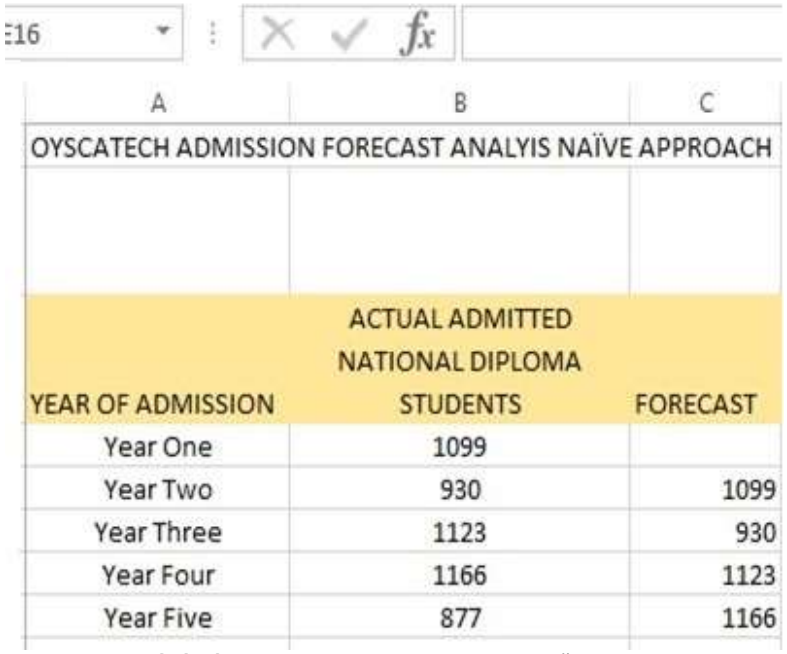
	A	B	C	D	E	F	G
1	ND 2020/2021 ADMISSIONS LIST						
2	S/N	REG NO.	NAME	GENDER	DEPT	STATE	LGA
3	1	21578094IA	Adedeji Waris Adeyemi	M	ABE	OYO	IBARAPA-CENTRAL
4	2	21859647CF	Adedokun Azeez Adebayo	M	ABE	OYO	OLUYOLE
5	3	21878361DA	Adetunji Abiodun Abdullahi	M	ABE	OSUN	ISOKAN
6	4	21950760IA	Adetunji Mukadas Tolani	M	ABE	OYO	KAJOLA
7	5	21890961BF	Adeyemo Rasheed Aremu	M	ABE	OYO	IBADAN NORTH EAST
8	6	21946985FA	Afolabi Samuel Kayode	M	ABE	OYO	SAKI-EAST
9	7	21879668FA	Akanni Uthman Ifedapo	M	ABE	OYO	IBADAN NORTH EAST
10	8	21850812CF	Akinleye Akintomiwa Phillip	M	ABE	OYO	AKINYELE
11	9	21853783IF	Akinnehin Babatunde Samuel	M	ABE	OYO	IBADAN NORTH EAST
12	10	21879446CA	Akintayo Olamide Anu	M	ABE	OYO	ONA-ARA
13	11	21850146DA	Alani Pelumi Sunday	M	ABE	OYO	EGBEDA

Figure 2.1: Screenshot of a section ND Admission 2020/2021 List

Microsoft Excell 2020 was used to develop different forecast models on the record. The models used include Naïve Approach, Linear Regression, 2-years Moving Average and Exponential Smoothing.

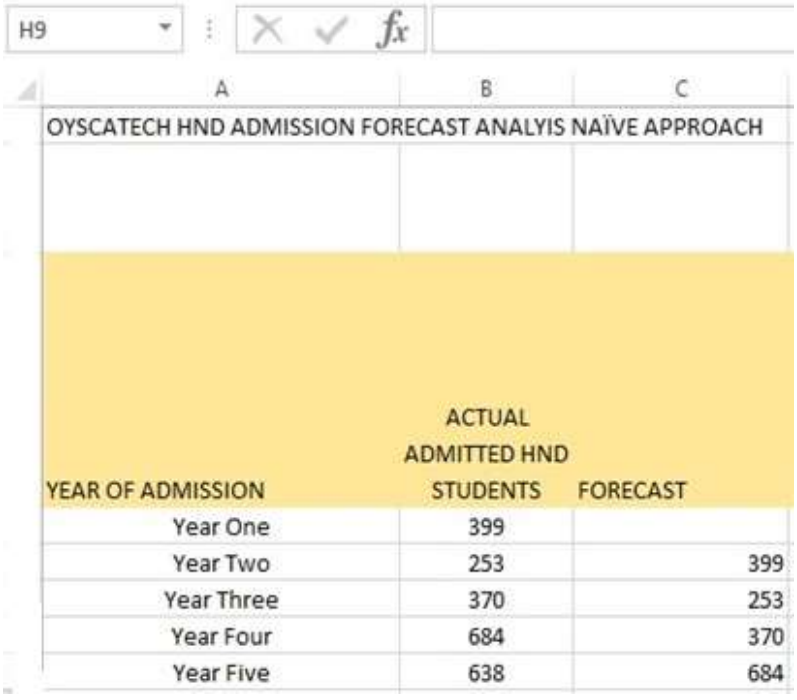
Naïve Approach

This forecasting model comes with the assumption that the record of the previous year is what is obtainable in the current year as shown in figure 2.2 and figure 2.3. Minimum amounts of effort and record data manipulation is employed here in arriving at a forecast. The summary of the record obtained for ND and HND are used to obtain a yearly forecast. The overall expectation is to obtain the accuracy. The accuracy value will determine whether the approach is good from the past record.



OYSCATECH ADMISSION FORECAST ANALYIS NAÏVE APPROACH		
YEAR OF ADMISSION	ACTUAL ADMITTED NATIONAL DIPLOMA STUDENTS	FORECAST
Year One	1099	
Year Two	930	1099
Year Three	1123	930
Year Four	1166	1123
Year Five	877	1166

Figure 2.2: Screenshot showing HND naïve approach



OYSCATECH HND ADMISSION FORECAST ANALYIS NAÏVE APPROACH		
YEAR OF ADMISSION	ACTUAL ADMITTED HND STUDENTS	FORECAST
Year One	399	
Year Two	253	399
Year Three	370	253
Year Four	684	370
Year Five	638	684

. Figure 2.3: Screenshot showing ND naïve approach

Journal, Advances in Mathematical & Computational Sciences
Vol. 11 No. 3, September, 2023
www.isteams.net/mathematics-computationaljournal

Exponential Smoothing Model

Exponential smoothing attempts to smoothen the line in the graph of the actual record, and to remove the various sharp ends present in the graph. The dampening factor was set at 70% for actual admission value, and 30% for the forecast value and the results obtained are shown in figure 2.4. The values of the forecast are affected by the settings of the dampening factor; there is a certainty the values of the evaluation metrics change when the dampening factors are altered.

D10

	A	B	C
1	ND FORECAST USING EXPONENTIAL SMOOTHING APPROACH		
2			
3	YEAR OF ADMISSION	ACTUAL ADMITTED NATIONAL DIPLOMA STUDENTS	FORECAST
4	Year One	1099	#N/A
5	Year Two	930	1099
6	Year Three	1123	980.7
7	Year Four	1166	1080.31
8	Year Five	877	1140.293

Figure 2.4: Screenshot showing exponential smoothening approach

Linear Regression Model

Linear regression makes use of the linear expression

$y = a + bx.$

Where y represents Dependent variable, a stands for intercept, b denotes slope while x is the independent variable. All these were used in obtaining a forecast for the admission record. Seasonality index was also calculated before the model was evaluated for performance measures. The period of years under consideration were converted to period as shown in figure 2.5, these were used to determine the variables of the linear expression.

Journal, Advances in Mathematical & Computational Sciences
Vol. 11 No. 3, September, 2023
www.isteams.net/mathematics-computationaljournal

J5 =H5/C5

	A	B	C	D	E	F
	OYSCATECH ND ADMISSION FORECAST ANALYSIS USING LINEAR REGRESSION APPROACH					
	PERIOD	YEAR OF ADMISSION	ACTUAL ADMITTED HIGHER NATIONAL DIPLOMA STUDENTS	FORECAST	SEASONALITY	LT + SEASONALITY
	1	Year One	1099	1081	1.06	1143.00
	2	Year Two	930	1060	0.90	948.62
	3	Year Three	1123	1039	1.08	1123.00
	4	Year Four	1166	1018	1.12	1142.66
	5	Year Five	877	997	0.84	841.89
	Year Four		1166		1123	
	Year Five		877		1166	

Figure 2.5: Screenshot showing Linear regression model

2-Years Moving Average

This approach assumes the average of the last two actual records as the forecasted value. So, we calculated the average of years one and two to obtain a forecast for year three, ditto for others. There was no forecast record for year one and year two as shown in figure 2.6.

Journal, Advances in Mathematical & Computational Sciences
Vol. 11 No. 3, September, 2023
www.isteams.net/mathematics-computationaljournal

I8

	A	B	C
1	FORECAST ANALYSIS USING TWO-YEARS MOVING AVERAGE APPROACH		
2			
3	YEAR OF ADMISSION	ACTUAL ADMITTED NATIONAL DIPLOMA STUDENTS	FORECAST
4	Year One	1099	
5	Year Two	930	
6	Year Three	1123	1014.5
7	Year Four	1166	1026.5
8	Year Five	877	1144.5

Figure 2.6: Screenshot Showing 2-Years Moving Average forecast

3. Exploratory Data Analysis (EDA) of the Obtained Record

The records were imported into the tableau for the purpose of graphical analysis and visualization, this will enable stakeholders to easily understand the deep insights locked up in the raw students' admission record. Figure 3.1 shows a pie chart showing the distribution of admitted students in 2021/2022 session. It can be clearly seen that 86% of the population are Oyo indigenes, and 14% are shared by other states. Osun and Ogun states which border Oyo state are seen with a great chunk of the non-indigenes portion. Though this is understandable, but efforts must be geared towards increasing the indigenes' population because the Oyo state government is highly subsiding the institution with taxes paid by Oyo state workers.

Figure 3.2 shows the distribution of admitted students within the 5-year period, admission year 2020/2021 attracted the highest number of 684 for the HND students and the reason for the drop of the figure to 638 in 2021/2022 admission year should be investigated because if the reason for the surge in students admitted was as a result of mounting additional programs, no program could be said to be unmounted in the following year.

Year 2017/2018 and 2018/2019 had a similar pattern of reduced number of admitted students too, this needs to be further investigated as there must always be a yearly improvement unless there is a need for control by the management based on certain conditions. Figure 3.3 showed the distribution of total admitted students with a surprising drop in the admitted student numbers from 1,850 in 2020/2021 session to 1,515 in 2021/2022 session, this should call for attention. Year 2017/2018 and 2018/2019 also had a similar pattern of reduced number of admitted students.

Journal, Advances in Mathematical & Computational Sciences
Vol. 11 No. 3, September, 2023
www.isteams.net/mathematics-computationaljournal

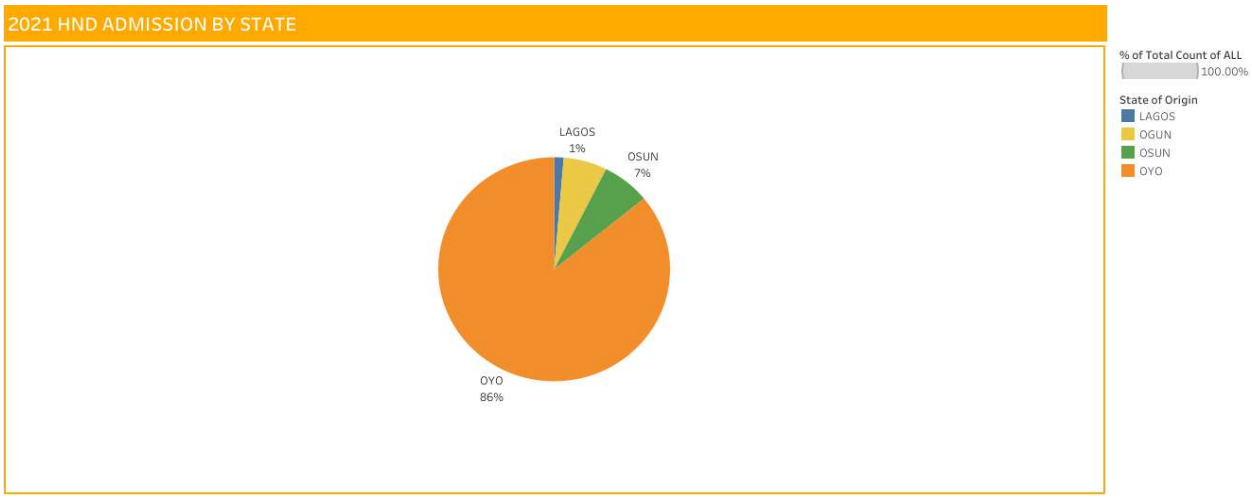


Figure 3.1: Pie chart showing the distribution of HND Admitted students by state of origin in 2021

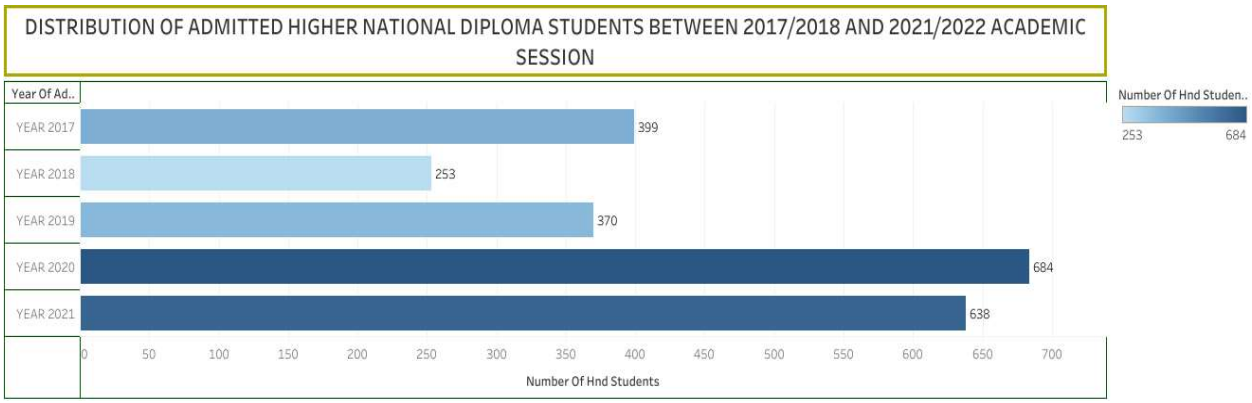


Figure 3.2: HND Admission distribution between 2017/2018 and 2021/2022 session

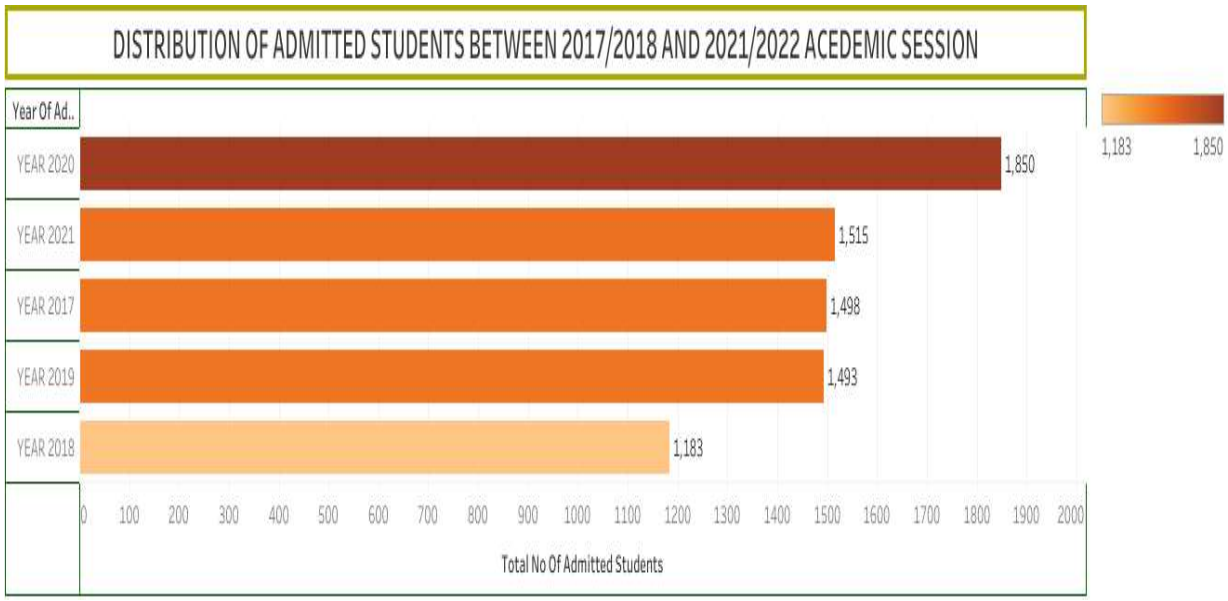


Figure 3.3: Total Student Admission Chart (2017-2021)

Figure 3.4 shows a sharp decline in ND students' admission from 1,166 in 2020/201 session to 877 in 2021/2022 session. This is quite alarming as it will have a multiplier effect in the future HND admission of the 2021/2022 ND class.

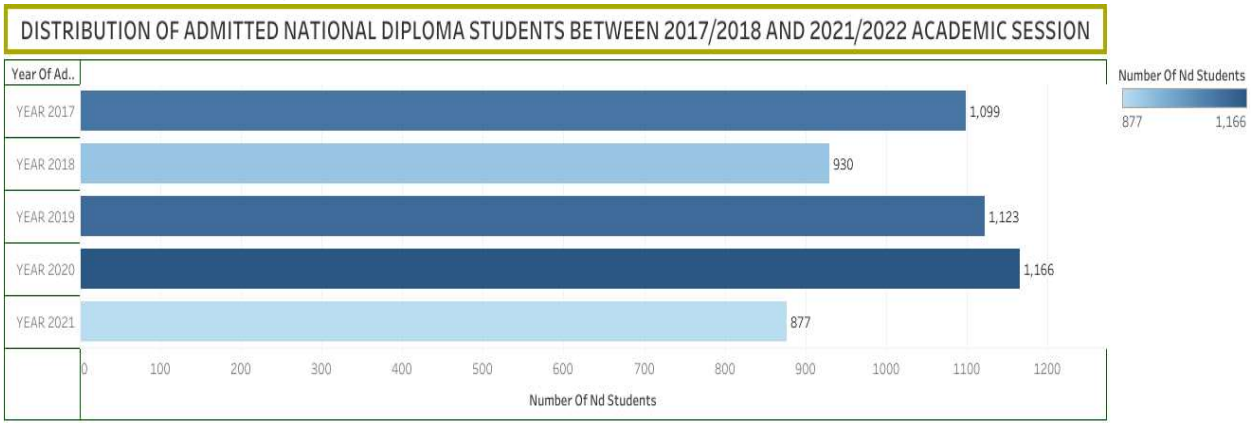


Figure 3.4: ND Admission Distribution between 2017/2018 and 2021/2022

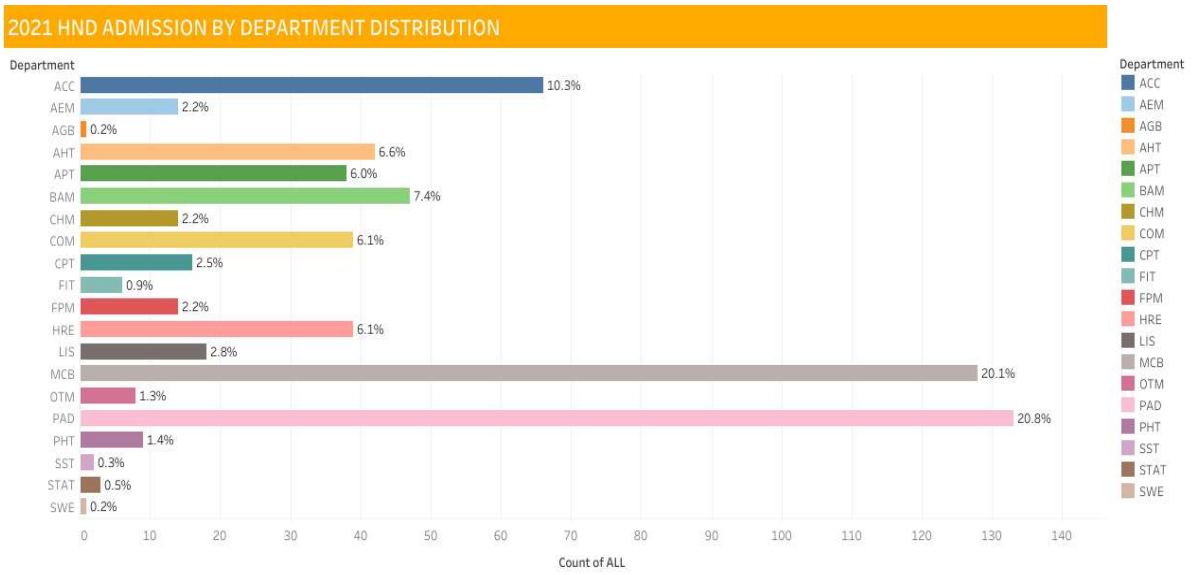


Figure 3.5 HND Admission Distribution by Department

Using the HND and 2021/2022 admission year as a case study, figure 3.5 describes the admission distribution by department. Just two (2) departments account for more than 40% of the total admitted students; this positions them as the departments as the golden fleece of the college. Soil water enginering, Statistics, Agric business and Soil science technology departments all contribute just around 1% of the admitted students population; they need to do more vigorous admission sensitisation in subsequent years.

4. RESULTS

Naïve approach model was used to analyze the data obtained within the 5-year period for the HND and ND students. The evaluation shows that Mean Absolute Deviation (MAD) for the period was 155.75 and 173.5 respectively meaning that the average error/deviation possible within those years was 155.75 and 173.5 more or less than the actual number of students admitted in those years than was eventually realized. The relationship between the ND actual record and forecasted record is depicted in fig 2.4 while that of the HND is shown in figure 2.5. Mean Squared Error (MSE) achieved 33,929.25 and 37,795 for ND and HND; however, the lower the mean squared error, the better the model. Mean absolute percent error is 36% and 18%, this makes the accuracy to be 64% and 82% respectively.

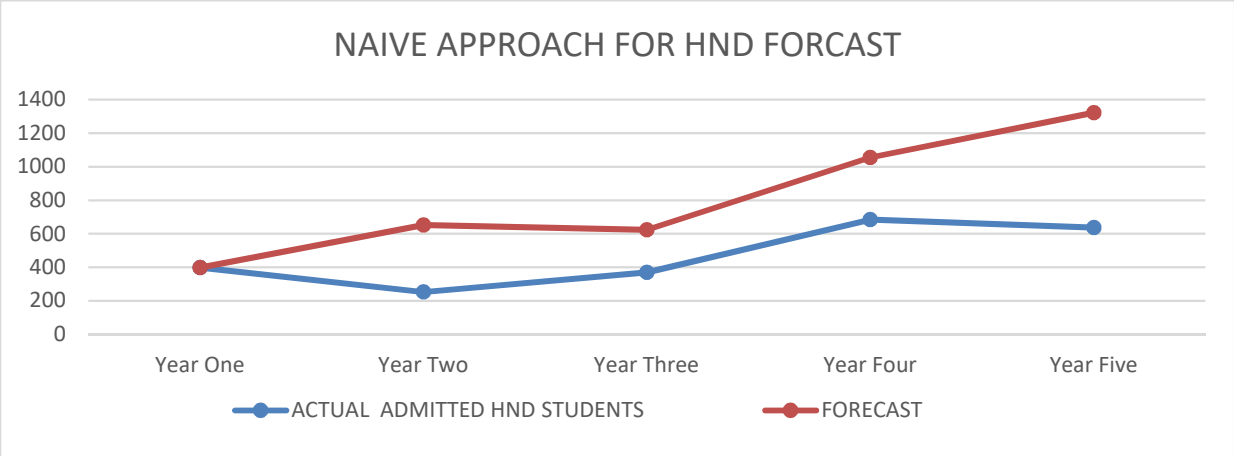


Figure: 4.1 Naïve approach for HND forecast

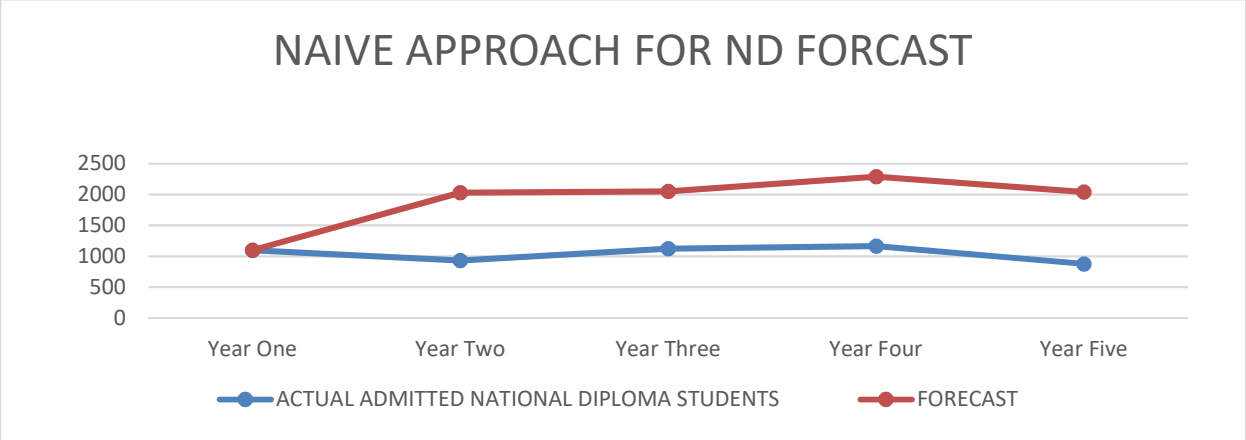


Figure 4.2 Naïve Approach for ND Forecast

4.1 Linear Regression Graph

When Linear Regression model was used to forecast the available record for HND and ND admission, the evaluation shows that MAD for the period was 116.8 and 24.2, meaning that the error/deviation possible within those years was 116.8 and 24.2 respectively more or less than the actual number of students admitted in those years than was eventually realized. Mean squared error achieved 21,031 for HND and 812 for ND. Mean absolute percent error is 23% and 2% which makes the accuracy to be 77% and 98% respectively. Figure 4.3 shows the convergence of the HND student population between the forecasted linear graph and the actual admitted students at the fifth year. For the HND record shown in figure 4.3, 'LT + Seasonality' graph helped to introduce seasonality into the linear graph, this follows the pattern of the actual graph. Figure 4.4 however shows that the actual graph and the 'LT + Seasonality' graph followed exactly the same pattern

Journal, Advances in Mathematical & Computational Sciences
Vol. 11 No. 3, September, 2023
www.isteams.net/mathematics-computationaljournal

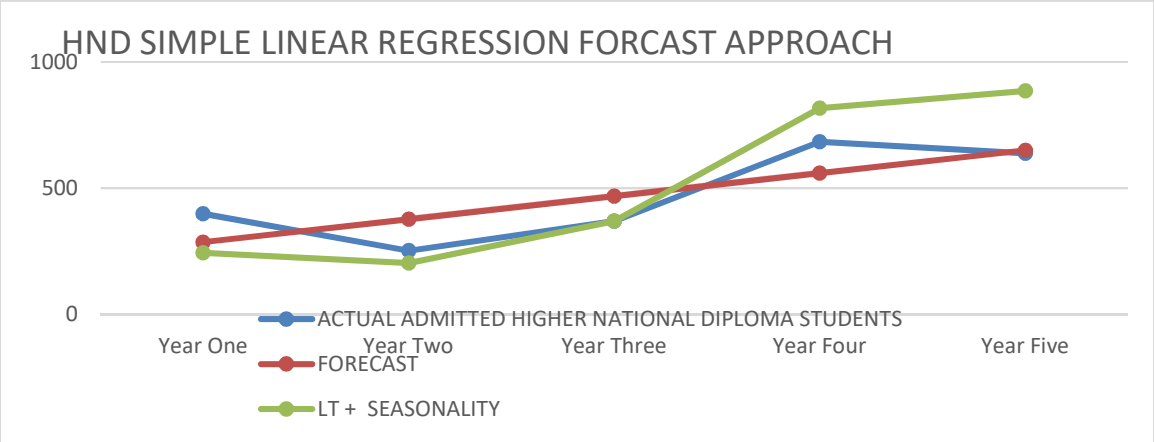


Figure 4.3 HND simple linear regression forecast

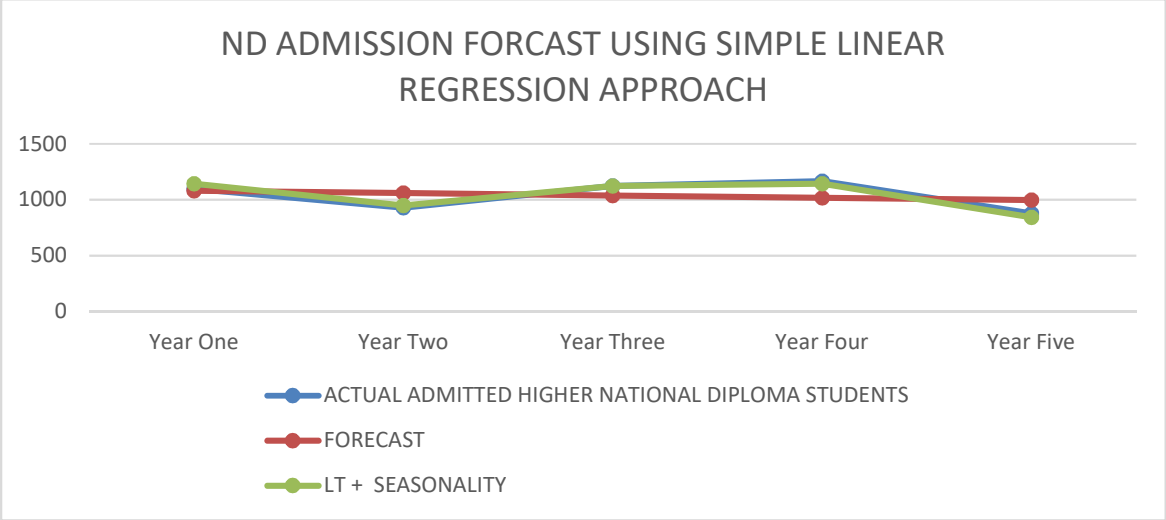


Figure: 4.4 ND linear regression forecast

4.2 Exponential Smoothing

Exponential Smoothing model evaluation for HND and ND show that MAD for the period was 154.6 and 163.7, meaning that the error/deviation possible within those years was either 154.6 and 163.7 respectively more or less than the actual number of students admitted in those years than was eventually realized. Mean squared error achieved 40,410 and 32,305 for both HND and ND admissions. Mean absolute percent error was 26% and 17% with the accuracy for both categories of admission being 74% and 83% in-turn. From figure 4.5, it can be seen that Graph for the forecast for HND record analysis show a close trend flow with the actual record. In fact there was a convergence at year one which continued for some time towards year two, this proves a level of correlation at the early stage. Figure 4.6 does not show so much of that correlation even though there was a convergence at year one.

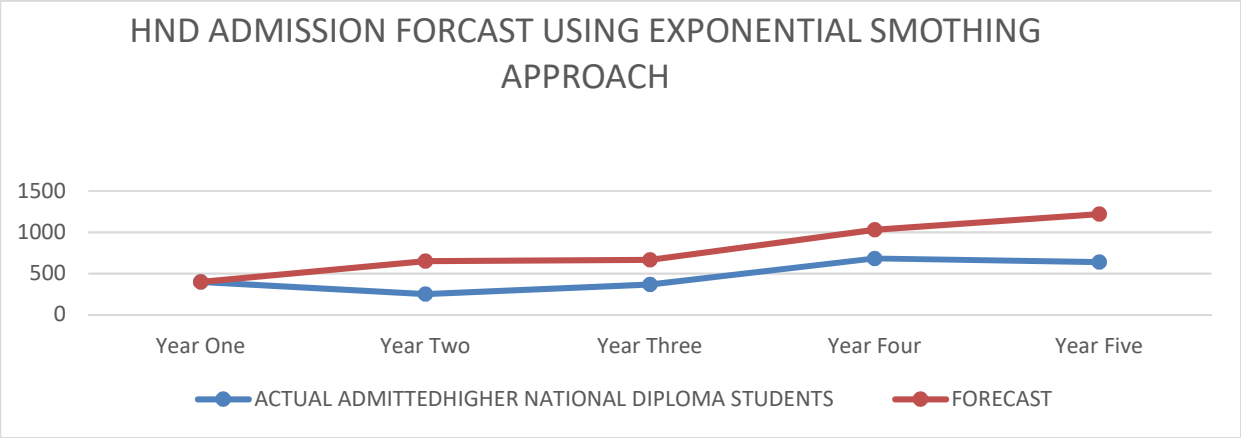


Figure: 4.5 HND exponential smoothing forecast

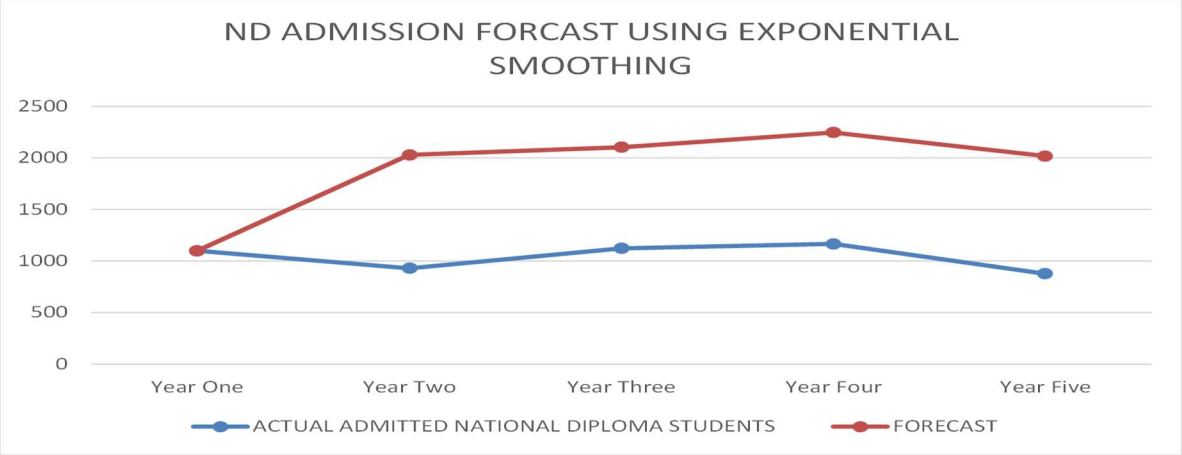


Figure: 4.6 ND exponential smoothing forecast

4.3 Two (2) -Year Moving Average (MA)

2 Years Moving Average model achieved a MAD of 175.8 and 171.8 for HND and ND respectively for the 5 year period. Mean squared error stood at 51,004 for HND and 34,263 for ND records. Mean absolute percent error is 28% and 17% respectively with an accuracy of 72% and 83% respectively. Figure 4.7 and figure 4.8 both have exact convergence trend between year one and year two, this must have accounted for the relatively high accuracy values. A wide divergence was observed at year 5 for the 2 set of records in both figure 4.7 and 4.8

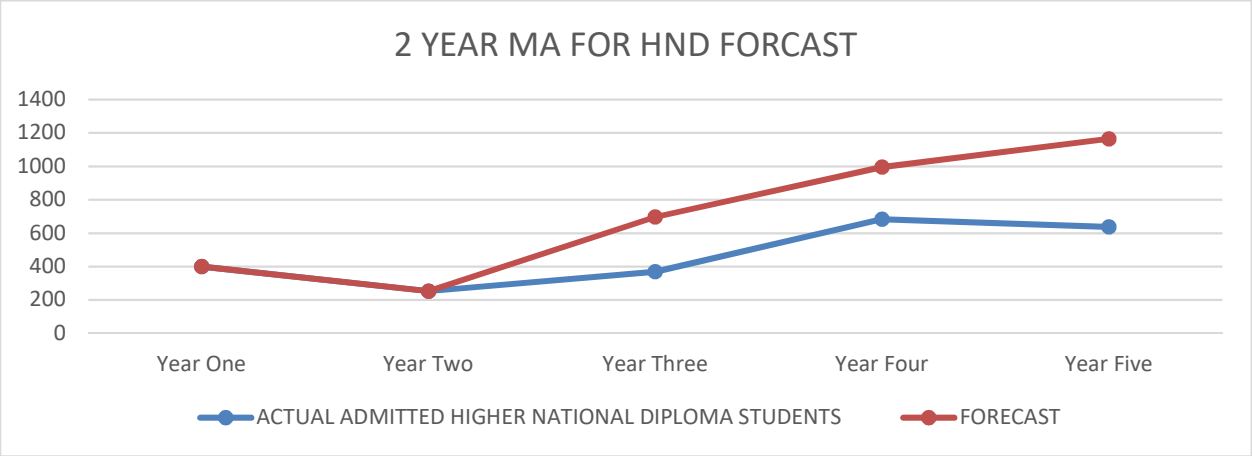


Figure: 4.7 HND 2-Year Moving Average forecast

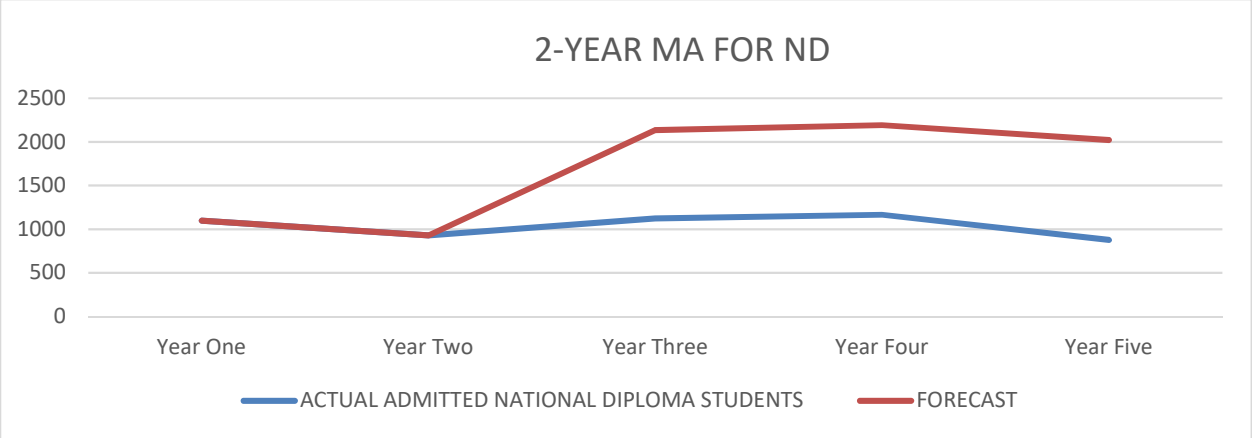


Figure 4.8: ND 2-year Moving Average (MA)

From Table 4.1 which represents HND record, Linear Regression (LR) has the highest accuracy of 77%, this implies that it is the most reliable forecasting model from among the evaluated models. Likewise, the MSE of 21,031 is the lowest obtained value which denotes it is the best model as the model with lowest MSE is adjudged the best model. This further corroborates the high accuracy value of the linear regression model as being reliable. Naïve approach is the worst performing model with accuracy value of 64%. Two year moving average model also performed poorly by having the highest MSE value of 51,004 from among the evaluated models. In Table 4.2 which summarizes the ND record, Linear Regression has both the highest accuracy and also the lowest MSE value, this confirms that LR is clearly the best model from the tested forecasting models. Naïve approach again proved to be the most unreliable model with the overall lowest accuracy of 82% and highest MSE of 37,795.

Table 4.1: Summary of HND Forecasting Analysis Result

Model Name	Mean Abosolute Error/Deviation (MAD/MAE)	Mean Square Error (MSE)	Mean Absolute Percent Error (MAPE)	Accuracy (%)
Naïve approach	155.75	33,929	36%	64
Exponential Smoothing	154.6	40,410	26%	74
Linear Regression	116.8	21,031	23%	77
Two-Years Moving Average(MA)	175.8	51,004	28%	72

Table 4.2: Summary of ND Forecasting Analysis Result

Model Name	Mean Abosolute Error/Deviation (MAD/MAE)	Mean Square Error (MSE)	Mean Absolute Percent Error (MAPE)	Accuracy (%)
Naïve approach	173.5	37,795	18%	82
Exponential Smoothing	163.761	32,305	17%	83
Linear Regression	24.21528393	812	2%	98
Two-Years Moving Average(MA)	171.833333	34,263	17%	83

5. CONCLUSION

We have attempted to forecast Naïve Approach, Linear Regression, 2-years Moving Average and Exponential Smoothing. The outcome of the evaluation confirms that Linear Regression is the best with both the HND and ND records with accuracy scores of 77% and 98% respectively, and so it is recommended to be adopted for forecasting future admission trends. Naïve approach and 2-years moving average proved to be unreliable for modelling of these records with MSE score of 37,795 and 51,004 in turn, they should therefore be avoided.

REFERENCES

1. Bernardita Calzon (2023): Data analysis methods and techniques, www.datapine.com (retrieved: 12th May, 2023)
2. Hyndman, R. J., and Athanasopoulos, G. (2018). Forecasting: principles and practice, Second edition, OTexts: Melbourne, Australia. OTexts.com/fpp2
3. Josephine, Malini M. Patil," A Review on Data Analytics for Supply Chain Management: A Case study", International Journal of Information Engineering and Electronic Business(IJIEEB), Vol.10, No.5, pp. 3039, 2018. DOI: 10.5815/ijieeb.2018.05.05 https://www.mecspress.org/ijieeb-v11-n1
4. Makridakis, S., Spiliotis, E., and Assimakopoulos, V. (2020). The M4 Competition:Results, findings, conclusion and way forward. International Journal of Forecasting, 36(1) 1-22. https://www.mecs-press.org/ijieeb-v11-n1
5. Rameshwar Garg, Shriya Barpanda, Girish Rao Salanke N S (2022): Machine Learning Algorithms for Time Series Analysis and Forecasting, DOI:10.48550/arXiv.2211.14387
6. Tim Stobierski (2021): It was originally published on August 26, 2019 Harvard Business School Online https://online.hbs.edu/blog/post/data-driven-decision-making (retrieved: 26th April, 2023)