Revenue Prediction and Donor Segmentation for NGOs

Abhishek Lalwani
Marathwada Mitra Mandal's
College of Engineering,
Pune, India

Harsh Gangawane
Marathwada Mitra Mandal's
College of Engineering,
Pune, India

Bhagyesh Hatwalne
Marathwada Mitra Mandal's
College of Engineering,
Pune, India

Rutuja Khire
Marathwada Mitra Mandal's College of Engineering,
Pune, India

Jitendra Chavan
Assistant Professor
Marathwada Mitra Mandal's College of Engineering,
Pune, India

ABSTRACT
Revenue Prediction and Donor Segmentation are vital to ensuring that any NGO has the right tools to promote itself in this digital era so that they can bring in more donations and have a better notion of what they might get, allowing them to serve more people. According to historical statistics, this data provides numerous insights on the kind of people who should be addressed, the target audience, and the predicted donations that may be expected in the coming months. These insights enable NGOs to improve their attempts to attract new donors. Because it expressly allows univariate time series data with a seasonal component, revenue prediction will be performed by utilizing the SARIMA (Seasonal Auto Regressive Integrated Moving Averages) Model on previous monthly arriving donations to estimate future month wise donations of the NGO. On the donor dataset, donor segmentation is accomplished by combining RFM (Recency, Frequency, and Monetary Value) Analysis with K-Means because RFM Analysis is a data-driven segmentation technique that allows the NGO to make tactical decisions, and the K-means clustering algorithm is used to find groups that have not been explicitly labelled in the data. Any additional data may be readily allocated to the correct group once the algorithm has been run and the groups have been formed. This document provides an overview of revenue forecasting and donor segmentation for non-profits, as well as a technique for doing so. It will assist NGOs in making well-informed decisions.

General Terms
Clustering, Time Series Algorithm, Prediction Analysis;

Keywords
NGOs, Machine Learning,SARIMA,RFM, Forecasting;

1. INTRODUCTION
Every year, the number of unfortunate people who rely on private groups such as NGOs to provide their fundamental necessities such as food, shelter, and education grows. According to recent surveys, over 20 corer Indians go to bed hungry every day, and over 8.8 lakh children under the age of five die of starvation. As a result, it becomes very expensive for any NGO to ensure that the quality of service that they provide now to an increased number of people remains the same or does not come at the expense of those in need. As a result, one of the most important aspects of an NGO's day-to-day operations is to keep accurate track of its incoming resources, potential targets to approach for additional funds and essential items, and an accurate estimate of next year's incoming to be able to assist more people in need in the coming years. This, in turn, aids them in meeting their stringent financial goals, keeping their budgets in check, and developing a marketing strategy to target a certain group of people who are more likely to donate to their cause based on their previous behavior.

Keeping track of an NGO's donors is a must for any organization. Food, books, and other basic supplies are donated instead of money by a variety of individuals. Some of them donate on a regular basis, while others just donate once in a while. The approach suggested first categorizes donors based on a variety of characteristics, then anticipate donation levels based on prior year data, allowing the NGO to better manage its overall revenue and plan its activities.

2. LITERATURE SURVEY
1. Yosua Alvin Adi Soetrisno et.al [1] Autoregressive Integrated Moving Average, or ARIMA, is one illustration of technique that generally used to figure cost or stock trade as univariate time-series information. Despite the fact that ARIMA could deal with information with pattern, it doesn't uphold time series examination for occasional merchandise like apple. SARIMA required for time series investigation, which is occasionally different. Apple cost forecast utilizing SARIMA could assist with checking the stock wellbeing level of apple, which is spoiled rapidly.

2. UZAIR ASLAM BHATTI et.al [2] Current advancement of Pakistan's economy, transportation and industry with the improvement of urbanization, natural contamination issues have step by step become noticeable, yet this is in opposition to individuals' vision of seeking after a great life. Presently the issue of cloudiness, photochemical issues in the air, and an Earth-wide temperature boost is as of now a central point of contention of worldwide concern. This is centered around the encompassing air nature of Lahore city of Pakistan. The review uncovers that the particulate matter in the Lahore season (PM2:5, PM10) surpasses Pakistan's National Environmental Quality Standards (NEQS).

3. Galina Goncharenko et.al [3] Advocacy non-legislative associations (NGOs) assume a significant part in the public eye by holding under tight restraints the force of enterprises and legislatures and uncovering privileges infringement. They contrast

from different NGOs as far as their plan, subsidizing structure and the partners they serve, and work in a setting portrayed by expanding requests for straightforwardness, responsibility and dependable promotion. This study looks at how the responsibility plan of promotion NGOs is formed by the need to keep up with autonomy, protect values and keep notoriety stainless when confronted with monetary and authenticity pressures.

4. Carlo Bertiet.al [4] The contemporary episode of traditional populism in mix with expanding transient streams toward Europe raises worries about the social development of relocations and travelers and the arrangement directions toward them. Research showed stressing propensities to see relocations adversely and condemn transients. Be that as it may, this paper centers around an alternate strategy took on by various libertarian powers: the criminalization of ocean salvage NGOs.

5. SeleshiSisaye et.al [5] GRI has given the establishment to coordinated detailing (IR). The two GRI and IR have natural and practical aspects. Manageability is practically inborn in the bookkeeping rule of materiality, when uncovered in outer announcing. The continuous worry of business accepts an association is fundamental and works as a living substance just when it can give economical execution that benefits partners and society.

6. YunusTurhan et.al [6] While Turkish religious NGOs have been associated with conveying philanthropic guide for quite some time, the relations of these associations with the state stand out. The primary reason for this article is to address this hole by asking which jobs NGOs play in Turkey's compassionate guide strategy and practice. In spite of the overall seclusion of Turkey's managerial construction because of a solid-state custom, this paper shows that NGOs endeavor to assume political parts in Turkish international strategy.

7. Ka Ho Mok et.al [7] in the previous ten years, the impact of worldwide powers for public area changes through the contribution of third area associations in friendly help conveyance through contracting out of friendly administrations has been developing. Understanding that the state alone can experience issues supporting and offering great quality social types of assistance, the Chinese government has made endeavors to connect with non-administrative associations to convey them. This article fundamentally inspects how the contracting out friendly help strategy took on in China has improved social assistance conveyance. This article additionally reflects upon the social administration changes in China and its consequences for the public authority NGOs relationship.

8. HE YU et.al [8] the upgraded ELM with GPS-EO-ABC, as well as the ARIMA, are used freely to created different anticipating results and joined by the weight-based parade. They affirm the presentation of the proposed superior ABC calculation by ten benchmark capacities, recreating the proposed gauging models by three monetary time-series datasets. That's what the outcomes demonstrate: (1) the proposed calculation shows remarkable limits in boundary improvement. The upgraded ELM produced more steady and exact outcomes contrasted and unique ELM, ABC-ELM, single LSTM, and ANN; (2) The proposed crossover model has viability as well as efficiencies in denoising information, amending exceptions, organizing both straight and nonlinear examples, its presentation in monetary time series gauging is more superb than existing half breed models.

9. Elvin Shava et.al [9] this article surveys the difficulties confronting both neighborhood and worldwide NGOs in Mwenezi locale of Zimbabwe in their endeavors to guarantee monetary manageability. Drawing from a subjective methodology that utilizes meetings and reports, the review uncovered that diminished subsidizing from the giver local area, changing needs of funders, unfortunate asset distribution and absence of straightforward constructions to draw in contributor subsidizing sway on the monetary manageability of NGOs. To guarantee their monetary supportability, NGOs should get rid of the giver reliance disorder and leave on other enterprising and vital raising money techniques.

10. Ashutosh Kumar Dubey et.al [10] Energy utilization estimating is fundamental for savvy network tasks as it works with power request the executives and utilities load arranging. In this paper information examination has been introduced on the gathered brilliant meter estimation and afterward foreseeing the energy utilization consistently utilizing (autoregressive coordinated moving normal) ARIMA, occasional ARIMA (SARIMA) and long momentary memory (LSTM). The investigation will in general comprehend the various elements which impact energy utilization, and help administrators to appropriately decide. It is useful in diminishing the blackout, and improving the situational attention to drive utilization consistently of the brilliant meters.

11. Yiqi Li et.al [11] NGOs are significant common entertainers in social orders' crisis and catastrophe reactions, and they meet up via web-based entertainment to distinguish conspicuous issues and direction issue reactions. This exploration investigates how U.S. NGO structure point driven networks via web-based entertainment to examine and fabricate authentic key organizations around issues connected with the COVID-19 pandemic. Drawing from Issue Niche Theory, they analyzed how NGOs' organizations and talk advanced when the overall population gave incredible consideration to the COVID-19 issue and how such examples changed across the entire issue specialty and sub-issue specialties. They dissected the development of Twitter-based organizations and talk of 2,588 NGOs in the initial five months of the COVID-19 episode in the United States. The examination uncovered significant elements that shape tie development designs in the NGOs' people group in this clever issue specialty. The discoveries show that NGOs' talks help to situate the authoritative local area to recognize most striking issues. At last, changes in the talk designs reflected changes in the correspondence networks in the NGO people group.

12. TulinDzhengiz et.al [12] Drawing on the writing on outlining, they investigate the enthusiastic outlining contrasts in revolutionary and reformative NGOs over the long run. They dissect the feeling of an

example of 5880 official statements gave by five NGOs situated diversely on the reformative extremist range and look at how they address enormous organizations. The discoveries uncover a rising polarization of opinion in these NGOs' outlining, with individual NGOs floating towards ideal-type revolutionary or reformative positions, separately. In arrangement with the distinctions in their outlining, they notice contrasts in their ways to deal with cross-area organizations. Policymakers need to take note of the ramifications of the noticed polarization for the viability and believability of cross-area organizations and multi-partner drives all the more by and large, given the gamble of co-optation (for reformative NGOs) as well as the gamble of previous critical financing and administration valuable open doors (for extremist NGOs).

3. PROBLEM STATEMENT

To develop a web-based application for an NGO to use Machine Learning to classify its donors and forecast revenue. The goal of this project is to create a platform that will allow NGOs to forecast revenue and categorize their contributors. To develop a software solution for predicting revenue using the SARIMA model. To provide a software system that uses RFM analysis to segment donors. To convert the data from the NGOs into a relational database for easy access and management.

4. PROPOSED SYSTEM

The system follows the architecture below, which describes a conceptual model that defines the structure, behavior, and other aspects of a system, such as a mapping of functionality to hardware and software components, a mapping of software architecture to hardware architecture, and human interaction with these components.

4.1 Dataset

From 2012 to 2021, the data set covers donations received by Advika Welfare Foundation throughout an eight-year period. The NGO has received donations ranging from Rs.50 to Rs. 25,000 in various amounts. The transactions indicate money received in a variety of ways, including cash, UPI, NEFT, and cheque.
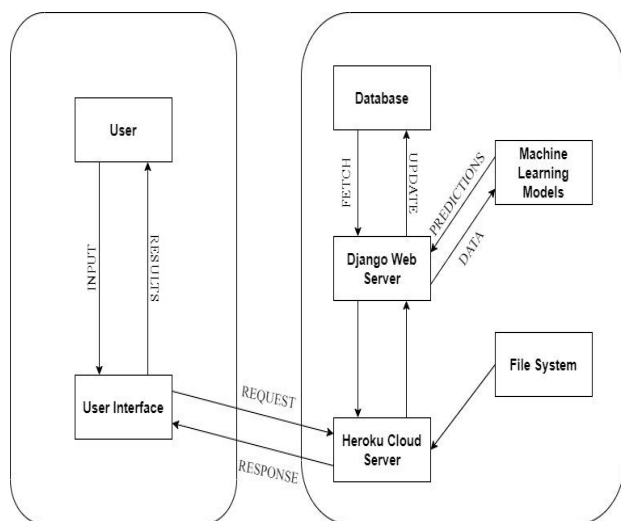

Fig.1:Architecture of proposed system

4.2 Pre-processing for Revenue Prediction

Any type of processing conducted on raw data to prepare it for another processing technique is referred to as data preparation. Data preprocessing, which is commonly used as a preliminary data mining strategy, converts data into a format that can be processed more readily and effectively for the user's purpose. Preprocessing is accomplished using a variety of tools and methods, such as sampling, which chooses a representative subset from a vast population of data transformation, which manipulates raw data to produce a single input denoising, which removes noise from data.

The initial preparation operation was to convert the month column to date time format after importing the dataset for preprocessing. Augmented Dickey Fuller Test (ADF Test) was used to check if the time series is stationary after type conversion. The qualities of a stationary time series are independent of the time at which it is viewed.

The time series is non-stationary, according to the ADF Test results. Then, to make the time series stationary, use the seasonal difference of the donation's column. To confirm whether the series is stationary after taking the seasonal difference, ADF test was used.

The seasonal component of the model for the Auto Regressive parameters, differences, and Moving Averages parameters is then determined by plotting the auto correlation and partial auto correlation.

4.3 Pre-processing for Segmentation

The RFM segmentation technique is based on the premise that by assessing three quantitative criteria, marketers may acquire a comprehensive picture of their donors. These are the following:

1. Recency:
When was the last time a donor engaged in an activity or made a transaction with the brand? The most common activity is a purchase, while other examples include the most recent visit to a website or the use of a mobile app. The more recently a donor has connected or transacted with a brand, the more likely the donor will be responsive to brand messages.

2. Frequency:
During a given time period, how many times has a donor transacted or interacted with the brand? Donors who participate in events on a regular basis are clearly more involved and perhaps more loyal than those who do so infrequently. One-time-only contributors are in an other league altogether.

3. Monetary:
This component, often known as "monetary value," represents how much a contributor has spent with the company over a period of time. Donors who spend a lot of money should be rewarded differently than those who give a little. The average buy amount is calculated by dividing monetary by frequency, which is a significant secondary statistic to consider when segmenting donors.

4.4 SARIMA Algorithm

An Auto Regressive Integrated Moving Average (ARIMA) model is a speculation of an Auto Regressive Moving Average (ARMA) model in measurements and econometrics, especially in time series investigation. For univariate time series information anticipating, ARIMA is quite possibly the most broadly utilized estimating calculation. Whenever information shows proof of non-fixed in the feeling of mean, an underlying differencing step can be applied at least multiple times to dispose of the non-fixed of the mean capacity, ARIMA models

are utilized (i.e., the pattern). Albeit the strategy can deal with information with a pattern, time series with an occasional part are not upheld.

Therefore, when irregularity shows up in a period series, occasional differencing can be utilized to eliminate the occasional part. Therefore, the Seasonal Auto Regressive Integrated Moving Average was created (SARIMA).

SARIMA is an ARIMA expansion that takes into account direct demonstrating of the series' occasional part. SARIMA (Seasonal ARIMA) is an alteration of ARIMA that expressly permits univariate time series information with an occasional part. It adds three new hyper boundaries for the occasional part of the series, including auto relapse (AR), differencing (I), and moving normal (MA), as well as an extra boundary for the irregularity time frame.

The execution is alluded to as SARIMAX instead of SARIMA since the "X" added to the technique name demonstrates that it likewise permits exogenous factors. These are equal time series varieties that aren't expressly displayed utilizing Auto Regressive, Integrated or Moving Average cycles, yet are made accessible to the model as a weighted info.

4.5 RFM analysis with K-means Clustering

To target specific groups of contributors with communications that are far more relevant to their individual habits, resulting in significantly higher response rates, as well as increased loyalty and donor lifetime value. RFM segmentation, like other segmentation methods, is a powerful tool for identifying donor groups that should be treated differently. Recency, frequency, and monetary are abbreviated as RFM. Marketers frequently have a wealth of data about their existing customers, such as purchase history, browsing history, prior campaign response patterns, and demographics, which can be leveraged to identify specific groups of contributors who can be targeted with highly relevant offers.

The K-means algorithm is an iterative method for dividing a dataset into K separate clusters. It attempts to make intra-cluster data points as comparable as possible while maintaining clusters as far apart as possible. It assigns data points to a cluster so that the sum of their squared distances from the cluster's centroid is as small as possible. Within clusters, the less variance there is, the more uniform the data points are. K-means clustering is a vector quantization technique originating in signal processing that seeks to divide n observations into k groups, with each observation belonging to the cluster with the closest mean. Regular Euclidean distances, on the other hand, are not. Both k-means and Gaussian mixture modeling use an iterative refinement technique, which is comparable to the expectation-maximization algorithm for mixtures of Gaussian distributions. To model the data, they both use cluster centers.

Then assign Recency, Frequency, and Monetary values to each contributor while developing the RFM model.

The period of time since the donor's most recent transaction is referred to as recency (most businesses use days, though for others it might make sense to use months, weeks or even hours instead).

The total number of transactions done by the donor is known as frequency (during a defined period).

The entire amount spent by the contributor across all transactions is referred to as monetary (during a defined period).

For each of the three dimensions, the donors were then sorted into tiered groups (Recency, Frequency and Monetary).

Finally, K-Means Clustering was used to construct donor clusters to whom certain types of messages will be sent.

5. RESULTS AND DISCUSSION
The Home Page of the NGO Admin Portal contains information about NGOs and a navigation bar with different functionalities like Login, Statistics, Revenue Prediction, Donor Segmentation, Payment Receipt and Logout.
a revenue forecast is an educated prediction or estimation for the upcoming year about how much money the NGO is likely to bring in.
Clustering is the process of dividing a population or set of data points into multiple groups so that data points in the same group are more similar than data points in other groups.

5.1 Result I: Revenue Prediction
The graph below shows the predicted donation amounts for the year 2022 as a result of the SARIMA algorithm.

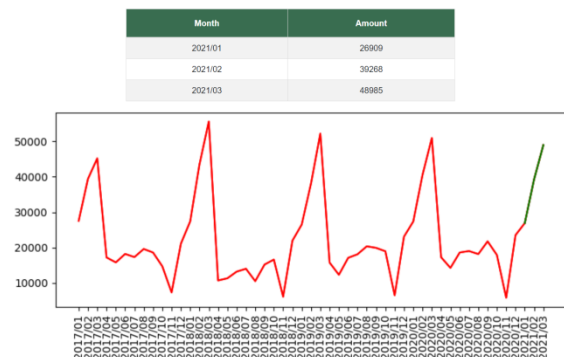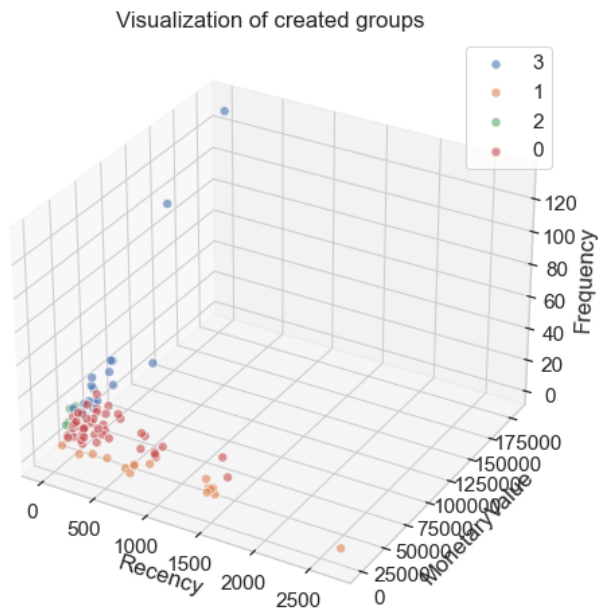The RMSE value of the algorithm is: 2.004

Month	Amount
2021/01	26909
2021/02	30268
2021/03	48985

Fig.2: SARIMA Prediction Graph

5.3 Result II: RFM Graph
The graph below shows the clusters of donors for the as a result of the K-Means Clustering algorithm using RFM analysis.

The silhouette score of the algorithm is: 0.502

Fig.3: RFM Cluster Graph

6. CONCLUSION

As a result, the developed software can anticipate revenue using SARIMA (Seasonal Auto Regressive Integrated Moving Averages). Using RFM, we've additionally segregated the NGOs' contributors (Recency Frequency and Monetary values). This ensures that NGOs have the best tools to advertise themselves in this digital age, allowing them to bring in more donations and have a better sense of how much money they will receive,allowing them to serve more people. This information provides various insights about the kind of people who should be approached, the target audience, and the expected donations in the coming months based on historical data.

The future scope includes adapting the system to include multiple NGO entries. Social media analytics can be integrated to reach to a larger audience using publicity campaigns targeting specific clusters. Also include KPIs for NGOs and use it to rate the NGOs and corelate it with the donor satisfaction and loyalty.

7. ACKNOWLEDGEMENTS

8. REFERENCES

[1] Chunli Huang, Wenjun Jiang, Jie Wu, and Guojun Wang. 2020. Personalized Review Recommendation based on Users' Aspect Sentiment. ACM Trans. Internet Technol. 20, 4, Article 42 (November 2020), 26 pages.DOI:https://doi.org/10.1145/3414841

[2] Slava Novgorodov, Ido Guy, Guy Elad, and Kira Radinsky. 2020. Descriptions from the Customers: Comparative Analysis of Review-based Product Description Generation Methods. ACM Trans. Internet Technol. 20, 4, Article 44 (November 2020), 31 pages. DOI:https://doi.org/10.1145/3418202

[3] Galina Goncharenko." The accountability of advocacy NGOs: insights from the online community of practice" ACCOUNTING FORUM 2019, VOL. 43, NO. 1, 135–

[4] SeleshiSisaye." The influence of non-governmental organizations (NGOs) on the development of voluntary sustainability accounting reporting rules". Received 13 February 2021 Revised 13 March 2021, 14 March 2021, 15 March 2021, Accepted 15 March 2021.

[5] YunusTurhan&ŞerifOnurBahçecik." The agency of faith-based NGOs in Turkish humanitarian aid policy and practice". Received 28 November 2019; Accepted 12 March 2020.

[6] Ka Ho Mok, Chak Kwan Chan and Zhouyi Wen." Searching for new welfare governance in China: contracting out social service and impact on government-NGOs relationship". Received 19 May 2020, Accepted 20 May 2020

[7] HE YU 1, LI JING MING2, RUAN SUMEI2, AND ZHAO SHUPING3." A Hybrid Model for Financial Time Series Forecasting Integration of EWT, ARIMA with the Improved ABC Optimized ELM". Received April 1, 2020, accepted April 5, 2020, date of publication April 13, 2020, date of current version May 18, 2020.

[8] ZHAO SHUPING3." A Hybrid Model for Financial Time Series Forecasting Integration of EWT, ARIMA with the Improved ABC Optimized ELM". Received April 1, 2020, accepted April 5, 2020, date of publication April 13, 2020, date of current version May 18, 2020.

[9] Elvin Shava." Financial sustainability of NGOs in rural development programmers". DEVELOPMENT IN PRACTICE 2021, VOL. 31, NO. 3, 393–403

[10] Ashutosh Kumar Dubey a, Abhishek Kumar a, Vicente García-Díaz b,*, Arpit Kumar Sharma c, Kishan Kanhaiya." Study and analysis of SARIMA and LSTM in forecasting time series data". Sustainable Energy Technologies and Assessments 47 (2021) 101474.

[11] Yiqi Li a,*, Jieun Shin b, Jingyi Sun a, Hye Min Kim a, Yan Qu c, Aimei Yang." Organizational sensemaking in tough times: The ecology of NGOs' COVID-19 issue discourse communities on social media". Computers in Human Behavior 122 (2021) 106838.

[12] Tulin Dzhengiz1 | Ralf Barkemeyer2 | Giulio Napolitano." Emotional framing of NGO press releases: Reformative versus radical NGOs". Received: 7 August 2020 Revised: 25 January 2021 Accepted: 12 February 2021