Proceedings of the 2019 IISE Annual Conference
H.E. Romeijn, A Schaefer, R. Thomas, eds.

Predicting Food Donor Contribution Behavior Using Support Vector Regression

Shubhra Paul, Lauren B. Davis
Department of Industrial and Systems Engineering
North Carolina A&T State University, Greensboro, NC, 27411, USA.

Abstract

Hunger and food insecurity are present in each American county. Government and non-government organizations are working to address food insecurity in the United States. Food banks are nonprofit hunger relief organizations that collect food and monetary donations from donors and distribute food to local agencies which serve people in need. Contributions come from retail donors, communities, and food manufacturers. The uncertainty of donation amounts and frequency is a challenge for food banks in the fight against hunger. In this research, we analyze local food bank donation data and propose a predictive model to forecast the contribution of different donors. Our study shows the necessary behavioral attributes to classify donors and the best way to cluster donor data to improve the prediction model. We also compare the accuracy of prediction for different conventional forecasting techniques with the proposed Support Vector Regression (SVR) model.

Keywords: Food Insecurity, Support Vector Regression, Forecasting, Donation Behavior, Humanitarian Supply Chain.

1. Introduction

1.1 Background

A household with limited or uncertain availability of food as a result of insufficient financial resources is food insecure [1]. In the United States, Food insecurity affects one out of eight Americans, which corresponds to 12 million children and 40 million households [2]. The Red Cross and Red Crescent Society define the food insecurity and hunger problem as a complex human-made hazard [3,4]. There are several federally funded programs for fighting against hunger such as Women Infants and Children (WIC) and Supplemental Nutrition Assistance Program (SNAP). Besides those programs, and there is a network of public and private organizations such as Feeding America (FA), independent food banks, soup kitchen. Under FA networks there are 200 food banks, and 60,000 food pantries operating [5].

In North Caroline (NC), over one million people face food insecurity, which is approximately 15.6% of the total NC population [6]. In addition, 479,220 children struggle with hunger with Mecklenburg County having the highest food insecure children (45,370 children), and Scotland County with the highest percentage (30.6 %) [7]. There are 7 Feeding America (FA) food banks that serve NC. Food Bank and Central of Eastern North Carolina is the largest among them [6]. Some food banks serve more than one county.

1.2 Related Literature

Forecasting of demand is extensively explored in the commercial and humanitarian supply chain literature using a variety of techniques. Prediction of supply, mainly food donations in the humanitarian sector is limited. Artificial Neural Network and Multiple Linear Regression were proposed to predict donations received by a food bank from the supermarket [8]. Data clustering and time series methods were used to predict food donation amounts based on the donor, location, and food type attributes; exponential smoothing, moving average and autoregressive integrated moving average (ARIMA) approaches were considered[9]. The authors also discuss the effect of data clustering on forecast accuracy and variability. Food donations have been considered in other contexts of food operations, beyond supply forecasting. For example, a mathematical model was introduced to integrate the effective use of donated food and optimized menu planning for the soup kitchen [11]. Within the context of humanitarian food relief activities, demand forecasting has been studied to identify factors that contribute to estimating food need for a local food bank. The authors proposed a regression model to predict future contributions, and evaluated model accuracy using the Mean Absolute Deviation (MAD) and Mean Absolute Percentage Error (MAPE) [10].

Support Vector Machine (SVM) is a machine learning approach for solving the define classification problem [12]. SVM can also solve the quantity prediction problem via Support Vector Regression (SVR) [13]. Support Vector Regression (SVR) is the supervised machine learning technique. It can learn data patterns from the training data set and can apply the pattern to validation data set. Authors in [14] were the first to use SVR to forecast in kind aggregate donations for the food bank. Results of their showed that a model based on 24 months past data provided the best result for training data.

Table 1: FA food bank overview in NC [6]

Food Bank	No of Counties Served	Meals Distributed each year
Food Bank of Central of Eastern NC	34	53,293,730
Second Harvest Food Bank of Metrolina	19 (14 in NC and 5 in SC)	39,041,689
Second Harvest Food Bank of North West NC	18	30,340,545
Manna Food Bank	16	12,624,347
Second Harvest Food Bank of Southeast NC	7	7,897,622
Food Bank of the Albemarle	15	4,980,418
Inter-Faith Food Shuttle	7	4,654,195

1.3 Research Objectives

In-kind donations are the primary supply for food banks. Both demand and supply are uncertain for the hunger-relief humanitarian supply chain. In this research, we have focused on supply prediction problem. To our best knowledge, this is the first work which considered the contribution behavior of different donors for SVR. The prior SVR work considered contributions by location [14]. The research questions driving this study are as follow.

1) How well does Support Vector Regression (SVR) perform over different time series model in predicting food donation based on the donor?
2) What is the best way to cluster donor data to improve the prediction model?
3) What are the necessary behavioral attributes to classify donors?

The remainder of the chapter is organized as follow; section 2 describes data collection, data analysis, section 3 discusses the results and comparison of models and section 4 concludes the chapter.

2. Method

2.1 Data Collection

We collected historical data from a local food bank for three the fiscal year 2015-2016 to 2017-2018. The data contained total 853,908 records and 109 variables. The data was filtered by donations, and all distribution and purchase records removed which reduced the observations to 105,433. Information regarding original and cleaned data are shown in Table 2. Table 3 contains the key donation fields and example values used in the study. In addition to the historical transaction data, a detailed list of donor information was obtained. Variables in this second list contain donor ID, name and location (address, city, and state).

Table 2: Information on original and cleaned data

Fiscal Year	2015 – 16	2016 - 17	2017 – 18
Variables	109	109	109
Observations (Original)	272,304	287,840	293,764
Observations (Cleaned)	49,511	50,777	50,845

Table 3: Descriptions of key variables in the dataset

Key Variables	Example Values	Key Variables	Example Values
Posting Date	07-01-2016	Donor ID	1529
Receiving Location	Durham, Raleigh	Product Type	Meats
Donor Affiliation	Local	Donor Trade Classification	Retail
Storage Classification	Dry	Gross Weight (lbs.)	100,000

2.2 Data Preprocessing:

Before analysis, many issues were founded in the dataset. For example, there was a high positive donation of a donor and negative donation of the same donor on the same posting date. This happens when contribution amounts need to be adjusted. For analysis, data were summarized by monthly total gross weight, so there is no error in frequency analysis. There was also evidence of missing data and potential data error during donations entry. Missing values had a significant impact on frequency analysis and future predictions. Therefore, the data needed to be preprocessed before analysis. For analysis, data were preprocessed by software "R." Missing values, outliers, significant data of donation were fixed. Unnecessary variables were filtered, removed and added together to get the final donation data set.

2.2.1 Individual Donor Behavior:

During preprocessing data set, an observation was found that reflects individual donor contributions. There was one donor id that aggregated in-kind donations made by individuals and therefore, detail information regarding the address, city; and the state are missing in the donor list. The contribution of all individual donors is shown in Figure1. The most substantial donations occur from October to December. The reasons for the high amount of donations during these months are attributed to the North Carolina State Fair, and holiday donations (Thanksgiving and Christmas).

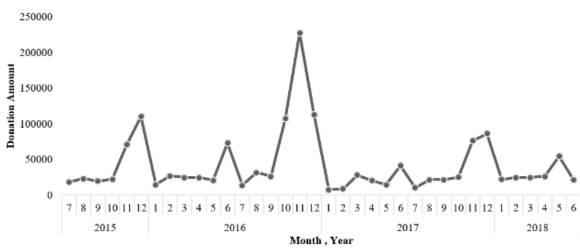

Figure 1: Individual donor's donation pattern

Statistical analysis was done on individual donor's donation data to determine the show minimum, mean, maximum and standard deviation of donation as 7,344 lbs., 41,669.67 lbs., 227,418 lbs., and 43,418.31 lbs. respectively. The variability of data has a coefficient of variation which is more significant than 1, and therefore considered high. Figure 2 summarizes the contribution of individual donors to total donations for each fiscal year. The fiscal year contribution of individual donors is less than 1%. Due to low participation and missing addresses, all individual donor contributions were removed from the donation data set.

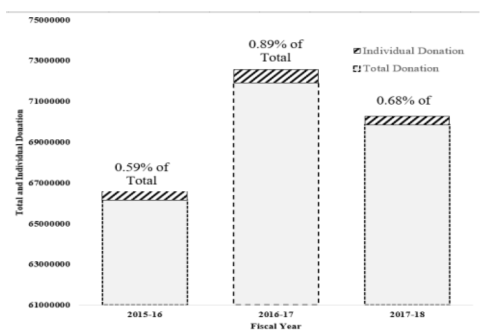

Figure 2: Individual donors' contribution

2.2.2 Donor Affiliation Category Data Analysis:

Donor Affiliation categorizes donors based on affiliation with Feeding America, State Government, Federal Government, and Agencies. Description of donor affiliation category shown in Table 4. The gross weight of the donation amount for each possible *Donor Affiliation* was plotted and shown in Figure 3. Most of the donations come from FA Local donors and Local donors. Total donation from FA Local and Local donation is 88.37 %, 76.50% and 81.30 % for the Fiscal Years 2015-16, 2016-17 and 2016-18 respectively.

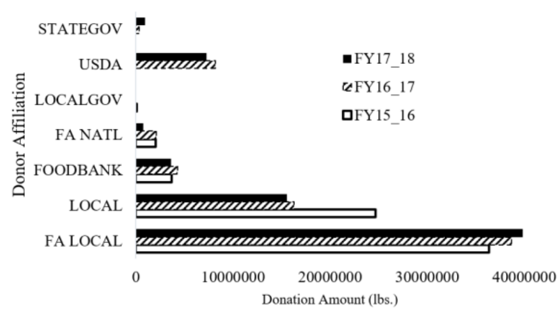

Figure 3: Donation amount vs. Donor Affiliation

The research focus was to find different donors contribution behaviors. Therefore data related to purchased food (FA Purchase data) FA National, Federal and State donation data was filtered from the donation data set. The Pareto analysis also shows that the law of the vital few are widely used to find approximately 80% contributions comes from 20% population. The Pareto study was done to see the percentage of total contributions of FA Local and Local donors which shows that most donations came from those two. The Pareto analysis result is shown in Table 5.

Table 4: Descriptions of key values for the Donor Affiliation in the dataset

Donor Affiliation Category	Description
FA Local	Donors have an affiliation with FA, but there may be a local location.
FA National	Donors have a relationship with FA the National Network and food banks need to send trucks to collects food.
Local	Donor within the Food Bank's service area with local ties, not a national company.
Local Govt.	Food obtained from the local department of agriculture.
State Govt.	Food Obtained through funds allocated by the state.
USDA	Food obtained through the government commodity program.

Table 5: Pareto Analysis results

Fiscal Year	Total contributions of FA Local and Local Donors
2015 – 16	88.37 %
2016 – 17	76.50 %
2017 – 18	81.30 %

2.2.3 Trend Analysis:

Figure 4 display the overall donation amount (in gross weight) for the last three fiscal year dataset. There is no increasing or decreasing trend of total donation amounts in the last three years' data. Minimum donations were 4,580,745 lbs. in September 2015, and maximum contributions were 7,442,772 lbs. in October 2016.

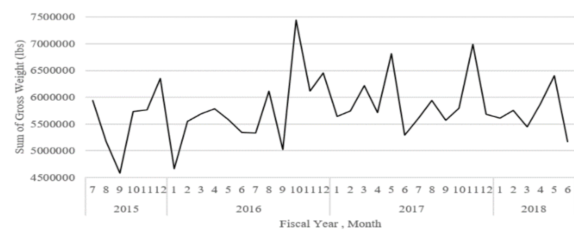

Figure 4: Trend analysis of donation amounts

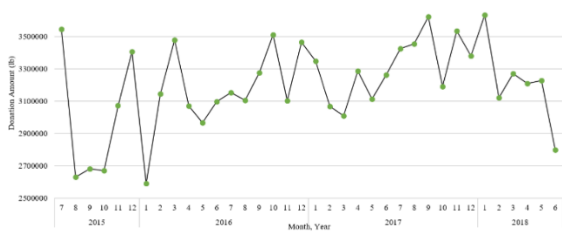

Figure 5: Trend analysis of FA Local donations

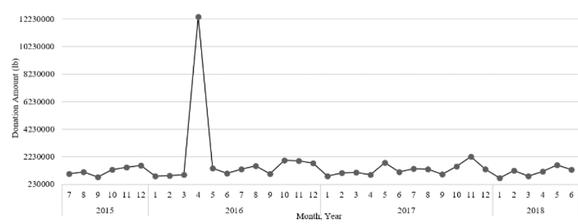

Figure 6: Trend analysis of Local Donor donations

During frequency analysis, we found that there were 757, 884 and 870 different donors that donated 49510 times, 50776 times and 50844 times during the fiscal year 2015 - 16, 2016 - 17 and 2017 -18 respectively. The number of donations and donors during these years are relatively the same. Figure 5 provides a snapshot of the trend analysis of monthly FA Local donors donation amount in lbs. Minimum, Maximum, Mean and Standard Deviation for FA Local donation amounts are 2591091 lbs., 3632182 lbs., 3191776.31 lbs., and 276148.24 lbs. There is no increasing or decreasing trend found in figure 5. Figure 6 shows less variability for local donations. The mean value is 1571262.17 lbs., and the standard deviation is 1892268.08 lbs.

2.2.4 Donation Frequency Analysis:

There are 258 Local donors and 512 FA Local donors during the last three fiscal years. From the frequency analysis, we observe that donations from FA local donors are more frequent. One hundred eleven donors donated once in three fiscal years among 258 donors whereas such value for FA local is only 14. 40.23 % of FA Local Donor donated every month whereas only 6.20% Local donor donated every month. Table 6 shows the comparison of donation frequency analysis of FA Local and Local donor.

3. Result and Discussion:

3.1 Moving Average:

Moving Average is a popular forecasting method to predict future values. Moving Average was done based on the last two months' average, three months' average, four months' average, five months' average, and six months' average. Three types of error – Mean Absolute Deviation (MAD), Mean Square Error (MSE) and Mean Absolute Percentage Error (MAPE) were calculated, and the results are shown in table 6. Two-month moving average provided the best prediction (i.e., smallest forecast error).

Table 6: Moving Average Error comparison concerning months

	Local		FA Local	
Donation Frequency	No of Donor	Percentage of donor	No of Donor	Percentage of donor
36	16	6.20%	206	40.23%
30-35	18	6.98%	120	23.44%
20-29	15	5.81%	23	4.49%
10-19	23	8.91%	35	6.84%
1-9	186	72.09%	68	13.28%

Table 7: Moving Average Error comparison concerning months

Error	2 Months	3 Months	4 Months	5 Months	6 Months
MAD	542287.86	725555.40	812566.12	865353.80	907216.69
MSE	1.84E+12	2.64E+12	3.12E+12	3.26E+12	3.49E+12
MAPE	9.80%	13.09%	14.17%	14.88%	15.53%

3.2 Support Vector Regression Model:

Support Vector Regression Model was used to predict donation amounts based on the type of donor. Support Vector Regression was done based on a one-month lag. Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) for SVR model for pre-tuned validation data are 456034.90 and 5.80% respectively. MAPE and RMSE comparison for SVR and two months moving average are shown in Figure 7 and Figure 8 respectively. SVR model provided a better result than MA.

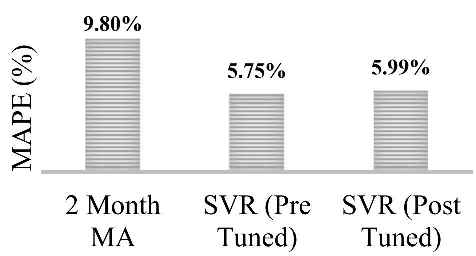

Figure 7: MAPE comparison for MA and SVR Model.

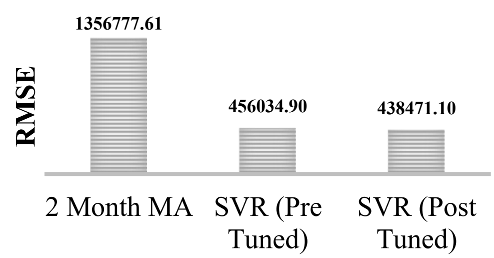

Figure 8: RMSE comparison for MA and SVR Model

3.3 Comparison between FA Local Donation, Local Donation and Cluster Donations prediction model:

We used SVR on FA Local Data, Local Data and Cluster of both data set. The comparison of MAPE is shown in Figure 9. Local data has the highest MAPE 27.53% and 24.91% for pre-tune and post tune validation data respectively, whereas cluster data provide less than 6% in both cases. MAPE for FA Local is 7.05 % for pre-tune data and 7.77% for post tune data.

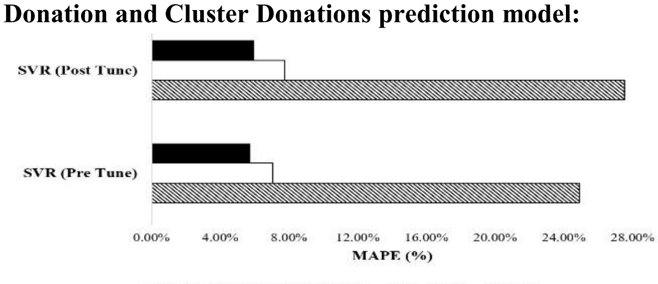

Figure 9: SVR MAPE comparison for Local, FA Local and Cluster Data

4. Conclusion:

The prediction accuracy for SVR is better than the Moving Average method as shown in Figures 7 and 8. With high standard deviation data set SVR also worked better. Future research will explore different predictive models such as Auto-Regressive Integrated Moving Average (ARIMA), Logistics Regression. We will also explore other variables for clustering data such as the type of food and expiration date. From this research, we conclude that donation frequency is a vital behavioral attribute to classify donor. To improve the prediction model, donor affiliation is a right way for clustering.

Acknowledgment:

- The authors would like to acknowledge the food bank of Central and Eastern North Carolina (FBCENC) for sharing data and cooperating for this research.
- This project was supported by NSF National Research Traineeship Project Improving Strategies for Hunger Relief and Food Security using Computational Data Science (Award No. DGE-1735258) and NSF Partnerships for Innovation Project Flexible, Equitable, Efficient, and Effective distribution (FEEED) (Award No. IIP - 1718672)

References

1. United States Department of Agriculture, Economic Research Service. Available at https://www.ers.usda.gov/topics/food-nutrition-assistance/food-security-in-the-us/definitions-of-food-security.aspx
2. Coleman-Jensen, A., Nord, M., Andrews, M., & Carlson, S. (2016). Statistical Supplement to Household Food Security in the United States in 2011. United States, Department of Agriculture Economic, Research Service.
3. International Federation of Red Cross and Red Crescent Societies, Complex/manmade hazards: complex emergencies. Available at: https://www.ifrc.org/en/what-we-do/disaster-management/about-disasters/definition-of-hazard/complex-emergencies/
4. The Magazine of the international Red Cross and red crescent, Food Security – a paradigm shift. Available at http://www.redcross.int/EN/mag/magazine2006_2/4-9.html
5. Feeding America, Delivering food and services. Available at https://www.feedingamerica.org/our-work/food-bank-network
6. Feeding America, Hunger in North Carolina. Available at https://www.feedingamerica.org/hunger-in-america/north-carolina
7. Feeding America, Child Food Insecurity in NC by county. Available at https://www.feedingamerica.org/sites/default/files/research/map-the-meal-gap/2016/child/NC_AllCounties_CDs_CFI_2016.pdf
8. Brock III, L. G., & Davis, L. B. (2015). Estimating available supermarket commodities for food bank collection in the absence of information. Expert Systems with Applications, 42(7), 3450-3461.
9. Davis, L. B., Jiang, S. X., Morgan, S. D., Nuamah, I. A., & Terry, J. R. (2016). Analysis and prediction of food donation behavior for a domestic hunger relief organization. International Journal of Production Economics, 182, 26-37.
10. Okore-Hanson, A., Winbush, H., Davis, L., & Jiang, S. (2012, January). Empirical modeling of demand for a local food bank. In IIE Annual Conference. Proceedings (p. 1). Institute of Industrial and Systems Engineers (IISE).
11. Buisman, M. E., Haijema, R., Akkerman, R., & Bloemhof, J. M. (2019). Donation management for menu planning at soup kitchens. European Journal of Operational Research, 272(1), 324-338.
12. Noble, W. S. (2006). What is a support vector machine?. Nature biotechnology, 24(12), 1565.
13. Vapnik, V. (2013). The nature of statistical learning theory. Springer science & business media.
14. Pugh, N., & Davis, L. B. (2017, December). Forecast and analysis of food donations using support vector regression. In 2017 IEEE International Conference on Big Data (Big Data) (pp. 3261-3267). IEEE.