Annals of Operations Research
https://doi.org/10.1007/s10479-021-04146-5

S.I.: DESIGN AND MANAGEMENT OF HUMANITARIAN SUPPLY CHAINS

An ensemble forecasting model for predicting contribution of food donors based on supply behavior

Shubhra Paul[1] · Lauren B. Davis[1]

Accepted: 29 May 2021

Abstract
Food banks are nonprofit hunger relief organizations that collect donations from donors and distribute food to local agencies that serve people in need. Donors consist of local supermarkets, manufacturers, and community organizations. The frequency, quantity, and type of food donated by each donor can vary each month. In this research, we propose a technique to identify the supply behavior of donors and cluster them based on these attributes. We then develop a predictive ensemble model to forecast the contribution of different donor clusters. Our study shows the necessary behavioral attributes to classify donors and the best way to cluster donor data to improve the prediction model.

Keywords Food insecurity · Forecasting · Humanitarian supply chain · Ensemble model · Clustering · ARIMA · Support vector regression

1 Introduction

A household with limited or uncertain availability of food as a result of insufficient financial resources is food insecure (USDA Economic Research Service 2020a). The Red Cross and Red Crescent Society define the food insecurity and hunger problem as a complex human-made hazard (International Federation of Red Cross and Red Crescent Societies (IFRC) n.d.). Globally, approximately 782 million people face hunger (Thome et al. 2018). Within the US, 40 million households and nearly 12 million children are food insecure (USDA Economic Research Service 2020b). Rates of food insecurity differ geographically as well as among states. For example, the states in the South have higher food insecurity rates (13.4%) than those in the Northeast (9.9%) (USDA Economic Research Service 2020b). One southern state which our study focuses on (North Carolina (NC)) has a food insecurity rate (15.6%) above the national average (11.5%) as of 2018 (Feeding America n.d.). The prevalence of food insecurity has prompted the creation of several federal, public, and private efforts to

✉ Lauren B. Davis
lbdavis@ncat.edu

[1] Department of Industrial and Systems Engineering, North Carolina Agricultural and Technical State University, Greensboro, NC, USA

mitigate the effects of hunger. In the US, examples of federally funded programs targeted for fighting against hunger are Women, Infants, and Children (WIC), and Supplemental Nutrition Assistance Program (SNAP) (Davis et al. 2016). There are also a network of public and private organizations assisting in hunger relief such as, Feeding America, independent food banks, and soup kitchens. Feeding America is a national hunger relief organization consisting of a network of 200 food banks and 60,000 food pantries (Davis et al. 2016). There are 7 Feeding America food banks that serve NC of which The Food Bank and Central of Eastern North Carolina (FBCENC) is the largest among them, providing food to hungry people in 34 counties for the last 35 years (Feeding America n.d.).

The primary role of food banks is to collect donations from small and large-scale donors, provide storage capability for the collected food items and subsequently distribute them to the food insecure population through local agencies (Davis et al. 2016). Food bank operations depend on both food and monetary donations from various sources. Monetary donations are used to purchase food items (Morello 2020). Purchased food amounts can vary between 10 and 62% of distributed quantity depending on different food banks (Chattanooga Area Food Bank n.d.; Everitt 2012; Chester County Food Bank n.d.). A significant portion of donations is food which comes from food drives, farmers, food manufactures, retail donors, and the government. The frequency and amount of donated foods are uncertain since donation patterns vary among donors. For example, local farmers may donate more fresh food during harvest time, whereas individual donations from the community are more prevalent during the holiday season. Retail donors may provide more food that are close to its expiration date. Some donors donate dry foods, and others donate perishable food. Understanding donation behavior is complex and significantly influences the supply available to distribute food efficiently, effectively and equitably. Since donated food is a significant supply source, this study aims to address the following research questions:

1. What are the key attributes that describe donor behavior ?
2. What is the best way to cluster donors based on their behavioral attributes?
3. What is the best way to predict food donations considering the donor behavioral attributes?

The study aims to address these questions by defining donor behavioral attributes and clustering them to predict food donations. This research extends the notion of supplier behavior in the for-profit supply chain to characterize donor behavior in a non-profit supply chain. Specifically, we propose seven metrics to characterize the behavior along the dimensions of quantity, quality, reliability, affiliation, service, and donation product variety. We use the metrics to construct profiles of supplier behavior and subsequently estimate in-kind food donations for donors clustered based on their supplier behavior. From our results, we are able to show which behaviors lead to better predictions of in-kind contributions. We are also able to show how prediction accuracy can be increased using an ensemble forecasting approach.

The results of our study can inform sourcing decisions of food bank decision-makers through improved identification of the donors that provide certain types of food, their contribution frequency and quality. They can also do early logistics plans based on the donation amounts received from different donors. Donated supply informs resource planning, transportation, warehousing, and allocation decisions.

The remainder of this paper is outlined as follows. The related literature is described in Sect. 2. Sections 3 and 4 discuss the methodology and experimental design. Section 5 shows data analysis and model results. Summary of key findings and our conclusions are presented in Sect. 6.

2 Related literature

Supply chain flows in non-profit hunger relief organizations (NPHRO) and profit-oriented organizations are similar, but their structure and goals are different. Profit-oriented organizations purchase products from suppliers and sell them to customers. NPHROs receive donations from various sources and distribute, rather than sell, those donations to people in need. Profit-oriented organizations select suppliers that will support their product needs and have a significant impact on business performance and success (Kannan and Tan 2002). NPHROs, however, typically solicit (as opposed to select and reject) donors because one of their prime goals is to increase monetary or in-kind donations to help achieve their mission of reducing hunger. Supplier selection and evaluation in a commercial organization is a complex decision-making problem with various criteria, and it is one of the main focuses for researchers (Benyoucef et al. 2003; Makhitha 2017). In our discussion of the related literature, we identify supplier selection criteria that can be used to evaluate donors. We focus on the criteria used in the selection process rather than the decision-making aspects of the problem to extract a useful set of measures for evaluating and analyzing the performance and behavior of donors. We believe that identifying behavioral attributes can help support operational decision-making, particularly within the area of estimating and allocating available supply. Our review also briefly summarizes the related literature in the area of humanitarian donation prediction.

2.1 Supplier evaluation criteria

A large stream of the supplier selection literature consists of empirical studies that assess the importance of various criteria in specific business enterprises (Dickson 1966; Swift 1995; Nair et al. 2015). The work of Dickson (1966) is one of the earliest studies to rank the importance of supplier selection criteria by sending questionnaires to 273 purchasing agents and managers from the US and Canada. The most important criteria identified from this study were quality of the product, on-time delivery, and performance history. The dimensions of quality,on-time delivery, reliability, flexibility and cost/price were evaluated in other settings such as automotive (Dweiri et al. 2016), hospitals (Ishtiaq et al. 2018), manufacturing industries (Nair et al. 2015) and small businesses (Makhitha 2017; Naude 2013). Besides these, retailers (i.e., food, supermarkets, etc.) also consider return policy, food safety, and trace ability for selecting their supplier (Hansen 2001; Sternquist and Chen 2006; Lin and Wu 2011).

The relative importance of the evaluation criteria varies by business segment, timeline, and market segment. For example, critical strategic factors identified in Just-In-Time production environments are culture, technology, relationships, cost, quality, time, and flexibility (Sarkis and Talluri 2002). In contrast, Swift (1995) notes that product, availability, dependability, experience, and price are viewed as vital by purchasing managers from chemical, electronic, and transportation industries. Verma and Pullman (1998) ranked quality, on-time delivery, cost, lead-time, and flexibility using a multi-nominal logit model. In their research, quality was given the highest priority and flexibility the lowest. Tavana et al. (2016) focused on four criteria: price, quality, delivery time, and technology. The author found quality, delivery time and technology, and combination delivery time, technology, and price became single, two, and three influential criteria.

In more recent years, the importance of factors such as the geographical position of the supplier (Weber et al. 1991), the communication medium (Halley 2000), and environmental and ethical sustainability (Goebel et al. 2012; Gualandris et al. 2015; Azadnia et al. 2015;

Savage et al. 2018; Alikhani et al. 2019) began to emerge. A review of 78 supplier selection and evaluation research papers published from 2005 to 2012 found that cost, quality, delivery, service, supplier's profile, technology, and capability are the most-cited criteria for supplier selection (Mukherjee 2016).

From the above discussion, a number of common criteria consistently emerge in the literature as important: quality, on-time delivery, consistency, and flexibility. In addition to the total number of products in the portfolio and service location, these criteria can also be applied to the non-profit donor evaluation problem posed in this study. We formally justify and define these data-driven attributes in the methodology section of this paper.

2.2 Humanitarian donation prediction

A significant function of humanitarian organizations is providing aid (i.e., relief items or supply) during natural disasters. Aid can take the form of food (Wang 2013; Brock and Davis 2015) or non-food items like medical supplies or money (Ülkü et al. 2015; Özpolat et al. 2015). Many of the papers have focused on demand forecasting (Van der Laan et al. 2016; Law et al. 2019; Kim et al. 2019; Singh et al. 2006; Sarhani and El Afia 2014; Syntetos et al. 2016). However, a growing number of papers focus on relief supply forecasting, particularly for food items in slow-onset environments like famine (Holguín-Veras et al. 2012; Ahire and Pekgün 2018; Altay and Narayanan 2020). When forecasting for food items, the approach can be done at a macro or micro-level. Examples of macro-level forecasts are predictions based on all donations received at a specific point in time, all donations received at a location, or all donations of a specific type (Davis et al. 2016; Pugh and Davis 2017; Brock and Davis 2015). Micro-level forecasts focus on contributions made by individual donors.

Several approaches have been considered in the literature for predicting relief item supply or donor intention. Ülkü et al. (2015) developed a mathematical model to determine an organization's likelihood of receiving cash versus in-kind donations. Özpolat et al. (2015) developed a donation calculator to enlighten donors about the benefits of sending cash versus in-kind donations. The calculator measured the tradeoffs between buying and shipping the product versus providing cash to an organization that can purchase the items locally. With respect to predicting the actual quantities of relief supply, the models have been more statistical in nature. Examples include support vector machines (Korolov et al. 2016), time series models (Davis et al. 2016), genetic algorithms (Wang 2013), and neural networks (Brock and Davis 2015; Nair et al. 2017).

The most relevant work for predicting in-kind food donations has been macro-level forecasts using time-series or neural network-based approaches (Davis et al. 2016; Brock and Davis 2015; Nair et al. 2017; Pugh and Davis 2017). Data were clustered along different dimensions representing characteristics of the product (e.g. dry, frozen storage), location (warehouse specific or network level), or type of donor (e.g. retail donor). However, no statistical-based clustering approach was used or specific attributes around donor behavior considered. In addition, we find no evidence in the literature of an ensemble-based approach for predicting relief supply.

2.3 Research contribution

Table 1 summarizes the relevant prediction problems considered in humanitarian supply chains. Previous researchers used different stand-alone forecasting approaches to predict humanitarian contributions. In this research, an ensemble prediction model is used to forecast the number of donations received from donors clustered by behavioral attributes.

To the best of our knowledge, supplier behavior within the context of non-profit hunger relief organizations, specifically donors, has not been addressed in the open literature. This research addresses the dearth of literature on food bank donations in general and donors' behavior in particular to create donor profiles informed by attributes from the for-profit supplier selection problem. Introducing this concept of a donor profile is a unique contribution to the literature and can be adapted by other humanitarian organizations.

An unsupervised machine learning approach has been used to cluster donors based on their behavioral attributes in an effort to improve prediction accuracy over other macro-level approaches considered in the literature. Cluster analysis is a data analysis technique (Rao and Reddy 2012), and one of the most commonly used data mining applications for grouping data based on their similarity (Venkateswarlu and Raju 2013). K-means clustering algorithm is a data pattern analysis technique (Lima et al. 2014) used in pattern recognition (Han et al. 2011), image processing (MacQueen et al. 1967), document classification (Dunham 2006), economic science (Tan et al. 2016). Within the humanitarian logistics literature, K-mean clustering is used in blood donor hemoglobin trajectories (Vinkenoog et al. 2019), mining blood donors information (Venkateswarlu and Raju 2013), blood donors behavior identification (Ashoori and Taheri 2013), improving donors selection (Bianchi et al. 2005), humanitarian logistics coordination during natural disaster (Lima et al. 2014; Santos Lima et al. 2014; Tatham et al. 2010; Altay and Pal 2014), and humanitarian operations (Altay and Pal 2014). We believe the use of cluster analysis to classify donor behavior for hunger relief organizations has not been addressed.

3 Methodology

3.1 Problem background

We use the same food bank mentioned in the study of Davis et al. (2016). Therefore, we refer the reader to that paper to understand the full supply chain structure and flows within the network. We limit our discussion here to the actual data description and modeling assumptions. We collected historical data for three fiscal years, 2015–2016 to 2017–2018. The fiscal year starts in July and ends in June of the following year. The data contained a total of *853,908* records and *109* variables. The data were filtered by donations, and all distribution and purchase records were removed, which reduced the observations to *105,433*. Table 2 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).

3.2 Research framework

To address the research questions mentioned in Sect. 1, we construct a research design framework that incorporates the following components. The first phase consists of defining the behavioral attributes. We use the data from the food bank to derive the attributes for each donor. The end result is a donor profile that consists of the the supplier behavior attributes and location. A supplier (also called vendor) profile is a common practice for commercial organizations (Bensaou 1999), and can contain their basic information (i.e., name, address,

Table 1 Research contribution

Donation type	Author	Prediction approach										
		SVM	EMD	ARIMA	GA based grey model	MA	EWMA	MLP-NN	Naive	Seasonal naive	Ensemble	Cluster analysis
Disaster product	Korolov et al. (2016)	✓										
Food	Wang (2013)				✓							
In-kind	Davis et al. (2016)			✓		✓	✓					
In-kind	Brock and Davis (2015)							✓				
In-kind	Pugh and Davis (2017)	✓										
In-kind	Research contribution	✓		✓		✓	✓		✓	✓	✓	✓

Table 2 Key donation fields and example

Key variable	Example values	Key variable	Example values
Posting date	07-01-2016	Donor ID	1529
Receiving location	Raleigh	Product type	Meats
Donor afiliation	Local	Storage classification	Dry
Product quality	F2E	Gross weight (lbs.)	100,000

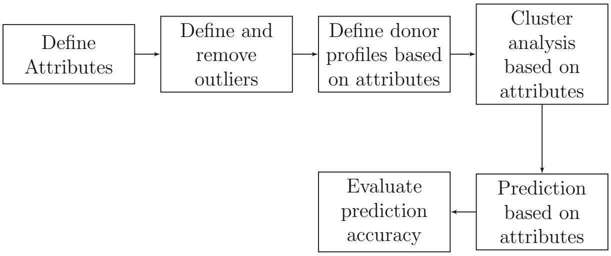


Fig. 1 Research framework

etc.) along with performance and risk score (Makhitha et al. 2013; Thompson 1990). In the second phase, we use different clustering methods to group donors based on their behavior. Lastly, we develop prediction models for food donations based on clustered donors. Our prediction model is an ensemble-based approach incorporating several time-series and machine learning forecast models. By clustering donors based on similar behavior, we aim to reduce the variability that may result when data are aggregated by other methods (as seen in Davis et al. (2016)). Figure 1 summarizes the research framework.

3.3 Definition of behavioral attributes

To answer research question 1, we proposed seven non-profit organization attributes which are summarized in Table 3. The proposed features are based on five of the most significant supplier evaluation criteria found from the literature review: On-time delivery, quality of the product, number of products in the portfolio, service, consistency, and flexibility. The interpretation and construction of these attributes for non-profit organizations like food banks are discussed in the subsequent sections.

3.3.1 Reliability

One of the critical attributes to evaluate a supplier is on-time delivery or shipment of goods (Benyoucef et al. 2003; Dickson 1966). On-time delivery can be further subdivided into two parts: always on time and sometimes late (Sarkis and Talluri 2002). A non-profit organization can similarly determine if a donor is reliable based on the delivery frequency. We propose a reliability*(Rel)* attribute for assessing donors on-time delivery performance based on donation frequency. We determine donor reliability according to Eq. (1) for a specific time period N. A donation event is a binary variable (E_p) that takes on the value of 1 if a donor donates at

Table 3 For-profit and non-profit attributes comparison

Proposed non-profit equivalent attributes	For-profit supplier attributes	References
Reliability	On-time delivery	Naude (2013), Hansen (2001), Sternquist and Chen (2006), Lin and Wu (2011), Tavana et al. (2016), Makhitha et al. (2013) and Kulkarni et al. (2004)
Waste percentage	Quality	Dickson (1966), Naude (2013), Hansen (2001), Sternquist and Chen (2006), Lin and Wu (2011), Verma and Pullman (1998), Tavana et al. (2016), Makhitha et al. (2013), Crow et al. (1980), Lehmann and O'shaughnessy (1982) and Segal (1989)
Donation product variety	Product portfolio (supplier profile)	Dickson (1966), Swift (1995), Mukherjee (2016) and Makhitha et al. (2013)
Received quantity	Consistency and flexibility	Nair et al. (2015), Naude (2013), Hansen (2001), Sternquist and Chen (2006), Lin and Wu (2011), Verma and Pullman (1998), Weber et al. (1991) and Makhitha et al. (2013)
Service score	Service	Dickson (1966), Weber et al. (1991), Mukherjee (2016), Crow et al. (1980), Lehmann and O'shaughnessy (1982), Segal (1989) and Luthra et al. (2017)
Affiliation	–	
Donation quality score	–	

least once during period p, and 0 otherwise.

$$Rel(\%) = \frac{\sum_{p=1}^{N} E_p}{N} \times 100 \tag{1}$$

In our study, we aggregate data by month for each donor. Therefore, as an example, if a donor donates 36 months in three consecutive fiscal years ($N = 36$), the reliability of that donor is $(36 \times 100)/36 = 100\%$.

3.3.2 Wastage percentage

Dickson's survey on supplier criteria explains that buyers focused on vendors' operational quality control systems, warranties, and claims handling policies to find the loss of product value (Dickson 1966). Similarly, if defective or damaged donations are received, sorting and disposal of unusable items require an investment of time and cost for non-profit organizations.

Furthermore, higher wastage amounts reduce efficiency. We propose to evaluate the level of waste obtained from a donor based on Eq. (2) over the time period considered in our study. For example, in the last 36 months, one donor donated 10,000 lbs. of which 250 lbs. was waste. Therefore, the wastage percentage score (Waste-Sco) for this donor is equal to $(250 \times 100/10000)\% = 2.5\%$.

$$\textit{Waste-Sco (\%)} = \frac{\textit{Total wastage amount (lbs)}}{\textit{Total donations (lbs)}} \times 100 \tag{2}$$

3.3.3 Donation product variety

The number of products in the portfolio is a crucial supplier attribute (Swift 1995). A commercial organization like a retail chain is always looking for a supplier with many products in its collection. They also give priority to those suppliers who have more product lines in their portfolio (Makhitha 2017). For non-profit organizations like food banks, we observe that donated products are classified by food type or storage type (Davis et al. 2016). The storage type classification takes on four distinct values: dry (D), frozen (F), produce (P), refrigerated (R). Therefore, we propose to measure donation product variety *(Don-Pro-Var)* as a function of the storage type classification as shown in Eq. (3). We define v_i, where $i \in \{D, F, R, P\}$, as a binary variable that takes on the value 1 if donor donated product type i at any point in time during the time period of interest, 0 otherwise.

$$\textit{Don-Pro-Var (\%)} = \frac{v_D + v_F + v_R + v_P}{4} \times 100 \tag{3}$$

3.3.4 Service score

Supplier geographical production and distribution facilities are a significant evaluation point in for-profit organizations (Luthra et al. 2017). Retailers evaluate suppliers based on their area served by a supplier called the service area. Retailers prefer to get their supply shipped direct to their local warehouses or retail stores since it increases flexibility and saves logistics cost (Zsidisin 2003). In a similar fashion, donors that can supply products to several locations will save distribution costs for a non-profit organization. Equation (4) defines our measure of service score for hunger relief organizations as the fraction of the NPHRO locations a donor can ship product to. For example, if a philanthropic organization has "10" branches, and one donor sends donations to "5" branches for a specific time period of N, then the service score is equal to $(5 \times 100/10)\% = 50\%$.

$$\textit{Ser-Sco(\%)} = \frac{\textit{Number of location donations donated}}{\textit{Total number of locations}} \times 100 \tag{4}$$

3.3.5 Quality score

Each supplier needs to maintain operational controls (including quality control and inventory control), meet quality specifications, and deliver high-quality products (Dickson 1966). Donors' quality score (Qual-Sco) defines the overall quality of donated food. A higher-quality score means better quality food given. In the donation dataset, food quality is divided into three categories—foods to encourage (F2E), salvage, and others. F2E is high-quality/nutritious food that the food bank may want to receive more of from donors. The quality score is defined as the percentage of F2E donations received from the donor, relative to the other total contributions over the specific time period of interest (N). Assume one donor gave 950 out of 1000 lbs of F2E donation in a defined period. For that donor, the quality score is 95%.

$$\textit{Qual-Sco (\%)} = \frac{\textit{F2E donations (lbs)}}{\textit{Total donations (lbs)}} \times \textit{100} \tag{5}$$

3.3.6 Received quantity

Due to the competitive world, consistency and flexibility are some of the vital supplier attributes. Changing the order quantity and delivery with consistent quality provides more weight for choosing suppliers in a commercial organization (Weber et al. 1991). Getting an even amount of donations from donors helps a non-profit organization to plan efficient distribution. In Eq. (6), we define received quantity *(Rece-Quant)* consistency in terms of the coefficient of variation. The period for the basis of this calculation is months but can be equivalently determined from weeks.

$$\textit{Rece-Quant} = \frac{\textit{Standard deviation of monthly donated amounts}}{\textit{Mean of donated amounts}} \tag{6}$$

3.3.7 Affiliation

In our dataset, there is a category that describes the affiliation of the donor. Some of the values indicated include State Government, Federal Government, and Feeding America (FA). A description of the donor affiliation values are shown in Table 4.

3.4 Construction of donor profiles

We utilized three fiscal years of data to analyze and create a profile for each donor. For the reliability, service score, donation product variety, and received quantity attributes, we aggregate every donor's donation amount to a monthly level. For the quality score, and wastage percentage analysis, donation data were aggregated over the whole time period as per the stated Eqs. (2) and (5). To analyze Affiliation, we categorize donors into six segments, as indicated in Table 4. The summary of data modeling is shown in Table 5.

3.5 Clustering methods

Cluster analysis helps to identify and partition the same groups of data into sub-classes (Venkateswarlu and Raju 2013). In this research, we defined seven attributes (Table 5), in which six attributes are numerical, and one attribute is categorical. K-means method is used when all numeric attributes are used. K-medoids method is used when the data contains both numeric and categorical attributes since the k-means clustering algorithm can not directly be used on categorical data.

K-means method partitions the data into k clusters by minimizing the sum of squared distances between each point and cluster centers (Fahim et al. 2006). Assume, there are $x_1, x_2, \ldots, x_n$ data points to be clustered in k numbers of clusters. In a cluster j, m_j is the centroid, x_i is object in cluster, and $d(x_i, m_j)$ is the euclidean distance between the object and the centroid. *k-means* clustering algorithm assigns data points to a cluster such that the average squared euclidean distance,$d(x_i, m_j)$ is minimized. Equation (7) illustrates the

Table 4 Descriptions of key values for the donor affiliation in the data-set

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
State Govt.	Food obtained through funds allocated by the state
USDA	Food obtained through the government commodity program
Food Bank	Food received from other food banks

Table 5 Data modeling from the donation database

Data source	Attributes	Measurement unit	Data type
Donation database	Reliability	Percentage	Numeric
	Service score	Percentage	Numeric
	Don-prod-variety	Percentage	Numeric
	Quality score	Percentage	Numeric
	Wastage	Percentage	Numeric
	Received quantity	Coefficient of variation	Numeric
Donor database	Affiliation	Donor type	Categorical

calculation for the average squared euclidian distance assuming $n_j < n$ data points are assigned to cluster j (Oyelade et al. 2010).

$$Minimize : \frac{1}{n_j} \sum_{i=1}^{n_j} [d^2(x_i.m_j)] \tag{7}$$

K-medoids clustering is a special type of k-means clustering that minimizes the sum of dissimilarities instead of squared Euclidean distance (Reynolds et al. 2006). Partitioning around medoids is one of three and the most popular K-medoids clustering approach. It is most robust to noise and outliers than *k-means* clustering approach (Reynolds et al. 2006). Euclidean distance only works for numerical data. For mixed data type, distance can be calculated through Gower distance (Gower 1971). It measures dissimilarity between two different variable types (i.e., nominal or ordinal). Nominal variables are converted into binary columns, and a linear combination using average weights is calculated for the final distance matrix. The distance is always a number between 0 to 1 where 0 means identical and 1 represents maximum dissimilarity. Assume, there are n objects having p variables. Define the jth variable of the object i as X_{ij} (where $i = 1, \ldots$, n, j = 1, …, p). If distance between object i and object j is d_{ij} (8) then v_j for object j is calculated as shown in Eq. (9) (Park and Jun 2009).

$$d_{ij} = \sqrt{\sum_{a=1}^{p} \left(X_{ia} - X_{ja}\right)^2}, i = 1, \ldots, n; j = 1, \ldots, n \tag{8}$$

$$v_j = \sum_{i=1}^{n} \frac{d_{ij}}{\sum_1^n d_i}, j = 1, 2, \ldots, n \tag{9}$$

v_j's need to be sorted in ascending order and the first k smallest values will be initial medoids. Then each object should be assigned to the nearest medoid, and the sum of distances need to be calculated from all objects to their medoids (Park and Jun 2009). The second step of the algorithm is to find new medoids from each cluster by minimizing the total distance in its clusters and update those medoids. In the next step, assign each object to the nearest medoids, create clusters, and calculate the sum of distances from all objects to their medoids. Lastly, we compare the sum of distances, if the new value is equal to the old one then stop the algorithm or repeat from the second step (Park and Jun 2009).

To find the optimal number of clusters (k) for both clustering approaches, we use the average silhouette width. Silhouette width is a measure of similarity of data points to its cluster compared to the other groups. It can be calculated from the distance metric. If *a(i)* is the average dissimilarity (i.e. distance) of object i to all other objects of its own cluster and *b(i)* average dissimilarity of object i to other data points in any other clusters then silhouette value can be represented as Eq. (10) (Rousseeuw 1987).

$$f(x) = \begin{cases} 1\text{-}a(i)/b(i), & \text{if } a(i) < b(i) \\ 0, & \text{if } a(i) = b(i) \\ b(i)/a(i)\ \text{-}1, & \text{if } a(i) > b(i) \end{cases} \tag{10}$$

From Eq. (10), the silhouette width range can be between − 1 to + 1, where a value close to + 1 implies the data point is matched to its own cluster point and dissimilar to other clusters. The cluster number (k) is optimal when *s(i)* is the highest (Rousseeuw 1987).

3.6 Prediction models

Prediction of time series data means forecasting the future based on historical observations (Larssen et al. 2007). There are several statistical models available for time-series predictions. Frequently, fitted models are used to directly predict several months, and without updating its parameters (Haggett and Cliff 2005). Less computation time and cost are the benefits for those models, whereas there is a high chance of increasing error. This research presents an ensemble-based forecasting approach that utilizes six time series models: moving average, naive forecasting method, seasonal naive, exponential smoothing, autoregressive integrated moving average, and support vector regression.

Moving average (MA) is the most simple forecasting method. The forecast for the next period is determined as the average of the most recent N observations. MA is formally defined by Eq. (11), where F_t is the forecast made in the period $(t-1)$, y_i is the observation in period i, and N is the total number of periods (Nahmias and Cheng 2001):

$$F_t = \frac{1}{N} \sum_{i=t-N}^{t-1} y_i \tag{11}$$

Based on preliminary research, $N = 2$ performs well for the data considered in this study (Paul and Davis 2019).

Naïve forecast model is determined from the last observation as defined by Eq. (12) (Hyndman and Athanasopoulos 2018).

$$\hat{y}_{T+h|T} = y_T \tag{12}$$

Seasonal naïve is similar to naïve. However, instead of using the last observed value, it uses the same value from the prior season. It is formally defined by Eq. (13) for the forecast for time $(T+h)$ where m is the seasonal period (m = 12 in this study), k is the integer part of $(h-1)/m$ (Hyndman and Athanasopoulos 2018).

$$\hat{y}_{T+h|T_s} = y_{T+h-m(k+1)} \tag{13}$$

Another popular forecasting method is exponential smoothing, where more recent observations are given the highest weights. Pagels, Gardner, and Taylor introduced methods with a multiplicative trend (Pegels 1969), damped additive trend (Gardner 1985), and multiplicative damped tend (Taylor 2003) respectively. Hyndman and Athanasopoulos (2018) summarized all models, methods, and equations based on the trend and seasonal component and proposed a state-space model for exponential smoothing (ETS) as shown in Table 6. They also provided all equations for each of the models in the ETS framework (Hyndman and Athanasopoulos 2018). In the state-space model there are two equations—state equation and observation equation. A state equation describes the state process dynamics, and observed series to a latent state process is described by a observation equation (Bai et al. 2013). The recursive nature of the models provide flexibility which is the main advantage of ETS models (Durbin and Koopman 2012).

Auto-Regressive Integrated Moving Average (ARIMA) is widely used for time series forecasting like exponential smoothing because of its flexibility. ETS model utilizes trend and seasonality for non-stationary data, whereas ARIMA mainly focuses on auto-correlations in the stationary data (Hyndman and Athanasopoulos 2018). ARIMA model consists of three parameters (p, d, q) representing the number of auto-regressive terms, the number of terms needed to convert a non-stationary time series, and past forecast errors respectively (Tabachnick et al. 2007). For ARIMA, d is defined first with the Dickey–Fuller unit root test, then the auto-correlation function (ACF) and the partial auto-correlation function (PACF) are then examined to determine the number for p and q. ACF and PACF plots can help determine the appropriate value for p and q (Hyndman and Athanasopoulos 2018).

Support Vector Machine is a popular machine learning approach originally developed for solving classification problems (Noble 2006). It was subsequently extended for regression as support vector regression (SVR) (Vapnik 2013). An SVR model's objective function is to minimize the coefficients vector, whereas linear regression minimizes the sum of squared errors. Another interesting fact of SVR is it can learn data patterns and make predictions based on its learning. Assume the training data set can be defined as follows where χ denotes the space of the input patterns (e.g., $\chi = \mathbb{R}^D$):

$$\{(\mathbf{x}_1, y_1)\ldots(\mathbf{x}_l, y_l)\} \subset \chi \times \mathbb{R} \tag{14}$$

We need to find the regression function $F(x) = \mathbf{w}^T\mathbf{x}_i + b$, where $\mathbf{w} \in \mathbb{R}^D$ represents the regression weights and $b \in \mathbb{R}$ is the intercept, assuming $F(x)$ is linear (Pugh and Davis 2017). If y_t, $\mathbf{w}$, $\mathbf{x}_i$, and ϵ are the future value, coefficient, predictors, and maximum error respectively, a simple representation of the model's objective function and constraints (Smola and Schölkopf 2004) are:

Table 6 A two-way classification of exponential smoothing (Hyndman and Athanasopoulos 2018)

Trend component	Seasonal component		
	N (none)	A (additive)	M (multiplicative)
N (none)	(N, N)	(N, A)	(N, M)
A (additive)	(A, N)	(A, A)	(A, M)
A_d (additive damped)	(A_d, N)	(A_d, A)	(A_d, M)

$$Minimize : \frac{1}{2}||\mathbf{w}||^2 \tag{15}$$

$$|y_t - \mathbf{w}^T \mathbf{x}_i - b| \leq \epsilon \text{ for } i = 1..l \tag{16}$$

The ϵ ensures that errors are small and do not affect the objective function. There is a possibility that the problem may be infeasible. Infeasibility can be handled by modifying constraints to include slack variables $\xi_i, \xi_i{}^*$. The resulting problem formulation is as follows (Vapnik 2013):

$$\text{Minimize} : \frac{1}{2}||\mathbf{w}||^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i{}^*) \tag{17}$$

$$y_t - \mathbf{w}^T \mathbf{x}_i - b \leq \epsilon + \xi_i \text{ for } i = 1..l \tag{18}$$

$$\mathbf{w}^T \mathbf{x}_i - y_t + b \leq \epsilon + \xi_i{}^* \text{ for } i = 1..l \tag{19}$$

$$\xi_i, \xi_i{}^* \geq 0 \tag{20}$$

The regularization parameter ($C > 0$) controls the trade-off between the flatness of $F(x)$ and the amount up to which deviations larger than ϵ are tolerated. If $F(x)$ is assumed to be linear then ϵ-insensitive loss function is $|\xi_\epsilon|$ derived as Eq. 21 (Vapnik 2013).

$$|\xi|_\epsilon = \begin{cases} 0 & \text{if } |\xi| \leq \epsilon \\ |\xi| - \epsilon & \text{otherwise} \end{cases} \tag{21}$$

If the SVR function is non-linear, it can be obtained from the kernel function. Regularization parameter (C), width of (ϵ)-insensitive tube , the kernel function (K), and the number of attributes are four parameters for non-linear SVR tuning. Based on a review of the literature, (Kulkarni et al. 2004; Pedregosa et al. 2011; Kowalczyk 2017), we use the most popular Gaussian radial basis function kernel for SVR fix the ϵ value at 0.1, and use grid search (with step size of 0.01) for tuning C and γ between $(1 - 2^9)$ and (0.1–0.9), respectively. A more comprehensive discussion about SVR can be found in the literature and we note the following references here for statistical learning equations (Cortes and Vapnik 1995), loss function (Cortes and Vapnik 1995), kernel functions (Smola and Schölkopf 2004), and tuning models (Pedregosa et al. 2011).

Bates and Granger (1969) first showed a combination of forecasts provides better forecast accuracy. Then Clemen (1989) claimed that averaging multiple forecasts can also significantly increase forecast accuracy. We propose an Ensemble Prediction Model (EM) by taking an average of the predictions generated by the six forecasting models discussed earlier. Equation (22) shows the ensemble model prediction where, $F_{em}(t)$, $F_m(t)$, $F_e(t)$, $F_\eta(t)$, $F_s(t)$, $F_\alpha(t)$, and $F_\nu(t)$ are forecast generated by Ensemble model, Moving average, ETS, Naive, Seasonal naive, ARIMA and SVR respectively.

$$F_{em}(t) = \frac{1}{6}[F_m(t) + F_e(t) + F_\eta(t) + F_s(t) + F_\alpha(t) + F_\nu(t)] \tag{22}$$

3.7 Evaluation of prediction accuracy

Evaluating prediction accuracy is essential for measuring the performance of the model in reducing supply uncertainty. Our approach consists of dividing the data into two sets: train data (in-sample) and test data (out-of-sample). The training data is used to estimate forecasting model parameters, whereas test data is used to evaluate model accuracy. The training data contains 30 data points, and test data contains one data point in each iteration. We use a rolling forecasting method in the sense that for each succesive time period, new training data is built by removing the first data point and adding the next month into the training set. The model parameters are then estimated again and a new one-step ahead forecast is constructed. Table 7 shows the train and test data set for the rolling forecast.

The mean absolute percentage error ($MAPE$) is then used on out-of-sample data for forecast accuracy. It is also used to compare the prediction accuracy of different clusters. A lower mean absolute percentage error implies greater forecast accuracy. MAPE is determined according Eq. (23) and calculated from the six *(T = 6)* one-step ahead forecasts generated. The actual value and predicted value come from test data and ensemble models, respectively.

$$MAPE = \frac{1}{T}\sum_{i}^{T}\left|\frac{Actual\ value - Predicted\ value}{Actual\ value}\right| \times 100 \tag{23}$$

4 Experimentation

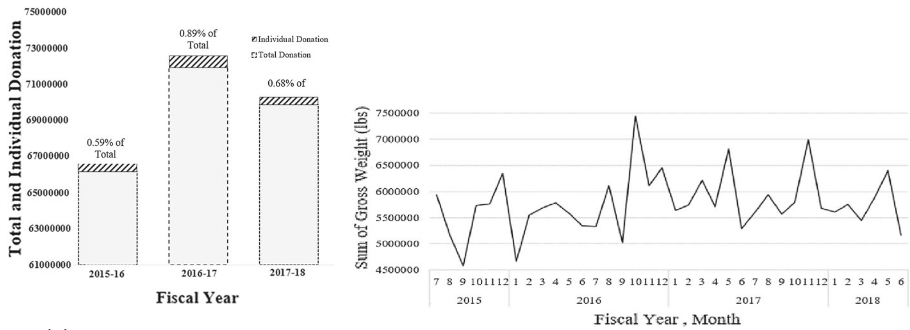
4.1 Data pre-processing

Before analysis, data were pre-processed in “R” to identify potential errors and outliers. We note a few of the cases here. For example, there was a large positive donation of a donor and a negative donation of the same donor on the same posting date. This happens when contribution amounts need to be adjusted. To minimize the effect of these adjustments, data is aggregated monthly for each donor. There was also evidence of missing data and potential data errors during donations entry. Missing values had a significant impact on frequency analysis and future predictions. Missing values and data errors were defined as outliers and removed from the data. Unnecessary variables were filtered, and removed to get the final donation data set.

During pre-processing we also found an observation that reflects individual donor contributions. One donor id aggregated in-kind donations made by individuals and, therefore, detailed information regarding the address, city, and state were missing in the donor file. The total contribution of all individual donors is shown in Fig. 2a, which is less than 1%. The most substantial individual donations occurred from October to December. The high amount of donations during these months is attributed to the North Carolina State Fair, and holiday donations (Thanksgiving and Christmas). All individual donor contributions can not be tracked from the data set, so those data are not included in our study.

Table 7 Train and test data split

Iteration number	Train data	Test data
1	(Jul 2015–Dec 2017)	Jan 2018
2	(Aug 2015–Jan 2018)	Feb 2018
3	(Sep 2015–Feb 2018)	Mar 2018
4	(Oct 2015–Mar 2018)	Apr 2018
5	(Nov 2015–Apr 2018)	May 2018
6	(Dec 2015–May 2018)	Jun 2018



(a) Individual donors' contribution

(b) Trend analysis of donation amounts

Fig. 2 Three fiscal years donation analysis

Monthly aggregated donation was plotted for trend and seasonal analysis in Fig. 2b. There was no increasing and decreasing trend for the last three fiscal years. We did note some evidence of a seasonality pattern from October to December each year.

4.2 Scenarios for clustering and prediction model

Seventeen scenarios are considered for clustering donors based on the seven attributes, as summarized in Table 8. *K-medoids* clustering is used for the first six scenarios because of the presence of the non-numeric attribute, affiliation, considered with the grouping. For the rest of the scenarios *k-means* clustering is used.

5 Result

5.1 Overview

Out of 2132 donors, 952 donors donated during the period we considered. When we aggregated the monthly donation amounts for each donor, 14 donors' donation amounts became zero. We consider those donations as outliers and exclude those donors from our list. Therefore, the total number of donors in our analysis is 938.

5.2 Research Question 1: Donor behavioral attributes

Figure 3 provides a graphical representation of the donor behavior across the attributes proposed. We note several observations with respect to each donor attribute.

5.2.1 Observation 1: Reliability

The reliability analysis for each fiscal year (Fig. 3a) shows the percentage of 100% reliable donors increased by 4% in 2016–2017 and 2017–18. However, based on three fiscal years of data, only 219 donors donated each month of 36 months (Fig. 3b) and approximately 32% of donors have reliability scores higher than 90%). Simultaneously, one-third of donors

Table 8 Clustering scenarios and methods

Scenario	Attributes							Clustering Method
	Affiliation	Reliability	Service Score	Donation Product Variety	Quality Score	Received Quantity	Wastage Percentage	
1	✓	✓	✓	✓	✓	✓	✓	K-Medoids Clustering
2	✓	✓						
3	✓	✓	✓					
4	✓	✓	✓	✓				
5	✓	✓	✓	✓	✓			
6	✓	✓	✓	✓	✓	✓		
7		✓						
8			✓					K-Means Clustering
9				✓				
10					✓			
11						✓		
12							✓	
13		✓	✓					
14		✓	✓	✓				
15		✓	✓	✓	✓			
16		✓	✓	✓	✓	✓		
17		✓	✓	✓	✓	✓	✓	

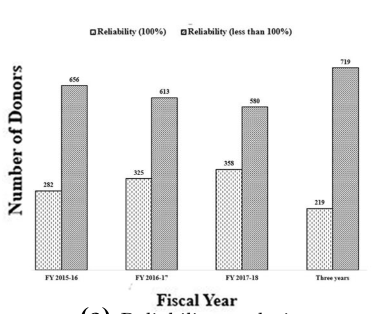
reliability is less than 10% which indicates that they contribute infrequently during the three fiscal years.

5.2.2 Observation 2: Wastage Percentage

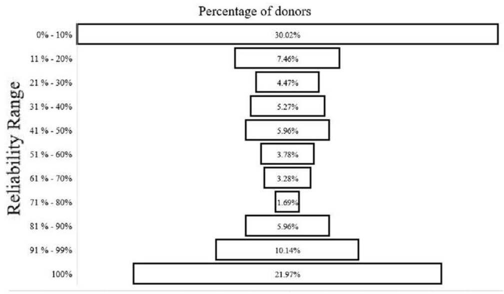
Approximately 28% of the 938 donors had some donations that were not usable and had to be discarded. Specifically, six donors had wastage amounts larger than there donation amounts. There are only four donors who had more than 50% wastage of donation amount. Approximately 210 donors had less than 10% wastage. There was no correlation found between wastage amounts, donation amounts, and wastage percentages after statistical analysis.

5.2.3 Observation 3: Quality score

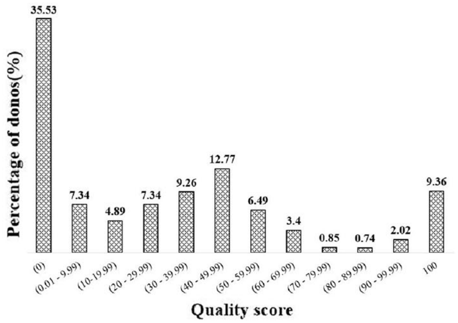
Out of 938 donors, 334 (approximately 35%) did not donate any F2E foods, whereas 88 donors donated all F2E foods during the 3 years. Approximately 13% of total donors contribute at least 50% of highly encouraged foods. The summary of the quality score analysis is given in Fig. 3c.



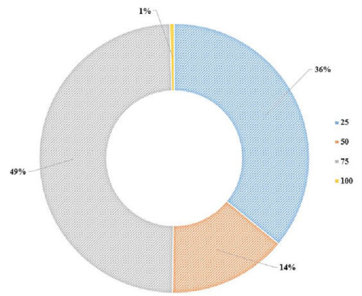
(a) Reliability analysis



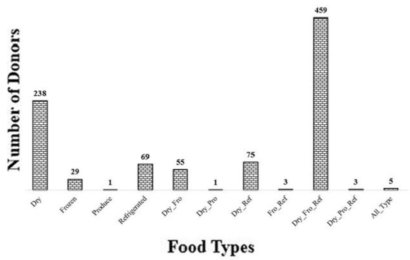
(b) Reliability range vs. number of the donors



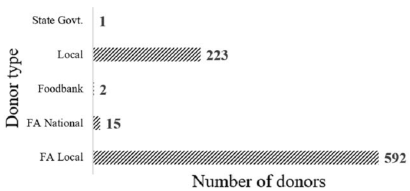
(c) Quality Score analysis



(d) Don-Pro-Var score analysis



(e) Number of donors vs. types of food donated



(f) Affiliation analysis

Fig. 3 Attributes analysis

5.2.4 Observation 4: Donation product variety

Figure 3d shows that almost 50% of donors provide three types of food, the majority of which was dry, frozen and refrigerated as seen in Fig. 3e. Only a small number of donors gave a combination of dry, produce and refrigerated. For the single product type donations, dry goods dominate as the type of food given.

5.2.5 Observation 5: FA affiliation

During the affiliation analysis, 105 donors had more than one type of affiliation. Those donors were considered outliers and removed from the dataset. Our results show that donations acquired from FA affiliated donors are higher than donors not affiliated with FA. Figure 3f also shows that the majority of donors are FA local and local donors.

5.2.6 Observation 6: Service score

FBCENC has six branches—Durham *(D)*, Greenville *(G)*, New Bern *(N)*,Raleigh *(R)*, Sandhills *(S)*, and Wilmington *(W)*. Raleigh is the main branch that also works as a distribution center. So, the denominator of Eq. (4) will be six. As per defined Eq. (4) three fiscal years of data were analyzed for service score, which is summarized in Table 9 for the 938 donors. 78.35% (735 total) donors donated to a single branch and therefore have a service score of 16.67%. There are only six donors who have a service score of 33.33%. They donated to a combination of any two branches. Raleigh branch received donations from most of the donors because it is the main branch.

Table 9 also summarizes the number of donors that contribute to each branch under each service score. Approximately 47% of donors that donated to single branches gave to the main branch Raleigh. Wilmington branch and Durham branch receive donations from 122 donors and 90 donors, respectively.

5.2.7 Observation 7: Received quantity

There were 91 donors with a received quantity of more than 1.5, which corresponds to higher donations variability. There are only nine donors who had scored less than 0.10.

5.2.8 Donor profile weighted score

To get a sense of the distribution of donors across the entire set of behavioral attributes we developed a weighted total score as shown in the equation below, where the weights w_1 to w_5 take on the values 1, 4, 3, 2, 1. We essentially give more weight to reliability and the inverse of waste. The upper bound of the weighted total score is 1100 which would correspond to a donor having perfect scores for all attributes (100 for all values and 0 for waste).

$$Z = w_1 * SerSco + w_2 * Rel + w_3 * (100 - WasteSco) + w4 * QualSco + w_5 * DonProVar \tag{24}$$

Figure 4 provides a frequency chart of the weighted score and only a few donors have scores above 1000.

Table 9 Service score analysis

Service score (%)	Donated branch	Num. of donors	Branch-wise					
			R	D	W	G	S	N
16.67	1	735	346	90	122	61	58	58
33.33	2	157	120	75	19	51	21	28
50.00	3	14	13	11	0	11	5	2
66.67	4	18	18	17	3	17	16	1
83.33	5	5	5	5	2	5	4	4
100.00	6	9	9	9	9	9	9	9
	Total donors	938	511	207	155	154	113	102

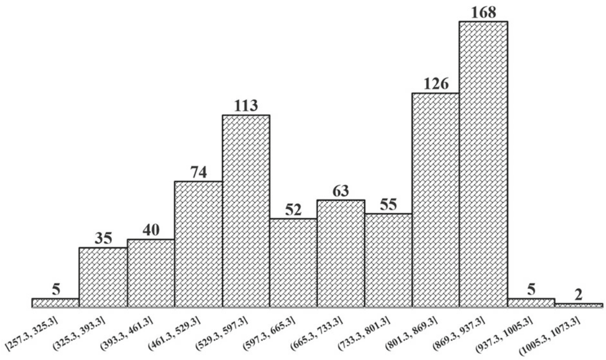


Fig. 4 Total weighted score distribution

5.3 Research Question 2,3: Clustering and prediction result

A summary of the results for each scenario are shown in Table 10. Recall that each scenario corresponds to all or a subset of the seven attributes as defined in Table 8. Our aim, with respect to research question 2 and 3 is to determine the best way to cluster donors and predict food donations, where best is evaluated by prediction accuracy (average MAPE). Each scenario defines the distinct number of donor profiles (i.e. clusters) created based on the behavioral attributes considered. Based on the results in the summary table, we make the following observations.

The highest number of clusters (i.e. donor profiles) created is 10, corresponding to scenarios 3 and 4. Scenario 3 uses three of the seven attributes, while scenario 4 uses four (refer to Table 3). In addition, scenarios 3 and 4 have an average prediction error of 20% and 49%, respectively. The average prediction error is calculated based on the average MAPE generated from each cluster. We also note that scenario 3 has a large number of non-predicted clusters.

Table 10 Clustering and prediction result

Scenario	Clusters	Non-predicted clusters	Average MAPE	Scenario	Clusters	Non-predicted clusters	Average MAPE
1	4	0	29.70%	10	3	0	10.24%
2	4	1	50.45%	11	2	0	16.85%
3	10	5	20.02%	12	2	1	11.46%
4	10	1	49.38%	13	2	0	12.86%
5	4	0	23.36%	14	2	0	12.14%
6	4	0	22.73%	15	3	0	16.42%
7	2	0	12.66%	16	3	0	12.99%
8	6	0	19.12%	17	3	0	12.36%
9	4	0	17.25%				

This implies that there were discontinuous gaps in the time series and therefore a prediction could not be made.This behavior was also observed for scenarios 2 and 4. Non-predicted clusters are excluded from the calculation of the average MAPE for the scenario.

We contrast the results of scenario 3 with those that have 2 clusters, (scenarios 7, 11, 12, 13, 14), which all have prediction errors below 20%. It is interesting to note that these scenarios do not attribute *affiliation* as part of the profile definition and most only use one attribute (scenarios 7, 11, 12).

The lowest average MAPE scenario belongs to scenario 10 where quality score is the only attribute considered . Using the quality score, three donor profiles (corresponding to the three clusters) can be used to partition the donors. The average MAPE for the three clusters along with their behavioral attribute values are shown in Table 11. The total number of donors in each cluster along with their affiliation is shown in Table 12. The first cluster has the highest quality score and the smallest number of donors which are primarily local donors (67%). This cluster of donors donated a high amount of F2E (94.03%). Cluster 2 has the highest number of donors and the lowest quality score, with the bulk of the donors being FA Local. It is interesting to note that cluster 3 has the lowest average MAPE. It has nearly the same amount of donors as cluster 2 but the coefficient of variation (as measured by Rece-Quant) is the lowest. Most interestingly, these donors have higher average reliability (85.33%) and donation product variety (72.61%) score that means they donated a combination of two or three types of products. 93% of donors in this group are FA affiliated local donors and they send their donations to a single branch.

We compare our proposed clustering approach with that of Davis et al. (2016), where clusters were developed based on the information available in the operations data. Donors were clustered using a field called *donor type* which contained values manufacturer, retail/wholesale, and grower. We replicate the same idea to cluster donors for our dataset and run our proposed ensemble model for them. The average MAPE for retail/wholesale, manufacturer, grower clusters, and overall are 6.15%, 22.85%, 72.37%, and 33.79%. However, it is interesting to note that predictions generated from clustering based on behavioral attributes results in a lower average MAPE (10.24%). Therefore, we conclude that quality score (behavioral attribute) is better than an information-based attribute (donor type) when creating a macro-level forecast based on supplier attributes.

5.4 Performance of ensemble approach

Figure 5 presents a stacked graphed for performance comparison between the proposed ensemble model and other forecasting models for the 17 scenarios. The x-axis represents the total number of scenarios and the y-axis represents performance comparison between existing forecasting models with the proposed ensemble model. The lighter bar shows how many scenarios existing models perform better and the darker bar represents how many scenarios the proposed ensemble model performs better. For example, MA performs better for only one scenario and the proposed ensemble model performs better for the remaining 16 scenarios. In general, the ensemble model outperforms the single forecasting models. The forecasting accuracy of the ensemble model is far better than moving average, naive, ARIMA models. For seven scenarios, the ETS model performed better than the ensemble model. But the average MAPE difference is less than 1.5%. Both SVR and Snaive forecast accuracy is better than the ensemble model for six scenarios. For SVR, the average MAPE difference with the proposed model is less than 1.8%.

Table 11 Attributes for the lowest average MAPE scenario

Cluster	Average MAPE (%)	Rel	Ser-Sco	Don-Pro-Var	Qual-Sco	Rece-Quant	Waste-Sco
1	12.13	32.62	34.03	46.47	94.03	0.75	6.91
2	13.06	53.09	20.96	48.51	4.10	0.66	0.92
3	5.53	85.33	21.71	72.61	43.29	0.44	0.67

Table 12 Donors' affiliation for the lowest average MAPE scenario

Cluster	Total donors	Donors' affiliation				
		FA Local	Local	FA national	Foodbank	State govt.
1	71	11	48	0	11	1
2	353	275	67	4	7	0
3	314	292	14	1	7	0

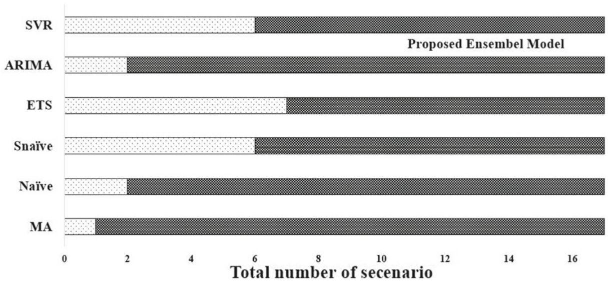


Fig. 5 Performance comparison between proposed ensemble model and existing forecasting models

6 Conclusion

6.1 Key findings

Our research addresses the management of uncertain donated supply in humanitarian supply chains and is motivated by food bank operations. We specifically contribute to the problem of predicting donor contributions and clustering donors based on their giving behavior. We define donor behavioral attributes to cluster donors, informed by the for-profit supplier evaluation literature. We aim to develop donor profiles that are subsequently used to predict donations. Our local food bank-specific results are summarized below:

- From Table 10 and Fig. 6, it can be seen that prediction accuracy is higher for k-mean cluster scenarios than k-medoids clusters scenarios. Non-numeric attribute affiliation is considered in k-medoids scenarios. Our results suggest it is better not to consider affiliation as an attribute when constructing donor behavior profiles for predicting contri-

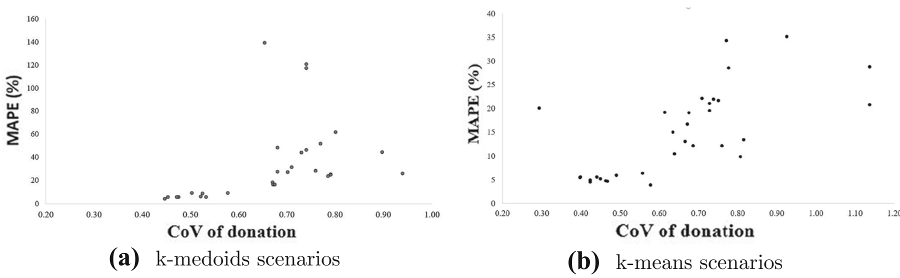


Fig. 6 Scatter-plot of COV by MAPE for scenarios

butions. It is best to report the affiliation as a descriptive attribute after the clusters have been defined.

- Scenario 2 prediction accuracy is only 50%, and the data in one of the clusters has the highest coefficient of variation (CoV) in comparison to all other scenario clusters. Our results suggest that a combination of affiliation, reliability, service score and donation product variety should not be used for clustering donors for forecasting donation accuracy.
- Figure 6 shows that there is a high positive correlation between the coefficient of variation(COV) of monthly donation and prediction accuracy. When CoV increases, MAPE also increases. Our findings are also consisting with current literature (Davis et al. 2016; Wang 2013).
- K-means clusters prediction accuracy performs better than K-medoids clusters.

Most humanitarian organizations do not analyze donor behaviors. To answer the first research question, we defined seven donor behaviors and created a profile for each donor. Defining donor profiles opens new area in the humanitarian research field. To handle the second research question, seventeen scenarios were created based on seven attributes. The results indicate that the quality score (donation of food to encourage) is the best way for creating donor profiles that result in minimum forecast error (or highest forecast accuracy). In essence, our approach finds a parsimonious minimum variance donor attribute grouping from which profiles can be created. Overall, in response to our second research question, it can be said that a single attribute model is an excellent way to cluster donors instead of using a multi-attribute model. Figure 5 showed summary performance of the proposed ensemble model. To answer the third research question, it can be concluded that the ensemble model can be a better choice to predict donations instead of traditional forecasting methods in the humanitarian supply chain.

6.2 Limitations and future research

There are several ways that we can extend our current results. We can modify our model such that it can predict inconsistent donations events and amounts simultaneously. We also want to explore the impact of defining predictions on different time scales and incorporating specific food types. These food type prediction will help the food bank make distribution and inventory decisions. Our ensemble model provides equal weights to all forecasting models we use. In the future, the weight can be given based on the accuracy of each model and updated each iteration by using deep learning algorithms. We also note that an exhaustive search of all attribute combinations were not considered, which could be explored in the future.

Acknowledgements We want to thank the Food Bank of Central and Eastern North Carolina for sharing data and continuous information for this research. This research is supported by NSF Partnerships for Innovation Project Flexible, Equitable, Efficient, and Effective Distribution (FEEED) (Award No. IIP-1718672).

References

Ahire, S. L., & Pekgün, P. (2018). Harvest hope food bank optimizes its promotional strategy to raise donations using integer programming. *Interfaces*, *48*, 291–306.

Alikhani, R., Torabi, S. A., & Altay, N. (2019). Strategic supplier selection under sustainability and risk criteria. *International Journal of Production Economics*, *208*, 69–82.

Altay, N., & Narayanan, A. (2020). Forecasting in humanitarian operations: Literature review and research needs. *International Journal of Forecasting*.

Altay, N., & Pal, R. (2014). Information diffusion among agents: Implications for humanitarian operations. *Production and Operations Management*, *23*, 1015–1027.
Ashoori, M., & Taheri, Z. (2013). Using clustering methods for identifying blood donors behavior. In *5th Iranian conference on electrical and electronics engineering (ICEEE2013)* (pp. 4055–4057).
Azadnia, A. H., Saman, M. Z. M., & Wong, K. Y. (2015). Sustainable supplier selection and order lot-sizing: An integrated multi-objective decision-making process. *International Journal of Production Research*, *53*, 383–408.
Bai, J., Ghysels, E., & Wright, J. H. (2013). State space models and MIDAS regressions. *Econometric Reviews*, *32*, 779–813.
Bates, J. M., & Granger, C. W. (1969). The combination of forecasts. *Journal of the Operational Research Society*, *20*, 451–468.
Bensaou, M. (1999). Portfolios of buyer-supplier relationships. *MIT Sloan Management Review*, *40*, 35.
Benyoucef, L., Ding, H., & Xie, X. (2003). Supplier selection problem: selection criteria and methods.
Bianchi, G., Bruni, R., Nucara, R., & Reale, A. (2005). Data clustering for improving the selection of donor for data imputation. In *Fifth conference on classification and data analysis (CLADAG), June 6–8*.
Brock, L. G., & I. I. I., & Davis, L. B., (2015). Estimating available supermarket commodities for food bank collection in the absence of information. *Expert Systems with Applications*, *42*, 3450–3461.
Chester County Food Bank. (n.d.). Your dollar + our buying power = a winning combo. https://chestercountyfoodbank.org/dollar-buying-power-winning-combo/.
Chattanooga Area Food Bank. (n.d.). How we source food. http://www.chattfoodbank.org/programs/sourcing.
Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, *5*, 559–583.
Complex emergencies - ifrc. Retrieved July 20, 2020, from https://www.ifrc.org/en/what-we-do/disaster-management/about-disasters/definition-of-hazard/complex-emergencies/.
Consumerbrochure.pdf. Retrieved January 19, 2021, from https://fns-prod.azureedge.net/sites/default/files/ConsumerBrochure.pdf.
Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*, 273–297.
Crow, L. E., Olshavsky, R. W., & Summers, J. O. (1980). Industrial buyers' choice strategies: A protocol analysis. *Journal of Marketing Research*, *17*, 34–44.
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.
Dickson, G. W. (1966). An analysis of vendor selection systems and decisions. *Journal of Purchasing*, *2*, 5–17.
Dunham, M. H. (2006). *Data mining: Introductory and advanced topics*. New Delhi: Pearson Education.
Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods*. Oxford: Oxford University Press.
Dweiri, F., Kumar, S., Khan, S. A., & Jain, V. (2016). Designing an integrated AHP based decision support system for supplier selection in automotive industry. *Expert Systems with Applications*, *62*, 273–283.
Everitt, M. (2012, December 10) A food banker explains how food banks work—The billfold. https://www.thebillfold.com/2012/12/a-food-banker-explains-how-food-banks-work/.
Fahim, A., Salem, A., Torkey, F. A., & Ramadan, M. (2006). An efficient enhanced k-means clustering algorithm. *Journal of Zhejiang University-Science A*, *7*, 1626–1633.
Feeding America. [n.d.]. Hunger in America. https://www.feedingamerica.org/hunger-in-america/north-carolina.
Gardner, E. S, Jr. (1985). Exponential smoothing: The state of the art. *Journal of Forecasting*, *4*, 1–28.
Goebel, P., Reuter, C., Pibernik, R., & Sichtmann, C. (2012). The influence of ethical culture on supplier selection in the context of sustainable sourcing. *International Journal of Production Economics*, *140*, 7–17.
Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, *27*, 857–871.
Gualandris, J., Klassen, R. D., Vachon, S., & Kalchschmidt, M. (2015). Sustainable evaluation and verification in supply chains: Aligning and leveraging accountability to stakeholders. *Journal of Operations Management*, *38*, 1–13.
Haggett, P., & Cliff, A. (2005). *Modeling diffusion processes, encyclopaedia of social measurement* (pp. 709–724). Richardson: Elsevier.
Halley, A. (2000). La contribution des fournisseurs privilégiés à l'intégration de la chaîne logistique: mythe ou réalité? *Logistique & Management*, *8*, 89–100.
Han, J., Pei, J., & Kamber, M. (2011). *Data mining: Concepts and techniques*. Amsterdam: Elsevier.
Hansen, K. (2001). Purchasing decision behaviour by Chinese supermarkets. *The International Review of Retail, Distribution and Consumer Research*, *11*, 159–175.

Holguín-Veras, J., Jaller, M., Van Wassenhove, L. N., Pérez, N., & Wachtendorf, T. (2012). On the unique features of post-disaster humanitarian logistics. *Journal of Operations Management*, *30*, 494–506.
Hyndman, R. J., Athanasopoulos, G. (2018). *Forecasting: Principles and practice*. OTexts.
Ishtiaq, P., Khan, S. A., & Haq, M.-U. (2018). A multi-criteria decision-making approach to rank supplier selection criteria for hospital waste management: A case from Pakistan. *Waste Management & Research*, *36*, 386–394.
Jensen, L. -M. (2010). Humanitarian cluster leads as fourth-party logistics providers. In *International Heinz Nixdorf symposium* (pp. 372–383). Springer.
Kannan, V. R., & Tan, K. C. (2002). Supplier selection and assessment: Their impact on business performance. *Journal of Supply Chain Management*, *38*, 11–21.
Kim, M., Choi, W., Jeon, Y., & Liu, L. (2019). A hybrid neural network model for power demand forecasting. *Energies*, *12*, 931.
Korolov, R., Peabody, J., Lavoie, A., Das, S., Magdon-Ismail, M., & Wallace, W. (2016). Predicting charitable donations using social media. *Social Network Analysis and Mining*, *6*, 31.
Kowalczyk, A. (2017). *Support vector machines succinctly*. Research Triangle: Syncfusion Inc.
Kulkarni, A., Jayaraman, V. K., & Kulkarni, B. D. (2004). Support vector classification with parameter tuning assisted by agent-based technique. *Computers & Chemical Engineering*, *28*, 311–318.
Larssen, T., Høgåsen, T., & Cosby, B. J. (2007). Impact of time series data on calibration and prediction uncertainty for a deterministic hydrogeochemical model. *Ecological Modelling*, *207*, 22–33.
Law, R., Li, G., Fong, D. K. C., & Han, X. (2019). Tourism demand forecasting: A deep learning approach. *Annals of Tourism Research*, *75*, 410–423.
Lehmann, D. R., & O'shaughnessy, J., (1982). Decision criteria used in buying different categories of products. *Journal of Purchasing and Materials Management*, *18*, 9–14.
Lima, F. S., de Oliveira, D., & Gonçalves, M. B. (2014). Methodology for clustering cities affected by natural disasters. In: New perspectives in information systems and technologies (Vol. 1, pp. 97–106) Springer.
Lin, P.-C., & Wu, L.-S. (2011). How supermarket chains in Taiwan select suppliers of fresh fruit and vegetables via direct purchasing. *The Service Industries Journal*, *31*, 1237–1255.
Luthra, S., Govindan, K., Kannan, D., Mangla, S. K., & Garg, C. P. (2017). An integrated framework for sustainable supplier selection and evaluation in supply chains. *Journal of Cleaner Production*, *140*, 1686–1698.
MacQueen, J., et al. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, pp. 281–297). Oakland, CA, USA.
Makhitha, K. (2017). Supplier selection criteria used by independent retailers in Johannesburg, South Africa. *Journal of Business and Retail Management Research*, *11*.
Makhitha, K. M. et al. (2013). *An investigation into buyer behaviour of craft retailers in South Africa*, Ph.D. thesis, University of Pretoria.
Morello, P. (2020, December 9). How food banks and food pantries get their food. Feeding America. from https://www.feedingamerica.org/hunger-blog/how-food-banks-and-food-pantries-get-their-food.
Mukherjee, K. (2016). Supplier selection criteria and methods: Past, present and future. *International Journal of Operational Research*, *27*, 356–373.
Nahmias, S., & Cheng, Y. (2001). *Production and operations analysis* (Vol. 6). New York: McGraw-hill.
Nair, A., Jayaram, J., & Das, A. (2015). Strategic purchasing participation, supplier selection, supplier evaluation and purchasing performance. *International Journal of Production Research*, *53*, 6263–6278.
Nair, D. J., Rashidi, T. H., & Dixit, V. V. (2017). Estimating surplus food supply for food rescue and delivery operations. *Socio-Economic Planning Sciences*, *57*, 73–83. https://doi.org/10.1016/j.seps.2016.09.004.
Naude, M. (2013). A perspective on the criteria used by SMEs in Pietermaritzburg to select suppliers. *Journal of Contemporary Management*, *10*, 1–19.
Noble, W. S. (2006). What is a support vector machine? *Nature Biotechnology*, *24*, 1565–1567.
Oyelade, O., Oladipupo, O. O., & Obagbuwa, I. (2010). *Application of k means clustering algorithm for prediction of students academic performance*. arXiv:1002.2425.
Özpolat, K., Rilling, J., Altay, N., & Chavez, E. (2015). Engaging donors in smart compassion: Usaid cidi's greatest good donation calculator. *Journal of Humanitarian Logistics and Supply Chain Management*, *5*, 95.
Park, H.-S., & Jun, C.-H. (2009). A simple and fast algorithm for k-medoids clustering. *Expert Systems with Applications*, *36*, 3336–3341.
Paul, S., & Davis, L. (2019). Predicting food donor contribution behavior using support vector regression. In: Romeijn, H. E., Schaefer, A. & Thomas, R. (Eds.), *Proceedings of the 2019 IISE annual conference* (pp. 390–395). Orlando: IISE.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in python. *The Journal of Machine Learning Research*, *12*, 2825–2830.
Pegels, C. C. (1969). Exponential forecasting: Some new variations. *Management Science*, *15*, 311–315.
Pugh, N., & Davis, L. B. (2017). Forecast and analysis of food donations using support vector regression. In *2017 IEEE international conference on big data (Big Data)* (pp. 3261–3267). IEEE.
Rao, N. K., & Reddy, G. S. (2012). Discovery of preliminary centroids using improved k-means clustering algorithm. *International Journal of Computer Science and Information Technologies*, *3*, 4558–1561.
Reynolds, A. P., Richards, G., de la Iglesia, B., & Rayward-Smith, V. J. (2006). Clustering rules: A comparison of partitioning and hierarchical clustering algorithms. *Journal of Mathematical Modelling and Algorithms*, *5*, 475–504.
Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65.
Santos Lima, F., de Oliveira, D., Buss Gonçalves, M., & Samed, Marcondes Altimari. (2014). Humanitarian logistics: A clustering methodology for assisting humanitarian operations. *Journal of Technology Management & Innovation*, *9*, 86–97.
Sarhani, M., & El Afia, A. (2014). Intelligent system based support vector regression for supply chain demand forecasting. In: *2014 second world conference on complex systems (WCCS)* (pp. 79–83). IEEE.
Sarkis, J., & Talluri, S. (2002). A model for strategic supplier selection. *Journal of Supply Chain Management*, *38*, 18–28.
Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., De Leeuw, C. A., et al. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature Genetics*, *50*, 912–919.
Segal, M. N. (1989). Implications of single vs. multiple buying sources. *Industrial Marketing Management*, *18*, 163–178.
Silhouette (clustering)—wikipedia. Retrieved July 27, 2020, from https://en.wikipedia.org/wiki/Silhouette_(clustering).
Singh, N., Olasky, S. J., Cluff, K. S., & Welch Jr., W. F. (2006). Supply chain demand forecasting and planning. US Patent 7080026.
Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, *14*, 199–222.
Sternquist, B., & Chen, Z. (2006). Food retail buyer behaviour in the People's Republic of China: A grounded theory model. *Qualitative Market Research: An International Journal*, *9*, 243–265.
Swift, C. O. (1995). Preferences for single sourcing and supplier selection criteria. *Journal of Business Research*, *32*, 105–111.
Syntetos, A. A., Babai, Z., Boylan, J. E., Kolassa, S., & Nikolopoulos, K. (2016). Supply chain forecasting: Theory, practice, their gap and the future. *European Journal of Operational Research*, *252*, 1–26.
Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics* (Vol. 5). Boston: Pearson.
Tan, P.-N., Steinbach, M., & Kumar, V. (2016). *Introduction to data mining*. New Delhi: Pearson Education.
Tatham, P., Pettit, S., Jahre, M., & Jensen, L.-M. (2010). Coordination in humanitarian logistics through clusters. *International Journal of Physical Distribution & Logistics Management*, *40*, 657–674.
Tavana, M., Fallahpour, A., Di Caprio, D., & Santos-Arteaga, F. J. (2016). A hybrid intelligent fuzzy predictive model with simulation for supplier evaluation and selection. *Expert Systems with Applications*, *61*, 129–144.
Taylor, J. W. (2003). Exponential smoothing with a damped multiplicative trend. *International Journal of Forecasting*, *19*, 715–725.
Thome, K., Meade, B., Daugherty, K., & Christensen, C. (June 2018). International food security assessment, 2018–2028, GFA-29, U.S. Department of Agriculture, Economic Research Service.
Thompson, K. N. (1990). Vendor profile analysis. *Journal of Purchasing and Materials Management, 26*, 11–18.
Ülkü, M. A., Bell, K. M., & Wilson, S. G. (2015). Modeling the impact of donor behavior on humanitarian aid operations. *Annals of Operations Research*, *230*, 153–168.
USDA Economic Research Service. (2020a). Definitions of Food Security. https://www.ers.usda.gov/topics/food-nutrition-assistance/food-security-in-the-us/definitions-of-food-security.aspx.
USDA Economic Research Service. (2020b). Key Statistics and Graphics. https://www.ers.usda.gov/topics/food-nutrition-assistance/food-security-in-the-us/key-statistics-graphics.aspx.
Van der Laan, E., van Dalen, J., Rohrmoser, M., & Simpson, R. (2016). Demand forecasting and order planning for humanitarian logistics: An empirical assessment. *Journal of Operations Management*, *45*, 114–122.
Vapnik, V. (2013). *The nature of statistical learning theory*. Berlin: Springer.
Venkateswarlu, B., & Raju, P. G. (2013). Mine blood donors information through improved k-means clustering. arXiv:1309.2597.

Verma, R., & Pullman, M. E. (1998). An analysis of the supplier selection process. *Omega*, *26*, 739–750.
Vinkenoog, M., Janssen, M., & van Leeuwen, M. (2019). Challenges and limitations in clustering blood donor hemoglobin trajectories. In *International workshop on advanced analysis and learning on temporal data* (pp. 72–84). Springer.
Wang, Z.-X. (2013). A genetic algorithm-based grey method for forecasting food demand after snow disasters: An empirical study. *Natural Hazards*, *68*, 675–686.
Weber, C. A., Current, J. R., & Benton, W. (1991). Vendor selection criteria and methods. *European Journal of Operational Research*, *50*, 2–18.
Zsidisin, G. A. (2003). Managerial perceptions of supply risk. *Journal of Supply Chain Management*, *39*, 14–26.