

Building a Prediction Model for Forecasting Adult Care Facility Quarterly Patient Demand

Sudhan Bhattarai¹, Yaneth Correa-Martinez², Ebisa Wollega¹ and Leonardo Bedoya-Valencia¹

¹Department of Engineering
Colorado State University - Pueblo
Pueblo, CO 81001, USA

s.bhattarai@pack.csupueblo.edu, ebisa.wollega@csupueblo.edu,
l.bedoyavalencia@csupueblo.edu

²Hasan School of Business
Colorado State University - Pueblo
Pueblo, CO 81001, USA
yaneth.correa@csupueblo.edu

Abstract

An Adult Care Facility (ACF) is a healthcare organization providing regular non-medical services to the disabled elderly people. The number of ACF and elderly care homes are rising in US. Forecasting the number of people in a facility based on other factors can be very useful for planning and scheduling resources. In this paper, machine learning models are developed with the purpose of predicting the total number of patients admitted in an ACF at the end of each quarter. In particular, this paper proposes two models: Linear Regression (LR) and Deep Neural Network (DNN). Both models were used to fit the quarterly data obtained from multiple ACFs. The performance of the models was evaluated by using the known R-squared score. Based on R-squared scores on training, validation, and testing, the LR model outperformed the complex DNN model as the best prediction model.

Keywords

Adult Care Facility, Demand Forecasting, Linear Regression, Deep Neural Networks.

1. Introduction

The Adult Care Facility (ACF) is a particular type of health care organization providing regular non-medical service to elderly patients. A group of elderly patients, those living independently but who are unable to take care of themselves choose to stay at the ACFs for regular care (Adult Care Facilities, 2018). The Adult Care, also known as assisted living facility is a popular health care model not only in the US but also around the world as there were 15,800 facilities providing services to elder and disabled in 2015 just in the USA (statista.com, 2016).

Similarly to many health care activities, having a good estimation of the demand for care plays an important role in the allocating of resources in adult care facilities. Proper scheduling of personnel such as physicians, physicians assistants, nurses, and any other type of health care providers can be fundamental to minimize the cost of care and maximize the quality of care provided. In this paper, two machine learning models are applied to predict the demand for care measured as the number of patients admitted each quarter in AFCs.

The rest of this paper is organized as follows: a literature review is presented in section 2, section 3 describes the data used to train, validate, and test the proposed models. After that, section 4 describes the implementation of the models and section 5 presents the results and compares the proposed models. Finally, section 6 discusses the conclusions and further research.

2. Literature Review

Recently, machine learning models such as LR and DNN have gained popularity for forecasting purposes. Energy planning is one of the main areas where these techniques have been used. Antonopoulos et al. (2020) performed a detailed review on the uses of artificial intelligence, neural networks, and machine learning in energy planning. They showed a big increase in the use of these models from 2015 to 2019. Ahmad and Chen (2020) experimented with different models to forecast electricity consumption. They compared the accuracy in the forecast of different structures for neural networks. They found that the backpropagation models performed really well for real-time forecast.

In the hotel industry, Sánchez-Medina and Sánchez (2020) used machine learning models to predict hotel booking cancellations. They used 13 independent variables to predict the most likely customer to cancel their reservation by using a neural network optimized by a genetic algorithm. Their model performed well reaching an accuracy of 98% allowing the user to identify particular attributes of the customers likely to cancel.

In supply chain management, forecasting is essential for managing, planning, operating, and optimizing the use of resources. In this area, novel models for forecasting have been abundantly used in order to forecast demand for products and services. Recently, Al Hajj Hassan et al. (2020) combined time series with machine learning models in order to reduce the prediction error when forecasting the demand for container shipment from a US intermodal company. Their model outperformed typical time series models used by the company when producing weekly and monthly forecasts. In the area of sustainable agriculture supply chain management (ASCM), Sharma et al. (2020) proposed a Machine Learning (ML) applications framework that identified the role of ML algorithms in providing real-time analytic insights for pro-active data-driven decision making in ASCM.

In any scheduling problem, predicting the demand as accurate as possible can facilitate finding an optimal or good resource allocation in order to meet a specific objective. An accurate forecast of patient visits and services requested in healthcare facilities is one of the key challenges for health care policy makers to better allocate medical resources and service providers.

Regarding health care applications, Hu et al. (2016) proposed a hybrid model intended to forecast patients' visits to emergency departments (EDs). They combined an autoregressive integrated moving average–linear regression (ARIMA–LR) model with an artificial neural network. Their hybrid model outperformed existing models in terms of forecasting accuracy when using real data sets from local EDs. Bacchi et al. (2020) developed a pilot study to show that machine learning, using natural language processing, may be able to assist with accurate length of stay (LOS) and discharge destination prediction for general medical hospital inpatients. They developed an artificial neural network capable of predict with about 80% accuracy the number of days a patient will stay in the hospital. They concluded that further research with larger, more detailed datasets from multiple hospitals is required to optimize and examine the accuracy that may be achieved with such predictions. Finally, Karnuta et al. (2020) developed a naïve Bayes machine learning model used to predict the payments, length of stay (LOS), and discharge disposition (next stage in the medical procedure) of patients following a dorsal and lumbar fusion. They also used the model to develop a risk-stratified payment scheme. They were able to show that their naïve Bayes machine learning model has good-to-excellent reliability and responsiveness for cost, LOS, and discharge disposition.

Based on the above literature review, up to now, there seems not to be any application of either LR or DNN for forecasting the number of patients requesting health care in ACPs every quarter. This paper is an initial attempt to bring these novel models in the management of ACPs. Next, the data used to implement the proposed models are discussed in detail.

3. Data Description

The data required to train, validate, and test the proposed models was obtained from the website: Health Data New-York (DOH, 2020). This website provides information coming from different Adult Care Facilities ranging from 2013 to 2019. After performing some cleaning, 3,311 data points (records) remained usable for the purpose of this paper. Each record consists of 14 columns as shown in Table 1.

Table 1. Data Description

Column Name	Type of Variable	Data Type
Reporting Year	Not included in model	Object
Reporting Quarter	Not included in model	Object
County	Not included in model	Object
Reporting Organization	Not included in model	Object
Reporting Organization ID	Not included in model	Object
Total Capacity (Certified)	Input (Predictor)	Quantitative
Beginning Census	Input (Predictor)	Quantitative
End Census	Output (Response)	Quantitative
Male Census	Input (Predictor)	Quantitative
Female Census	Input (Predictor)	Quantitative
# of Residents Age 66-80	Input (Predictor)	Quantitative
# of Residents Age 81+	Input (Predictor)	Quantitative
Total Admissions	Input (Predictor)	Quantitative
Total Discharges	Input (Predictor)	Quantitative

Reporting Year, Quarter, County, Organization and Organization ID show the general information about each ACF. Total capacity is the capacity of facility measured in the number of patients. Beginning census and end census show the quarterly starting and ending number of patients in the facility. Male and Female census show the number of male and female patients in the facility for that quarter. Number of residents aged 66-80 and 81+ indicate the number of patients on each age group. Total admissions and discharges indicate the number of patients admitted and discharged during each quarter. The next section presents proposed methodology.

4. Models Implementation

Two different prediction models are used to forecast the number of ACFs' patients at the end of each quarter as a measure of the demand for care in those facilities. Eight numeric variables were used as the input predictors or independent variables. There is a single numeric variable named 'End Census' used as the dependent variable which acts as the response in both models. As mentioned in the previous section, 3,311 records will be used to train, validate, and test the proposed models. The problem is then to build a prediction model such that the model will take 8 independent variables for a specific quarter in order to predict the response variable 'End Census' for that quarter.

The first proposed model is a linear regression model. In a typical linear prediction model, each independent variable is included in a linear model intended to estimate the dependent variable. Each independent variable has its respective weight. In this paper, a linear regression model with 8 independent variables (predictors) and 1 dependent variable (response) is proposed. Equation 1 shows the general linear model used:

$$\hat{y} = W^T X,$$

where

W = A weights vector, $[W_0, W_1, W_2, W_3, W_4, W_5, W_6, W_7, W_8]$ where, W_0 is a dummy weight.

X = Input (predictor) vector, $[X_0, X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8]$, X_0 is unity and \hat{y} is the response.

In this first proposed model, the linear regression the problem is to find W such as the value of \hat{y} is most accurate.

The second proposed model is a Deep Neural Network (DNN) model. A DNN is a sophisticated prediction model in which there are multiple hidden layers for activation between input and outputs. In this paper, 8 input variables will be used. In a DNN model, the first step is to initialize multiple linear solutions with different rates for the first hidden layer. Then, the DNN model connects the output of first layer to subsequent layers with different rates. Finally, from the last hidden layer, the DNN model obtains the one single output. For a linear prediction, we choose a linear activation function between hidden layers to produce different solutions on each layer and to combine them all together to obtain an output for each layer. Figure 1 shows a generic DNN model with three input variables, one response, and two hidden layers.

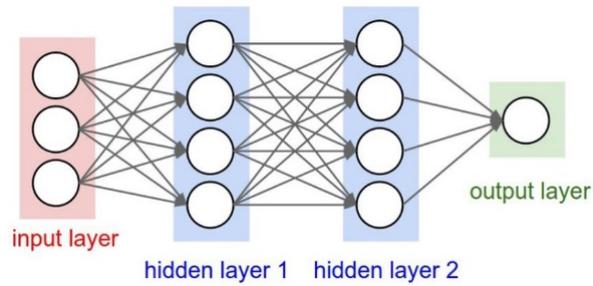


Figure 1. Generic Representation of a DNN (i2tutorials.com, 2019)

The LR uses a Gradient Descent method in order to calculate the most appropriate weights in each model by gradually decreasing weights to minimize error by using the equation presented below. In the gradient descent method, the partial derivative of cost function w.r.t the weight is calculated first and then subsequently moves in the direction of improvement from initial random value to momentum (Goh, 2017).

$w_j \leftarrow w_j - \alpha * \frac{J(W)}{\partial w_j}$, w_j : weight corresponding to j^{th} independent variable, α = learning rate. The gradient (the partial derivative of the cost function) is calculated first and then subsequently subtracted from previous weight to get minimum weight.

In this paper, the Mean Squared Error (MSE) is used to evaluate the loss function (average squared of difference between real output and predicted output). The Gradient of MSE gives minimum weight when the gradient is zero. With this method, the weights are gradually decreased step by step by subtracting a cost function gradient multiplied by learning rate. The cost function is given by the equation presented below.

$$J(\mathbf{w}) = \frac{1}{m} \sum_{i=1}^m (\hat{y}_i - y_i)^2, \text{ where } m = \text{number of training examples, } y_i = \text{actual output, and } \hat{y}_i = \text{predicted output.}$$

The solution method of the Deep Neural Network (DNN) is also similar to the one used for the linear regression model. But, in DNN since all the weights are not directly connected to the output, back-propagation approach using the chain rule of derivatives is applied to compute the gradient in every hidden layer. As the classic gradient descent method is prone to getting stuck in a local minima, the adaptive gradient descent is chosen in this paper. Unlike the classic gradient descent for which the learning rate is to the same for all the weights, in adaptive gradient descent, the learning rate is different for every weight. The algorithm adopts to lower learning rate for frequently utilized feature while to higher learning rate for less frequent features (Ruder, 2016). In order to use this model, the 3,311 data points were split into 80% for training purposes and 20% for testing the model.

5. Analysis of Results and Comparison of the Models

The data are split into 80% for training and 20% is left out for testing purpose to run experiments. The 80% is also further divided into four folds to study the behavior of the model on the training data set. The experiments are run in Python 3 programming environment by using Scikit-learn (to estimate R-squared and MSE for the linear regression model) and TensorFlow (to evaluate DNN). Figures 2 and 3 show the predictive accuracy of the linear regression model. It can be seen from Figure 2 that the linear regression model provides fairly stable results at about 99% across all the four folds, particularly for one predictor, male census. The weights that are estimated by the 80% of the data also give a similar accuracy level of 99% when tested on the 20% dataset as can be seen in Figure 3. The weights used to estimate the test error are given in Table 2.

cross validation rsquare score:
 [0.99471085 0.99655494 0.99606465 0.99575197]
 actual output: [96 104 179 ... 134 67 131]
 predictions: [95 103 178 ... 133 64 131]
 train rquare score: 0.9958949317991365

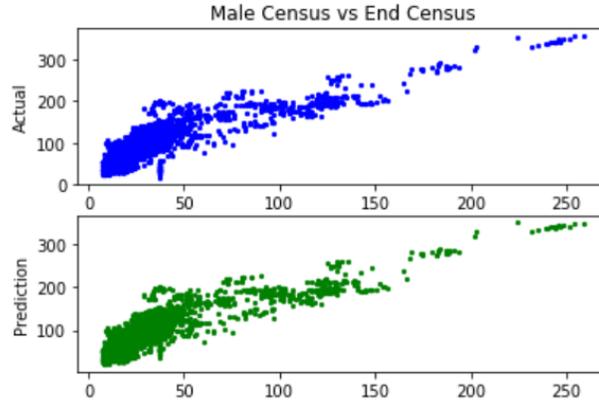


Figure 2. LR training result

test score: 0.9928975698581191
 actual output: [47 251 34 114 168] ...
 predictions: [45 249 31 112 169] ...

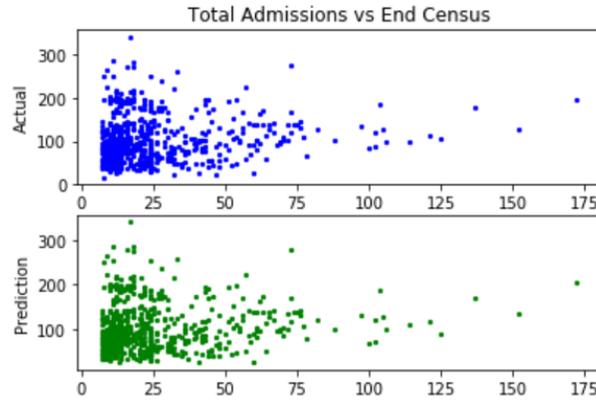


Figure 3. LR testing result

Table 2. LR Model Coefficients

Predictor	Coefficient
Intercept	-2.52
Total Capacity (Certified)	0.02
Beginning Census	0.29
Male Census	0.66
Female Census	0.68
Number of Residents Age 66-80	0.05
Number of Residents Age 81+	0.03
Total Admissions	0.10
Total Discharges	-0.10

The DNN's set of experiments is run using Adam's optimizer, learning rate of 0.01, activation function of rectified linear unit (ReLU) for hidden layers, 10 batches, and 100 epochs. The DNN is trained on 2 layers and 3 layers with different hyper parameters. Mean squared error and mean absolute error (MAE) are used to evaluate the accuracy of the model. Each experiment is terminated if there is no improvement for five consecutive epochs.

2-Layer DNN: A 2-layer DNN is fitted with 128 neurons in first layer following 20% dropout and 16 neurons in second layer following 20% dropout to omit overfitting. Figure 4 shows the results. The left-hand side shows the predicted versus the true value, where a complete overlap of the blue and green dots imply a perfect prediction. The right-hand side shows the mean absolute deviation and the MSE loss comparisons. The top orange lines show the test errors while the bottom blue lines show the training errors.

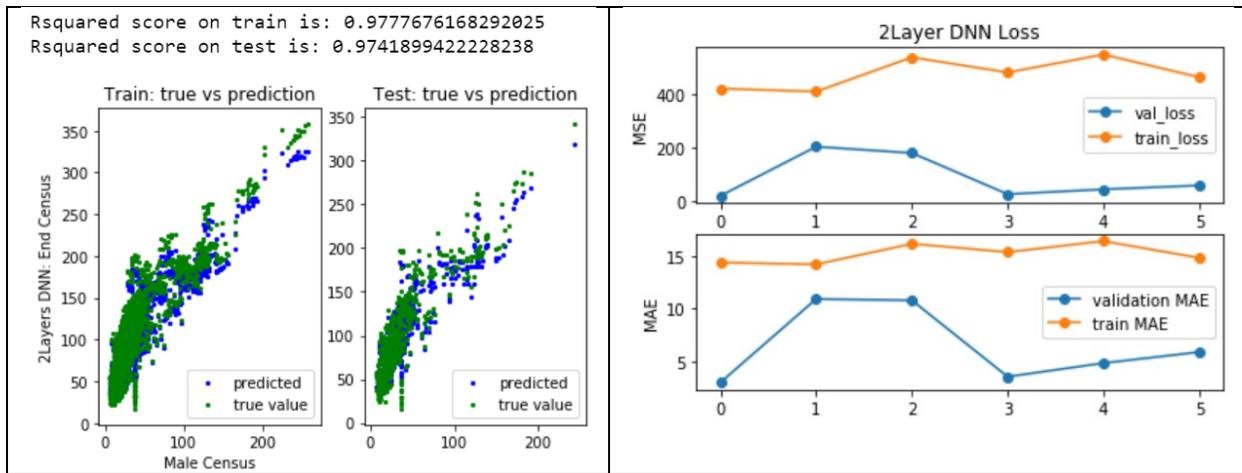


Figure 4. 2-Layer DNN results

3-Layer DNN: A 3-layer DNN is fitted with 128 neurons in first layer following 20% dropout, 32 neurons in second and 8 neurons in third layers following 20% dropout in each layer to omit overfitting. Figure 5 below shows the results. It can be seen that both the R-squared values for the training data and the testing data deteriorated as the number of layers increases from 2 to 3.

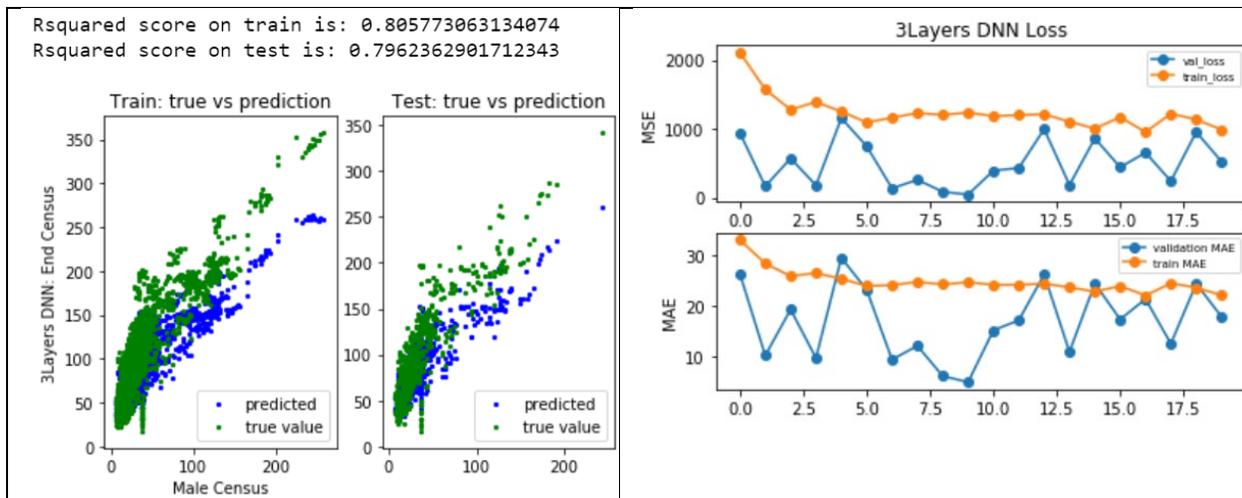


Figure 5. 3-Layer DNN results

Further experiments are run on multiple optimizers as shown in Tables 3 and 4 below. The DNN algorithm is executed for 2 layers and 3 layers networks. For the 2-layer network, 256 neurons are used in the first hidden layer

and 128 neurons in the second hidden layer, each layer followed a 20% dropout to avoid overfitting. Also, for the 3 layers network, 512 neurons are used in the first, 256 neurons in the second, and 128 neurons in the third hidden layer each followed by a 20% dropout layers. The output neuron is a single unit neuron. The rectified linear unit (ReLU) and hyperbolic tangent (Tanh) are used as the activation functions in hidden layers for both DNN models one after another. The activation function for the output layer is always linear. Adagrad and Adam are used as the optimizers one after another for the same network setup with learning rate of 0.0001 for each setup. Ten data points are used as the batch size which represents the number of samples to process in one run, and 100 epochs which represent number of times the whole dataset is processed. Running 100 epochs is not necessary as a reasonable solution may be obtained in fewer epochs. Hence, a validation loss callback function is used to terminate the algorithm once the loss function is not improved in 5 successive epochs and returns the result. The R-squared, MSE and MAE metrics are recorded for all the neural networks with different hyper-parameters setup. The last five columns of Table 3 and the last three columns of Table 4 portray the experimental values of these metrics for the training data and testing data, respectively. For the underlying optimizers, model, and activation functions, it can be seen that the linear activation function over performs the remaining activation functions for both the training and testing data sets.

Table 3. Training and Validation losses and scores

Optimizer	Model	Activation function	Epoch	Train MSE	Train MAE	Val MSE	Val MAE	R2 Train
GD	Least Square	linear	-	-	-	11.16	1.75	0.996
Adam	DNN 2 Hidden Layers	ReLU	5	91.43	6.91	43.67	5.89	0.969
		linear	20	105.20	7.59	17.01	2.64	0.992
		tanh	58	299.57	10.82	122.98	4.53	0.940
	DNN 3 Hidden Layers	ReLU	29	116.14	7.75	264.58	14.38	0.946
		linear	21	123.64	8.30	18.25	2.61	0.993
		tanh	27	230.98	10.72	65.63	3.50	0.970
Adagrad	DNN 2 Hidden Layers	ReLU	52	213.87	10.69	44.57	4.49	0.985
		linear	26	331.18	13.29	20.99	2.89	0.991
	DNN 3 Hidden Layers	ReLU	59	243.07	11.14	49.77	5.43	0.982
		linear	63	353.63	13.72	15.87	2.55	0.994

Table 4. Testing losses and scores

Optimizer	Model	Activation function	Epoch	Test MSE	Test MAE	R2 Test
GD	Least Square	linear	-	18.64	2.03	0.993
Adam	DNN 2 Hidden Layers	ReLU	5	47.03	5.35	0.966
		linear	20	29.61	3.40	0.987
		tanh	58	105.23	4.68	0.960
	DNN 3 Hidden Layers	ReLU	29	162.99	11.11	0.941
		linear	21	36.10	4.05	0.989
		tanh	27	44.23	3.23	0.983
Adagrad	DNN 2 Hidden Layers	ReLU	52	46.17	4.56	0.982
		linear	26	34.09	3.46	0.987
	DNN 3 Hidden Layers	ReLU	59	57.96	5.92	0.978
		linear	63	27.17	2.82	0.990

The r-squared scores achieved for this problem are very close for the LR and the smaller size DNN (2 hidden layers). However, the scores get worse as the number of layers is increased. This happens mainly due to the under-fitting and over-fitting of data. Depending upon the size and linearity of the data, single neuron DNN, in other words a linear regression model, seems to be performing better than complex multilayers and multi-neurons DNN. A DNN model consists of many parameters and hyper parameters which take numerous different values. Consequently, there is a lot of flexibility to modify our model and keep the track of good model. This property works better for larger and non-linear data sets.

6. Conclusions and Recommendations

An adult care facility is a health care organization providing ubiquitous non-medical services to disabled elderly. Adult care is a rising business and building a prediction model for this business is very useful to allocate scarce and expensive resources in order to minimize costs and maximize quality of care. Linear regression and a deep neural network models are implemented and evaluated to predict the number of patients admitted in a facility at the end of each quarter. Neural networks are very powerful on modeling big data but for problems with smaller data sets, the complex DNN might perform poorly because of overfitting problems. Based on the R-squared scores, mean square error, and mean absolute error, the linear regression model seems to perform better than the neural network.

References

- Adult Care Facilities. (2018, May). Retrieved from Health Data New York: https://www.health.ny.gov/facilities/adult_care/ accessed on June 15, 2020.
- statista.com. (2016, 12 20). Total number of assisted living facilities in the U.S. from 2006 to 2015. Retrieved from statista.com: <https://www.statista.com/statistics/323232/number-of-assisted-living-facilities-in-the-us/> accessed on June 15, 2020.
- Antonopoulos, I., Robu, V., Couraud, B., Desen, K., Norbu, S., Kiprakis, A., Flynn, D., Elizondo-Gonzalez, S., and Wattam, S., Artificial intelligence and machine learning approaches to energy demand-side response: A systematic review, *Renewable and Sustainable Energy Reviews*, September 2020, Vol. 130, 2020.
- Ahmad, T., and Chen, H., A review on machine learning forecasting growth trends and their real-time applications in different energy systems, *Sustainable Cities and Society*, March 2020, Vol. 54, 2020.
- Sánchez-Medina, A., and Sánchez, E., Using machine learning and big data for efficient forecasting of hotel booking cancellations, *International Journal of Hospitality Management*, August 2020, Vol. 89, 2020.
- Al Hajj Hassan, L., Mahmassani, H., and Chen, Y., Reinforcement learning framework for freight demand forecasting to support operational planning decisions, *Transportation Research Part E: Logistics and Transportation Review*, May 2020, Vol. 137, 2020.
- Sharma, R., Kamble, S., Gunasekaran, A., Kumar, V., and Kumar, A., A systematic literature review on machine learning applications for sustainable agriculture supply chain performance, *Computers & Operations Research*, July 2020, Vol. 119, 2020.
- Xu, Q., Tsui, K-L., Jiang, W., and Guo, H., A Hybrid Approach for Forecasting Patient Visits in Emergency Department, *Quality and Reliability Engineering International*, October 2016, Vol. 30, 2016.
- Bacchi, S., Gluck, S., Tan, Y., Chim, I., Cheng, J., Gilbert, T., Menon, D. K. Jannes, J., Kleinig, T., and Koblar, S., Prediction of general medical admission length of stay with natural language processing and deep learning: a pilot study, *Internal and Emergency Medicine*, January 2020.
- Karnuta, J. M., Golubovsky, J. L., Haeberle, H., Rajan, P. Navarro, S. Kamath, A., Schaffer, J., Krebs, V., Pelle, D., and Ramkumar, P. N., Can a machine learning model accurately predict patient resource utilization following lumbar spinal fusion?, *The Spine Journal*, March 2020, Vol. 20.
- DOH, O. D. (2020, 03 24). Adult Care Facility Quarterly Statistical Information Report: 2013-Present. Retrieved from Health Data New-York: <https://health.data.ny.gov/Health/Adult-Care-Facility-Quarterly-Statistical-Informat/h5s5-hcxg> accessed on June 15, 2020.
- Goh, G., Why momentum really works. *Distill*, 2017, Vol. 2.
- Ruder, S. (2016). An overview of gradient descent optimization algorithms. arXiv preprint arXiv, 1609.04747.
- statista.com. (2016, 12 20). Total number of assisted living facilities in the U.S. from 2006 to 2015. Retrieved from statista.com: <https://www.statista.com/statistics/323232/number-of-assisted-living-facilities-in-the-us/> accessed on June 15, 2020.

Biographies

Mr. Sudhan Bhattarai is a graduate student at Colorado State University-Pueblo with major in Master of Science in Industrial & Systems Engineering (MSISE). Sudhan's research interest is primarily in the fields of data science, optimization and artificial intelligence with the application in healthcare analytics. Sudhan's profile can be located at: <https://github.com/sudhan-bhattarai>.

Dr. Yaneth Cristina Correa-Martinez is an Assistant Professor of Production and Operations Management for the Hasan School of Business at Colorado State University-Pueblo. She earned her Ph.D. from Old Dominion University and her M.Sc. and her Bachelors in Engineering from the National University of Colombia. Her research interests are mainly in the Supply Chain Design and Management area. Her work has focused on how the structure and operational decisions in a supply chain network impact its response to disruptions. She also has research interests on supply chains of highly regulated goods and on healthcare operations management. She teaches statistics, supply chain management, and project management, among others. Professor Correa-Martinez is always looking for innovative ways to teach quantitative topics to business students. Finally, she is a proud mother of three and wife of one that enjoys hiking and camping in the National Parks and traveling the world with her family.

Dr. Ebisa Wollega is an Assistant Professor of engineering at Colorado State University Pueblo. His research areas include applying large-scale optimization methodologies and artificial intelligence algorithms in wide range of applications. His profile can be found at <https://www.csupueblo.edu/profile/ebisa-wollega/index.html>.

Dr. Leonardo Bedoya-Valencia is an associate professor of engineering at Colorado State University Pueblo. His research areas include applying simulation and optimization methodologies in wide range of applications including health care, energy planning, and manufacturing. His profile can be found at <https://www.csupueblo.edu/profile/leonardo-bedoya-valencia/index.html>.