# **Predicting Customer Spend in the Hospitality Group Procurement Industry**

**Nilabh Chaturvedi; Sarnav Chauhan; Supriya Malla; Payal Pandya; Tanveer Singh Sapra;**

**Dr. Yang Wang**

Purdue University, Department of Management, 403 W. State Street, West Lafayette, IN 47907

chaturvn@purdue.edu; chauha37@purdue.edu; malla2@purdue.edu; [pandyap@purdue.edu](mailto:pandyap@purdue.edu); tsapra@purdue.edu

## **ABSTRACT**

*This study aims to identify and predict the annual spending capacity of various hotels across USA and Canada. We analyzed spending patterns of 500 hotels, each belonging to 10 different market segments. Data across 6 years revealed that hotels spent the maximum within 5 product/service categories. Correlation analysis revealed that spending within a combination of a market segment and product category was influenced by unique variables and factors. For each set of this combination, we implemented a customized forecasting model, and accurately predicted a hotel’s annual expenditure across various product categories. Through feature selection and variable significance analysis, we were also able to identify key hotel characteristics that were strongly related to larger annual spending amounts.*

***Keywords:*** *Spend, prediction, machine learning models, feature importance, hospitality, hotels*

## **INTRODUCTION**

The hotel procurement industry serves a crucial role in the strategy and operations of a hotel. A hotel procurement agency partners with suppliers, negotiates purchasing deals, establishes service level agreements with hotels and offers discounts and incentives to client hotels. While many different types of companies use purchasing programs to maximize efficiencies and savings, the hospitality industry has been particularly successful at improving their bottom line using collective purchasing and negotiation. The revenue for the procurement industry is collected from their partner suppliers to whom the agencies guarantee valuable and profitable business. They also benefit from subscription plans and contracts that are signed with partnering hotels. It is crucial for the hospitality procurement agencies to ensure that partner hotels make the most purchases from their supplier network to guarantee maximum value to both suppliers and hotels.

This project was motivated by the need for a leading hospitality procurement agency in North America to analyze the expenditure that partner hotels make with their supplier networks. By analyzing spend patterns across 5 different product/service categories, the agency wished to accurately predict the future spend by these client hotels. The Total Potential Spend (TOPS) was calculated as the annual spend that a hotel makes with both agency and non-agency suppliers. The procurement wished to maximize their share in the TOPS of a hotel. Thus, they needed to identify future opportunities by studying historical spending patterns by hotels. The agency had been using a 20-year-old forecasting model to predict TOPS. Based on the data they previously used to train their model, they used only three variables to predict the spend, namely: if the hotel was a resort, if the hotel was an all-inclusive property, and the number of rooms in a hotel.

Further analysis of spending revealed that more variables influenced the spending capacity of hotels and that different variables influenced the spend on different products and service categories. The agency wanted to identify these variables and use a more sophisticated and accurate forecasting model to predict the TOPS of their client hotels. We built a forecasting model to accurately predict the TOPS for a given client hotel by implementing a unique model for each of the 30 combinations.

Each hotel was categorized into a distinct market segment. Additionally, 5 different product categories were identified that recorded the maximum spending. Correlation analysis on every combination of a market segment and product category revealed that the spend variable within each combination was influenced by a different set of variables. This led us to conclude that every combination of market segment and product category must have separate trained models.

By using the linear regression model used by the legacy model as a baseline, we were able to achieve 66% greater accuracy than the legacy model. We also extracted the important variables from each trained model and were able to identify those variables that contributed the most to the spend for a given combination of market segment and product/service category. We also referred to additional work done in the field of predictive modelling to learn about newer technologies and algorithms that we could implement. As most of our data points were categorical variables, we also implemented models that can predict more accurately and optimally.

We generated graphs on variable importance and concluded that important variables most impacted the spend for that combination. This information could be leveraged by the hospitality procurement agencies to better identify features of a hotel that contribute to the most amount of spend. It could also be used to identify potential clients and prospective partnerships with desirable features that would maximize the revenue of the hospitality procurement agencies. Our work is a novel endeavor as no work has been done in the field of predicting the spending patterns of a hotel by analyzing a unique set of variables for a combination of market segment and product/service category. Our unique implementation of variable importance extraction for each model also is a novel work in the field of hospitality.

The following paper is divided into comprehensive sections for easy understanding of our problem, statement, our work to address the business problem and our revelations and results. In Literature Review, we analyze the various works done in the field of hospitality procurement vis-a-vis predicting the spend of the customer hotel. We also analyze various could use to implement for our own business problem. In the Data section we discuss the three main data sources that we were given by the clients. We also discussed the type of variables that we worked with the challenges associated with it. In the Methodology section we discuss the entire workflow that we implemented to process the data and train a model to predict customer spend. In the Models section we speak about the many models that we used and their categories. We also discuss how each of these models are different and differently predict for customer spend. In the Results section we discuss our findings and statistical conclusions. Finally, in the Conclusion section we address our main business problems, the methods that we implemented, our unique findings and results, and how our work is a novel work in the field of hospitality procurement.

**LITERATURE REVIEW**

The hospitality industry is growing at an incredible speed across various geographies with a wide consumer base. This has led to the accumulation of large data sets and complex service networks. The Hospitality Procurement Industry is a subset of a larger supply-chain and procurement network. However, the industry has yet to explore the potential of Big Data and Machine Learning to aid decision making or to perform budget and revenue forecasting.[[1](https://docs.google.com/document/d/1-22CuQEeXGLp4tjEyki8FLUB5pcmQhWYFe3yOiMH010/edit#bookmark=id.lldnq8ydexb8)] In recent years considerable research and work has been done to integrate big data analytics to supply chain networks.[2] However the hospitality industry largely remains under-benefitted from these technologies. E-procurement is a growing need for the hospitality industry to integrate innovation and streamline hotel operations/functions. Studies performed by [3] have revealed the growing interest among managers and suppliers of hotels to integrate e-procurement to functions of the hospitality industry.

Our study aims at predicting the Total Opportunity Potential Spend (TOPS) at an annual level for a hotel. As stated earlier, limited work has been done to predict or forecast the annual spend by a hotel on supplies and services. Work done by Mathew et al [4] implements an LSTM model to forecast spend by a hotel. The LSTM algorithm had an architecture of 8 layers and 500 epochs. This forecast model could achieve an RMSE for the LSTM model was 0.0137, which was a significant improvement over the baseline RMSE of 9221.876. The model forecasted 6 months of spend and quantity of supplies for a hotel in UAE. The objectives of our work differs from the mentioned LSTM approach as we did not aim to forecast the spend of a hotel based on periodic trends alone. Additionally, this approach did not analyze the seasonal effects and descriptive attributes of the hotel in question. Our dataset contained detailed customer hotel attribute details which did not allow us to apply the LSTM approach. We also made separate models based on the spend entries for a given product/service category and market segment group. The partition of the dataset in these groups reduced the number of data-points needed to effectively train an LSTM. More detailed data collected over greater hotels and time periods may render using an LSTM feasible.

We were tasked to replace a 20-year-old legacy model that utilized multivariate regression [9] and poorly performed across various product categories. We therefore used the linear regression model as a baseline model to benchmark the performance of all the model that we would train on our dataset. This approach has been implemented by Bartol et al [10] where a multiple linear regression model is trained over public datasets used for body measurement estimation. The linear model is a simple and interpretable solution and was compared against the accuracy of the more sophisticated deep-learning models that were trained on similar datasets. It was noticed that the linear regression model performed as good as sophisticated state-of-the-art deep-learning models, or even better. They concluded that any model built for the estimation of body measurement should never perform worse than the benchmarked linear regression model.

To fit the best model for each combination of market segment and product category, we began exploring various models. Elastic Net [11] is a regularized regression model that performs feature selection by penalizing less significant features. In regularized regression, a tuning parameter controls the degree of shrinkage applied to the regression coefficients, and penalties that induce sparsity shrink many coefficients to exactly zero, performing in effect model selection. [12] Due to many categorical variables in our dataset, we attempted to use Elastic-Net to find the most important variables to best fit the spend for a particular combination.

Random Forest is a machine learning process that consists of several uncorrelated decision trees. All decision trees have grown under a certain type of randomization during the learning process. For a classification, each node in the structure is allowed to be decided and the class with the most votes decide the final classification (majority principle). Random Forests can also be used for regression analysis and problems [17].

Gradient boosting is a powerful machine-learning technique that achieves state-of-the-art results in a variety of practical tasks. For a number of years, it has remained the primary method for learning problems with heterogeneous features, noisy data, and complex dependencies: web search, recommendation systems, weather forecasting, and many others [13, 14, 15]. As the first gradient boosting algorithm, we implemented the XGBoost algorithm. The extreme gradient boosting was proposed by Friedman et al in 2001.[5] Our preliminary implementation of Decision Tree [6] for regression caused issues of overfitting due to limited data for each category. XGBoost could help in overcoming the overfitting of data by performing bagging and boosting. XGBoost has outperformed several Tree based algorithms with implementations of gradient tree boosting and gives 87% increased accuracy compared to other tree-based prediction algorithms.[7] The primary focus of boosting is to merge a set of weak learner to a strong one, in an iterative fashion. XGBoost is a widely used algorithm for many data science competitions.[8] Our dataset mostly comprised of categorical variables with 0 or 1 as values. These categorical variables were predicting spend which was an interval variable. We therefore needed a model that could best fit over categorical variables.

A series of experiments were conducted by Dorogush et al [16] and they concluded that CatBoost is a new gradient boosting algorithm that successfully handles categorical features and takes advantage of dealing with them during training as opposed to preprocessing time. Another advantage of the algorithm is that it uses a new schema for calculating leaf values when selecting the tree structure, which helps to reduce overfitting. As a result, the new algorithm outperforms the existing state-of-the-art implementations of gradient boosted decision trees (GBDTs) XGBoost.[5]

We also implemented an ensemble model to achieve the best outcome from the three tree-based algorithms. In an ensemble model, multiple models are combines and the average of their prediction is considered as the final prediction.[18] Ensemble learning is inspired from the principles of bagging or the boosting approach. In bagging, all the individual distinct models are built in parallel, and outputs of each model are considered independently. The output is then averaged by employing the majority voting method. However, the output from the ensemble model will not be optimized as a deviation among models will impact the average of results of models.[18]

Stacked generalization is a flexible method for multiple classifier systems in which the outputs of the base-level classifiers are viewed as data points in a new feature space and are used to train a combiner function [20]. Caruana et al. [21] evaluated stacked generalization with logistic regression with thousands of classifiers on binary classification problems, and reported that stacked generalization tended to overfit, resulting in poor overall performance. In this paper proposed by Reid et al [19], the overfitting of stacked generalization method has been remedied to improve overall generalization accuracy through regularization. They proposed that regularization can be performed by penalization of the L2 norm of the weights (Ridge regression), L1 norm of the weights (lasso regression) or a combination of the two (elastic net regression). L1 penalties yield sparse linear models; in stacked generalization, this means selecting from a small number of classifier posterior predictions.

## **DATA**

In this study we were provided with product and service line-item information from suppliers and the actual spend for each customer property. We had to gain an understanding of attributes associated with each property that is being serviced by our client. We also had access to the client’s original forecasting model and the predictions that it was generating.

We were given below data sources through our client’s PURS and eRM systems -

1. Customer File – This data source contained information regarding various attributes of customers’ properties. Examples of this included Qualitative factors like Customer Street Address, Customer State, etc. as well as Quantitative attributes like Number of rooms, whether the property has a restaurant or not etc.
2. Customer Spend File – This data source contained the spend captured for each customer. The information was on our client’s services/products on a monthly level.
3. Supplier Data – This contained information about all the suppliers that our client has and the product/service item they provide.
4. Product Service Mapping File – Provided us with a hierarchical relationship between the market program name, as the parent, and the product/service categories as its children.

Our objective was to model the customer’s total overall potential spend stratified by hospitality segment type and geography. To do that we first consolidated the information for customer spend which was available to us from 2017 onwards in multiple files, into a single source to bring them to a consistent structure that we could use further for evaluation.

We integrated the consolidated customer spend file with the supplier data and product service mapping to get spend information at the product/service category level. We included all three indicators from the previous model – Number of Rooms, All Inclusive, and Resort. We gradually added all the remaining binary variables (e.g.: Casino Flag, Restaurant flag, etc.) to our model to analyze the impact on spending.

For our modelling purposes, we choose to drop the descriptive variables like customer address, etc. since they would play no role in determining the customer spend. However, since we were interested in understanding the impact of location on the spend at a broader level, we created a dummy variable for Census location derived through the ZIP code of the property. Our dummy variable comprised of categories – Suburban, Urban, and Airport.

We additionally scraped Ratings and Number of reviews triggered through a Google Search Query with the customer’s name and location. We included weighted rating as a factor in our modelling to check the impact on TOPS.

## **METHODOLOGY**

The primary goal of this research is to predict the total overall potential spend made by the hospitality industry using the data from 2017 to 2022(excluding 2020). The models produced in this study have been used by many leading industries and are comprehensive and reliable.

**Exploratory Data Analysis:** After data collection, in order to clearly understand the nature of data, an exploratory data analysis was conducted. Firstly, data manipulation was performed to change the data type to correct formats as initial data did not have appropriate data types for respective variables.

**Inflation Factor:** Inflation can be described as the decline of purchasing power over time. It enables a single value representation of the rise in the cost of goods and services over time in an economy. By accounting for inflation, we can identify real growth. Additionally, we can highlight cyclical patterns in the data and/or stabilize the variance of random or seasonal oscillations. [23]

**Census Location:** As the spend of hospitality industry depends on the location of property, classification of the census location was performed based on the zip code of the property. Data was divided into multiple locations – urban, suburban, small town/metro, resort, interstate, airport. Usually, the hotels near to the airport are busier, hence it would be expected to have a higher spend.

**Business Rules:** Data of the project was filtered for US/Canada locations only as it was the specified scope. Data was further filtered out for only customers and not the prospects.

**Additional Features:** The research firm Oxford Economics discovered that a significant amount of expenditure was influenced by online reviews and ratings. According to the study on the impact of internet ratings and reviews, consumers are more likely to believe negative evaluations than positive ones. Bad ratings are reliable no matter how many reviews there are, but good ratings are only reliable when there are a lot of reviews. In summary, guests are influenced by the opinions of past visitors, and negative evaluations are unquestionably detrimental to a company's reputation. Travelers place value on a hotel's internet reputation, and the rating will determine the kind of business that can subsequently flow into the establishment. As a part of research, online ratings and reviews are scraped from google and an average of all the ratings from multiple travel websites like TripAdvisor, Booking.com are web scraped. In order to standardize, the KPI is divided by the number of rooms in the property to get an accurate sense of measure.

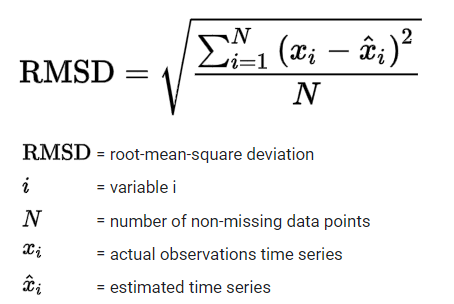
**Filtering:** The data for 2020 was removed from the analysis as the world was hit by covid-19. Due to the global pandemic and government rules for travel restrictions, the hospitality industry suffered the most. This data would not be appropriate to be used for modelling as it would create a bias in the model and would not represent the actual numbers.

**Modelling:** To build models for each granular product category, a Gridsearch cross validation is conducted to tune hyperparameters that are fed into the model. The technique selects the best parameter from the grid search and applies it to the user-selected model.

**Partition of Data:** Supervised machine learning techniques typically require splitting the data into multiple blocks for training and testing sets**.** The data in this research paper is partitioned by using 80-20 rule. This is chosen in accordance with pareto principles that suggests 80% of effects come from 20% of causes.

**Handling categorical variables:** To incorporate categorical variables into models, they must be converted to dummy variables. Since the only categorical variable in this research is census location, dummy variables for it are created. Furthermore, the dummy variable where location was unknown is dropped as it would not help to interpret the model.

**Prediction:** The Root Mean Square Error (RMSE) is a commonly used metric for determining how accurate a model is at predicting quantitative data. Basically, it normalizes the gap between predicted and actual values. A square root of RMSE helps to standardize the error in the same units.



Graphical user interface

Description automatically generated with medium confidence

## **MODEL(s)**

Post processing the data and creating an analytical data set, we needed to select the best model for each combination of Market Segment and Product/Service Category. We selected a subset of models to be implemented based on firstly, the distribution of data we had (21 categorical fields and 1 continuous field), secondly the literature explored with regards to different regression model used in the industry to forecast spend, and thirdly the cross-validation method for multiple categories.

We started with the existing model i.e. Multivariate Linear Regression, for each of the categories and fixed it as the baseline model for benchmarking the other models selected in the subset. For each combination of market segment, the test file (containing new/prospect customer information) is taken as the input, and the best model selected out of the subset of models based on lowest cross validated RMSE value is then used to predict the average monthly spend for those set of customers. We implemented a pipeline of following models to be considered for final prediction of the test dataset:

**Multivariate Linear Regression:**  
Model based on OLS optimization algorithm, where co-efficient of each feature is determined by minimizing the square of regression error. The model exploits the linear correlation between the response and the predictor variables very well and is a highly explainable model. Since for some categories correlation of SPEND and other variables were significant (>0.2), Multiple Linear Regression outperformed the other complex models that tend to overfit on such data because of inherent low bias.

**Polynomial Regression**:   
We tried to incorporate non-linearity in the model using Polynomial regression model, where each of the predictors were raised to a power to predict the response.

**Elastic Net Grid Search**:   
For some product/service categories we observed that Linear regression was highly overfitting, and one of the reasons could have been the presence of irrelevant features, since we did not perform any feature selection prior to implementing this model. We came across some literature regarding feature selection Regularized Linear Models where the performance of Linear Regression model can be enhanced by introducing regularization terms in the optimization objective function of OLS. The L1 regularization tries to minimize the absolute sum of co-efficient of predictors, by forcing the coefficient of non-significant predictors to be 0. While L2 regularization performs groupwise selection of correlated features and selects only one of them, thereby decreasing the overfit on training set. Alpha is the ratio of regularization factor of L1 to L2.

For taking out the best performing Elastic net model we performed grid search cross validation for a various values of α(0.1 to 1 with an interval of 0.1) and are calculating the Test RMSE based on the best model parameters(α).

**KNN Grid Search CV**:

Distance based algorithm that assigns a data point(SPEND) based on the closest distance Neighbor or a group of neighbors. This algorithm is used for regression analysis where the features are parsley distributed and have do not have much variance in terms of the response variable. To select the optimized number of neighbors, we implemented the model inside the Grid Search Cross Validation framework, with Number of Neighbors varying from (40 to 120). The best KNN model is then selected based on the least RMSE.  
We observed that KNN is overfitting on a lot of categories, especially for less Number of Neighbors and specifically for the categories where there were not enough number of data points.

Based on our literature review, we came across Tree based models that can be tuned based on our training set.

**Random Forest Grid Search CV**:   
An ensemble of decision trees where each tree is trained on a subset of data, and the mean of prediction of each tree is the final output. The combination of prediction from multiple trees makes this algorithm robust to overfitting while decreasing the variance of the model.

Graphical user interface

Description automatically generated with low confidence

Different hyperparameters such as number of tree, max depth, number of features were fine tuned to obtain the best model.

**XGBoost/CatBoost:**

These are implementation of Gradient Boosting Algorithm, that grow decision tree in a sequential manner instead of parallel. Each decision tree is fit on the residual of the previous tree and the process continues until all the trees(pre-specified) are trained. There is a chance of over-fitting if there are large number of trees. Mathematically, Gradient Boosting Algorithm can be described in following steps:

* Initialize the model with a constant prediction, such as the mean of the target variable.
* Train the first decision tree to predict the residual error of the model from step 1.
* Update the model by adding the predicted residual error from step 2.
* Repeat steps 2 and 3 until the desired number of trees is trained.

We implemented Grid Search Cross validation to determine the optimal number of trees and other hyperparameters to get the best models for each combination of market segment and product category

“CAT Boosting” algorithm is a special case of Gradient Boosting which has in-built parameters of recognizing the categorical features, an appropriate boosting algorithm for our data which had 18 categorical features. CAT Boost was the best algorithm for majority of the product category and market segment combination.

**Ensemble Modelling:**

To further improve the model performance, we decided to combine the best models implemented above and observe the results. We implemented two types of Ensembles using the best Random Forest, CAT Boost and XG Boost model :

* Simple Averaged Ensembles: The prediction from each model is averaged out to get the final predicted output

**Final Prediction=(PredCATBOOST+PredXGBOOST+PredRANDOMFOREST)/3**

* Stacked Generalization: The prediction from each models serve as as input to a meta model(CAT Boost in our case), and final predicted output is a the prediction of that meta model for each record

**FinalPrediction=MetaModel(PredCATBOOST,,PredXGBOOST,PredRANDOMFOREST)**

Stacked Generalization method gave better RMSE for category-market segment combination, where individual models were giving very high RMSE.

## **RESULTS**

The results shown in Table 2 is for one of the market segments that we predicted the spend for and corresponding four product service categories. Multivariate Linear Regression RMSE is the baseline RMSE, against which we are benchmarking our best models. One point to note here is that for some combination of product service category and market segment, Multivariate Linear Regression comes out to be the best model, and we are going with it to make our final prediction.

The average actual vs predicted spend is for the overall category and is based on the predictions we obtained from the best model. Here for some categories such as Maintenance, we see a very high percentage of error in the model, most likely because of the few data points we had for that product category.

Based on the best model we obtained, we also tried to observe the feature importance for each category, and observe which predictors are contributing the most in prediction of spend. The below graph shows the feature importance based on CAT Boost model for one of the categories we predicted spend for.

**Table 2:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Market Segment + Product Category** | **Best Model Name** | **Multivariate Linear Regression RMSE** | **Best Model RMSE** | **Average Actual Spend** | **Average Predicted Spend** | **Percentage Error (Actual vs Predicted)** |
| Food & Beverage | Simple Ensemble | 8290.31 | 2801.55 | 6044.59 | 6364.70 | 5% |
| Operations | XG Boost | 2299.17 | 353.87 | 894.22 | 771.90 | -14% |
| Sanitization Activities | XG Boost | 1257.16 | 287.37 | 769.27 | 763.49 | -1% |
| Maintenance | CAT Boost | 4613.61 | 1750.57 | 1879.00 | 1002.00 | -47% |

## **CONCLUSIONS**

Despite the adoption of Big Data and Data Analytics across industries, the hospitality industry is yet to fully benefit from the advances of these fields. This study was based on accurately predicting how a customer hotel of a hospitality procurement agency will spend with specific suppliers provided by that agency. We were tasked to improve on a 20-years old legacy model that implemented linear regression to predict the customer spend. Variable correlation analysis found that a unique set of variables dictated the spend of a particular hotel for a specific product or service. As a novel work in the field of predicting spend of a hotel, we trained and optimized a unique model for each combination of market segment and product category. On average our models performed 66% better than the legacy linear regression model. Using the SHAP python library [22], we extracted important variables that had the most impact on the spend for each model. The business could use these variable importance weights to ascertain which variables impact the spend to the greatest degree and what variables must a prospective hotel have or focus on to maximize their spend with the agency.

**REFERENCES**

1. Mathew, Elezabeth. *Integration of Artificial Intelligence in E-Procurement of the Hospitality Industry: A Case Study in the UAE*. Diss. The British University in Dubai (BUiD), 2020.
2. Wang, Gang, et al. "Big data analytics in logistics and supply chain management: Certain investigations for research and applications." *International journal of production economics* 176 (2016): 98-110.
3. Johnson, Michael. "A study of E‐market Adoption Barriers in the Local Government Sector." *Journal of Enterprise Information Management* (2012).
4. Mathew, Elezabeth, and Sherief Abdulla. "The LSTM technique for demand forecasting of e-procurement in the hospitality industry in the UAE." *IAES International Journal of Artificial Intelligence* 9.4 (2020): 757.
5. Friedman, Jerome H. "Greedy function approximation: a gradient boosting machine." *Annals of statistics* (2001): 1189-1232.
6. Myles, Anthony J., et al. "An introduction to decision tree modeling." *Journal of Chemometrics: A Journal of the Chemometrics Society* 18.6 (2004): 275-285.
7. Stadler, Timo, Amitrajit Sarkar, and Jan Dünnweber. "Bus demand forecasting for rural areas using XGBoost and random Forest algorithm." *International Conference on Computer Information Systems and Industrial Management*. Springer, Cham, 2021.
8. Vanichrujee, Ukrish, et al. "Taxi demand prediction using ensemble model based on rnns and xgboost." *2018 International Conference on Embedded Systems and Intelligent Technology & International Conference on Information and Communication Technology for Embedded Systems (ICESIT-ICICTES)*. IEEE, 2018.
9. Katipamula, Srinivas, T. Agami Reddy, and David E. Claridge. "Multivariate regression modeling." (1998): 177-184.
10. Bartol, Kristijan, et al. "Linear Regression vs. Deep Learning: A Simple Yet Effective Baseline for Human Body Measurement." *Sensors* 22.5 (2022): 1885.
11. Hastie, Trevor, et al. *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. New York: springer, 2009.
12. Liu, Jie, et al. "Data integration by multi-tuning parameter elastic net regression." *BMC bioinformatics* 19.1 (2018): 1-9.
13. Caruana, Rich, and Alexandru Niculescu-Mizil. "An empirical comparison of supervised learning algorithms." *Proceedings of the 23rd international conference on Machine learning*. 2006.
14. Wu, Qiang, et al. "Adapting boosting for information retrieval measures." *Information Retrieval* 13.3 (2010): 254-270.
15. Zhang, Yanru, and Ali Haghani. "A gradient boosting method to improve travel time prediction." *Transportation Research Part C: Emerging Technologies* 58 (2015): 308-324.
16. Dorogush, Anna Veronika, Vasily Ershov, and Andrey Gulin. "CatBoost: gradient boosting with categorical features support." *arXiv preprint arXiv:1810.11363* (2018).
17. Moroff, Nikolas Ulrich, Ersin Kurt, and Josef Kamphues. "Machine Learning and statistics: A Study for assessing innovative demand forecasting models." *Procedia Computer Science* 180 (2021): 40-49.
18. Zaghloul, Mohamed, Mofreh Salem, and Amr Ali-Eldin. "A new framework based on features modeling and ensemble learning to predict query performance." *PloS one* 16.10 (2021): e0258439.
19. Reid, Sam, and Greg Grudic. "Regularized linear models in stacked generalization." *International Workshop on Multiple Classifier Systems*. Springer, Berlin, Heidelberg, 2009.
20. Wolpert, David H. "Stacked generalization." *Neural networks* 5.2 (1992): 241-259.
21. Caruana, Rich, et al. "Ensemble selection from libraries of models." *Proceedings of the twenty-first international conference on Machine learning*. 2004.
22. Mokhtari, Karim El, Ben Peachey Higdon, and Ayşe Başar. "Interpreting financial time series with SHAP values." *Proceedings of the 29th Annual International Conference on Computer Science and Software Engineering*. 2019.
23. Schneider, Harold Cunningham. *An empirical investigation of different approaches to accounting for changing prices*. University of Missouri-Columbia, 1981.

## **APPENDIX**