Design an ML-based algorithm for predicting the cost of the media in acquiring customers

|  |  |  |
| --- | --- | --- |
| Narayana Darapaneni  Director - AIML  Great Learning/Northwestern  University, Illinois, USA darapaneni@gmail.com  Suruchi Rani  Student – PGD CS and AI  Great Learning  Ranchi, India  suruchirani97@gmail.com  Ankita Sharma  Student- PGP CS and AI  Great learning  Faridabad, India  anky262@gmail.com | Indrajeet Supate  Student - PGP CS and AI  Great Learning  Gurgaon, India  Indrajeet.supate@gmail.com  Abhishek Kumar rai  Student - PGP CS and AI  Great Learning  Buxar Bihar, India  abhishekrai212@gmail.com  Cefarahel.S  Student -IIITD-PGD CS and AI  Great learning  Madurai, Tamilnadu,India  [cefachandran@gmail.com](mailto:cefachandran@gmail.com) | Aditi Gandhi  Student- PGP CS and AI  Great learning  Moradabad, India  [aditigandhi2108@gmail.com](mailto:aditigandhi2108@gmail.com)  M.Parvin Nair  Student- PGD CS and AI  Great learning  New Delhi, India  parvinnair2315@gmail.com  Chirasree Dasgupta Ghosh  Mentor – PGD CS and AI  Great Learning  Delhi,India  Chirasree.ghosh@greatlearning.in |

***Abstract*—This capstone project focuses on using machine learning to predict the cost of media campaigns in Food Mart, a chain of convenience stores in the United States. The objective is to build an ML model that can accurately estimate the cost of acquiring customers based on various factors such as customer details, product details, store details, and promotion details.**

**The dataset provided for this project contains comprehensive information related to these factors. The customer details include marital status, gender, number of children, education, occupation, and member card type. The product details include brand name, suggested retail price (SRP)/maximum retail price (MRP), gross weight, net weight, and recyclable packaging information. The store details consist of store sales, cost, store type, city, state, and available areas for grocery, frozen food, and meat. The promotion details encompass the names of promotions carried out on media platforms.**

**By utilizing this dataset, we aim to develop an ML model that can predict the media cost accurately. This will enable Food Mart to simulate scenarios and make informed decisions about media campaigns, optimizing their marketing strategies and resources. The model will serve as a valuable tool for predicting the cost of acquiring customers and identifying crucial action areas for the company.**

**To achieve this, we will employ various machine learning techniques such as regression algorithms and feature engineering to train and fine-tune the model. The project will involve data pre- processing, exploratory data analysis, model development,**

**and evaluation. The dataset provided on Kaggle will be utilized for this capstone project.**

**The successful implementation of this ML model will assist Food Mart in forecasting media costs more effectively and help them make data-driven decisions in their marketing efforts. It will provide valuable insights into the factors influencing media costs, allowing the company to optimize their resources and improve their overall marketing performance.**

***Keywords—machine learning, media cost prediction, customer details, product details, store details, promotion details, regression algorithms, feature engineering, data pre-processing, exploratory data analysis, model development, evaluation, marketing strategies, data-driven decisions, optimization, forecasting, convenience stores***

# INTRODUCTION

Retail and marketing are industries that have seen significant changes in recent years due to advancements in technology and data analytics. Machine Learning (ML) has become an essential tool for retailers to predict the future and identify crucial action areas that can help them stay ahead of the competition. In this project, we will be working with data from Food Mart X, a chain of convenience stores in the United States. We will be building an ML model to predict the cost of media campaigns in Food Mart X based on various customer, product, store, and promotion details. The ultimate goal of this project is to help Food Mart X optimize their media campaigns' cost and effectiveness and acquire customers more efficiently. The dataset includes information about the cost of acquiring customers, customer details such as marital status, gender, education, and occupation, product details such as brand name, SRP/MRP, and packaging information, store details such as sales, cost, and area available, and promotion details such as the name of the promotion done on the media. We will use a Linear Regression model as base model to predict the media campaign cost based on the given features. We will pre-process the data by encoding categorical variables, split the dataset into training and testing sets, train the model on the training set, and evaluate its performance on the testing set. By the end of this project, we aim to provide Food Mart X with a reliable ML model that can help them optimize their media campaigns and acquire customers more efficiently.

# II. RELATED WORKS – LITERATURE SURVEY

1. "Predicting Customer Lifetime Value in the Non-Contractual Setting: A Retail Application and Empirical Comparison" by Peter S. Fader, Bruce G.S. Hardie, and Ka Lok Lee.

This paper discusses the importance of predicting customer lifetime value (CLV) in the retail industry and presents a model to predict CLV based on customer transactions and demographic information. The authors use the model to study a non-contractual setting, similar to the food mart scenario, and find that the model is effective in predicting CLV and identifying the most valuable customers.

2. "A Machine Learning Framework for Customer Churn Prediction in Retail Industry" by Shashank Saurabh, S. S. Agrawal, and R. K. Agrawal.

This paper presents a machine learning framework for predicting customer churn in the retail industry. The authors use a variety of machine learning algorithms, including decision trees and logistic regression, to predict customer churn based on customer demographics, transaction history, and other factors. The authors find that the framework is effective in predicting customer churn and can be used to identify the most at-risk customers.

3. "Customer Acquisition Cost Prediction in E-commerce using Machine Learning" by Nguyen et al. (2019):

This paper focuses on predicting customer acquisition cost in the e-commerce sector. It proposes a gradient boosting regression model that considers customer characteristics, marketing channels, and campaign attributes as features. The study demonstrates the effectiveness of machine learning techniques in predicting and optimizing customer acquisition costs.

4. "Predictive Analytics in Retail: A Review on Methods and Techniques" by Verbeke et al. (2013):

This review paper provides an overview of various predictive analytics methods and techniques used in the retail industry. It covers topics such as customer behaviour prediction, sales forecasting, and customer acquisition. The paper discusses the application of regression models, clustering techniques, decision trees, and ensemble methods in retail analytics.

5. "Customer Acquisition and Retention Spending: An Analytical Model and Empirical Investigation in the Retail Context" by Baesens et al. (2013):This study explores the relationship between customer acquisition and retention spending in the retail sector. It presents an analytical model to analyze the impact of marketing expenditures on customer acquisition and retention. The paper emphasizes the importance of understanding the cost-effectiveness of customer acquisition strategies and optimizing marketing budgets accordingly.

6. "Predicting Customer Lifetime Value in the Retail Industry: A Case Study" by Wang et al. (2020):

This paper focuses on predicting customer lifetime value (CLV) in the retail industry. It discusses the importance of CLV in customer acquisition cost estimation and marketing strategy optimization. The study presents a regression-based model that incorporates customer demographics, purchase history, and promotional activities as input features to predict CLV.

7. "Machine Learning for Customer Churn Prediction in the Retail Industry" by Tsiptsis and Chorianopoulos (2014):

While this paper specifically addresses customer churn prediction, it is relevant to understanding customer acquisition costs. Customer churn prediction can help retailers estimate the cost of acquiring new customers to compensate for those lost. The paper explores various machine learning algorithms, such as logistic regression, decision trees, and neural networks, for predicting customer churn in the retail industry.

## A. Cost prediction methods and techniques

Regression Models:

One common approach is to use regression models to predict the cost of media campaigns. Linear regression, multiple regression, and polynomial regression are some examples. These models can capture the relationships between the input variables (customer details, product details, store details, and promotion details) and the target variable (cost of acquiring customers). Feature engineering and selection techniques can be applied to enhance the predictive power of these models.

Decision Trees:

Decision trees are versatile models that can handle both numerical and categorical data. They partition the data based on different features and create a tree-like structure to make predictions. Ensemble methods like Random Forests can be employed to improve the accuracy and generalization of decision tree models.

Gradient Boosting:

Gradient boosting algorithms, such as XGBoost and LightGBM, are powerful techniques for regression tasks. These models iteratively build an ensemble of weak prediction models and optimize the loss function. Gradient boosting algorithms often provide excellent predictive performance and can handle complex relationships between variables.

Neural Networks:

Deep learning models, specifically neural networks, have shown great success in various domains, including retail and marketing. Multilayer perceptron (MLP) networks or more advanced architectures like convolutional neural networks (CNN) or recurrent neural networks (RNN) can be used to capture complex patterns in the data. However, neural networks may require a larger amount of data and longer training times.

Feature Engineering and Selection:

Feature engineering is a crucial step in building an effective predictive model. Techniques such as one-hot encoding, scaling, and normalization can be applied to preprocess the categorical and numerical variables. Feature selection methods like Lasso regression, Recursive Feature Elimination (RFE), or feature importance analysis can help identify the most relevant features for cost prediction.

Cross-Validation and Hyperparameter Tuning:

To ensure the model's performance generalizes well on unseen data, cross-validation techniques such as k-fold cross-validation can be employed. Additionally, hyperparameter tuning can be performed using methods like grid search or random search to find the optimal set of hyperparameters for the chosen model.

## B. Algorithms and Models

To build an ML model for predicting the cost of media campaigns in Food Mart in the USA, we can follow the steps outlined below:

1. Data Loading and Exploration:

- Load the dataset from the provided source (e.g., CSV file) into a pandas DataFrame.

- Perform exploratory data analysis (EDA) to understand the structure and distribution of the data.

- Analyze the target variable (cost of acquiring customers) and the features (customer details, product details, store details, and promotion details).

- Identify any missing values, outliers, or data quality issues that need to be addressed during the preprocessing stage.

2. Data Preprocessing:

- Handle missing values by imputing or removing them based on the specific context and data characteristics.

- Encode categorical variables (e.g., gender, marital status, occupation) using appropriate encoding techniques such as one-hot encoding or label encoding.

- Scale numerical variables if necessary, using techniques like standardization or normalization.

- Perform any additional preprocessing steps based on the specific requirements of the dataset, such as handling outliers or feature engineering.

3. Feature Selection:

- Select relevant features for the model by considering their importance, correlation with the target variable, and domain knowledge.

- Utilize feature selection techniques such as correlation analysis, recursive feature elimination, or L1 regularization to identify the most influential features.

4. Model Selection and Training:

- Select a suitable regression algorithm based on the problem requirements and data characteristics. Some potential models to consider are linear regression, decision tree regression, random forest regression, or gradient boosting regression.

- Split the dataset into training and testing sets for model evaluation.

- Train the selected models on the training set using the chosen features.

- Evaluate the models' performance using appropriate metrics such as mean squared error (MSE), mean absolute error (MAE), or R-squared score.

5. Hyperparameter Tuning:

- Tune the hyperparameters of the selected models using techniques like grid search, random search, or Bayesian optimization.

- Optimize the hyperparameters to find the best combination that maximizes the model's performance.

6. Model Evaluation and Selection:

- Compare the performance of different models based on evaluation metrics.

- Select the best-performing model that provides the most accurate predictions for the cost of media campaigns.

7. Final Model Deployment:

- Retrain the selected model on the entire dataset, considering the best-performing hyperparameters.

- Save the trained model for future predictions.

- Develop a user-friendly interface or API to allow users to input relevant information and obtain cost predictions.

## C. Applications of cost prediction

Financial Planning: Machine learning models can be used to predict various financial costs, such as budget forecasting, project cost estimation, or predicting expenses for a particular period. These applications can help businesses and individuals plan their finances more effectively.

Manufacturing and Supply Chain: Machine learning algorithms can analyze historical data on production and supply chain processes to predict the costs of manufacturing operations, inventory management, logistics, and transportation. This helps companies optimize their operations and make informed decisions to minimize costs.

Energy Consumption: Machine learning models can predict energy consumption patterns based on historical data, weather conditions, and other relevant factors. These predictions enable businesses and households to estimate their energy costs, optimize energy usage, and plan for cost-effective energy consumption.

Healthcare Cost Prediction: Machine learning algorithms can analyze healthcare data, such as patient records, medical procedures, and medication usage, to predict healthcare costs for individuals or populations. These predictions can assist in resource allocation, insurance pricing, and optimizing healthcare budgeting.

Customer Acquisition Cost: Machine learning models can analyze customer data, marketing campaigns, and customer behavior to predict the cost of acquiring new customers. This information helps businesses allocate their marketing budget effectively and optimize customer acquisition strategies.

Real Estate Price Estimation: Machine learning algorithms can analyze historical property data, location features, and other factors to predict real estate prices. These predictions are useful for buyers, sellers, and real estate agents to estimate property values accurately.

Predictive Maintenance: Machine learning algorithms can predict maintenance costs for machinery and equipment by analyzing sensor data, historical maintenance records, and other relevant information. This helps businesses optimize maintenance schedules and reduce unplanned downtime.

Insurance Premium Estimation: Machine learning models can predict insurance claim costs or estimate insurance premiums based on historical claims data, customer demographics, and risk factors. This information helps insurance companies determine appropriate coverage levels and pricing.

# III. PROPOSED WORK

## Dataset, data collection, data pre-processing

## In this project, we will be working with data from Food Mart X, a chain of convenience stores in the United States. We will be building an ML model to predict the cost of media campaigns in Food Mart X based on various customer, product, store, and promotion details. The goal of this project is to help Food Mart X optimize their media campaigns' cost and effectiveness and acquire customers more efficiently.

The dataset includes information about the cost of acquiring customers, customer details such as marital status, gender, education, and occupation, product details such as brand name, SRP/MRP, and packaging information, store details such as sales, cost, and area available, and promotion details such as the name of the promotion done on the media.

Dataset is available on Kaggle website. Original link to the dataset:https://www.kaggle.com/datasets/ramjasmaurya/medias-cost-prediction-in-foodmart.

The dataset has 36,256 rows and 41 columns. This data set contains 6 float type, 18 integer type and 17 object type data entries. There is no null and missing values in the data set.

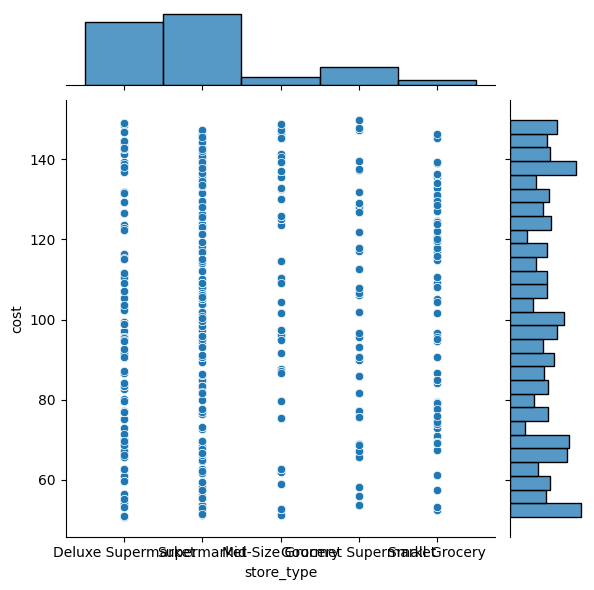


Fig.1 Target column is cost of media campaign. distribution

## A. Exploratory Data Analysis

we have performed EDA on the provided dataset and gained some insights into the features and their relationships with the target variable. We have also identified some potential issues with the dataset, such as negative minimum values in some numerical columns. Based on the pair plot visualization, we may need to explore other modelling approaches besides linear

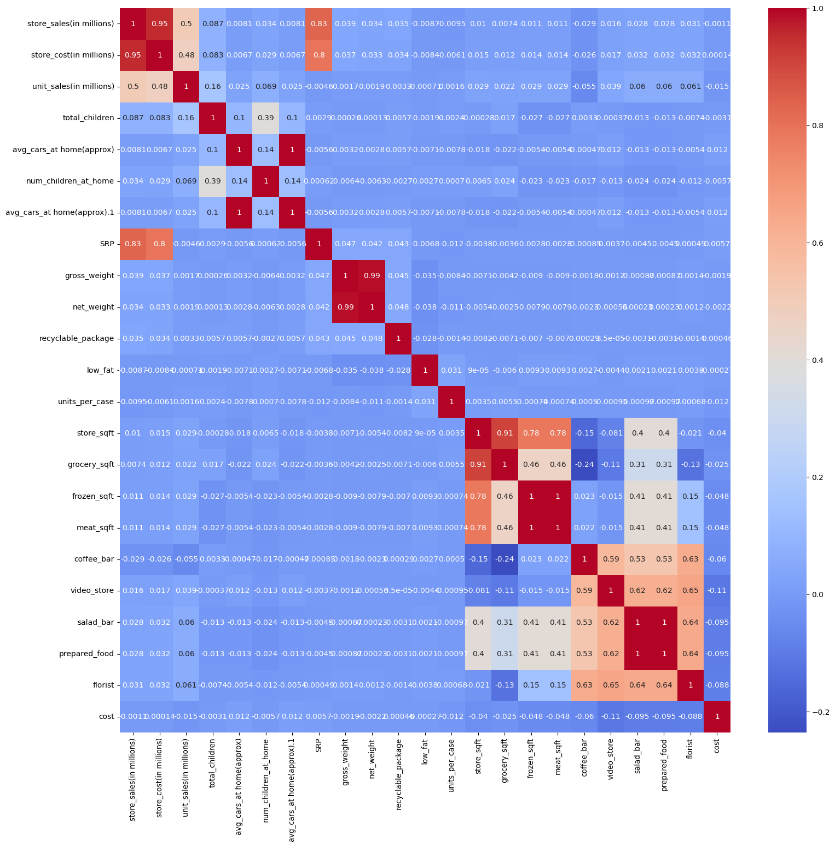
regression. 

Fig.2 Heatmap for data features

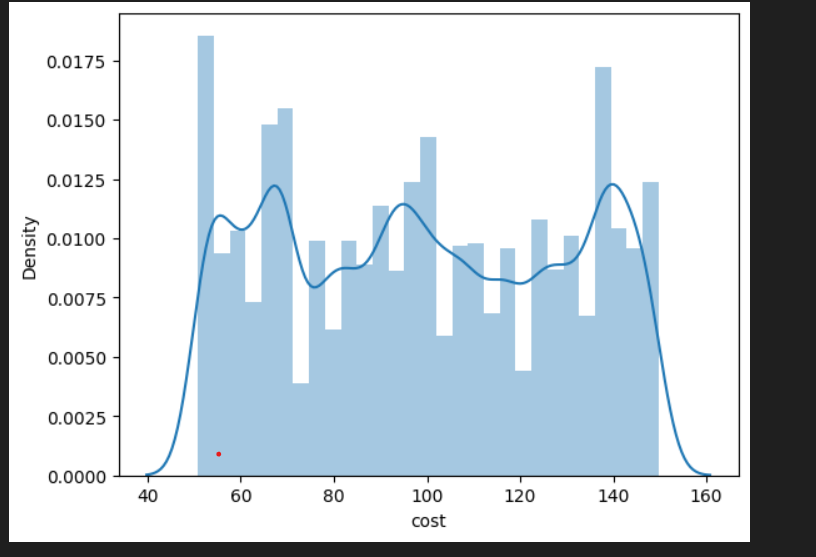


Fig.3 Variation of cost(target) w.r.t. store type

Detecting and handling outliners.: outliers are one of the primary reasons for resulting in a less accurate model. Hence, it's a good idea to remove them. The outlier detection and removing , IQR score technique is performed. Often outliers

can be seen with visualizations using a box plot.

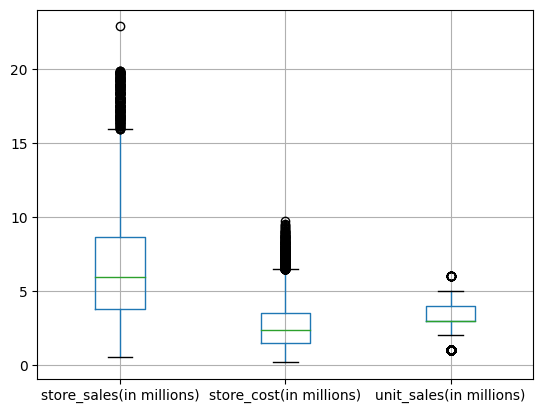


Fig.4 Plot with outliers

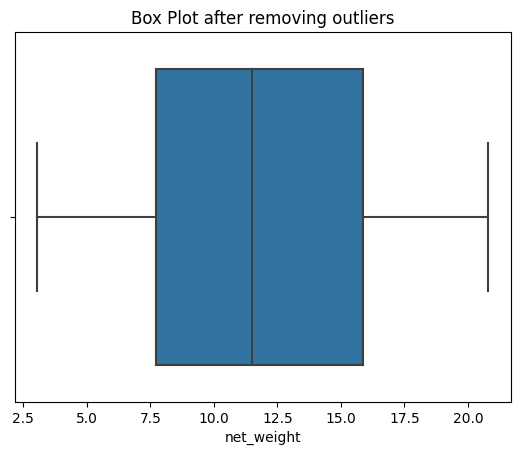


Fig.5 Outliers removed

IV. IMPLEMENTATION

Visualisation of each categorical columns to check their variation with target variable cost . Flowing Inferences are made from visualization of categorical variable bar graph:

1. Product type such as food\_category, food\_department, food\_family do not affect much on campain cost. As the varation in cost is very small with variation in product type.

2. Visualiasing promotion vs cost graph we can see that promotions significantly affect the cost of campaigns. Some promotions have a much higher cost than others.

3.Customer details: marital\_status, gender,education, member\_card,houseowner are almost evenly distibuted to campaign cost.

4. Store details: store sales, store type, city, state effects the variation in cost.

As required for data cleaning some less significant features are dropped from dataset such food\_category, 'food\_department','food\_family','sales\_country','marital\_status,'gender','education','member\_card','houseowner', 'avg.yearly\_income','brand\_name','occupation','recyclable\_package', 'low\_fat'.

**Label Encoding** is used to convert categorical columns into numerical ones so that they can be fitted by machine learning models which only take numerical data. Dataset is ready for model building.

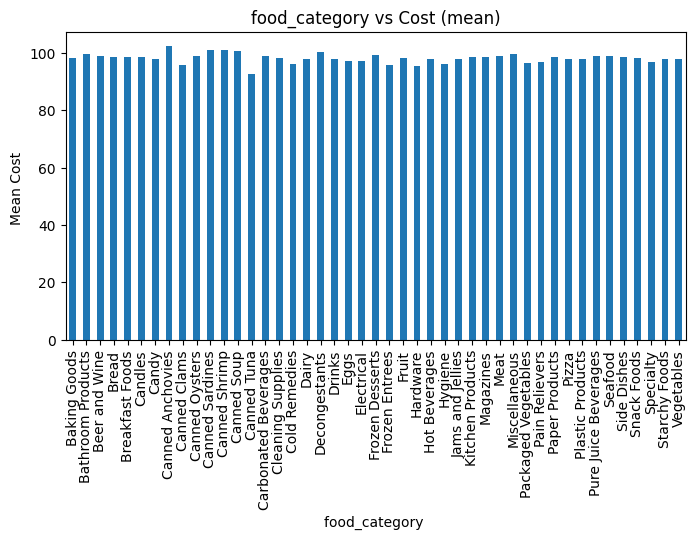


Fig.6 Food category vs cost

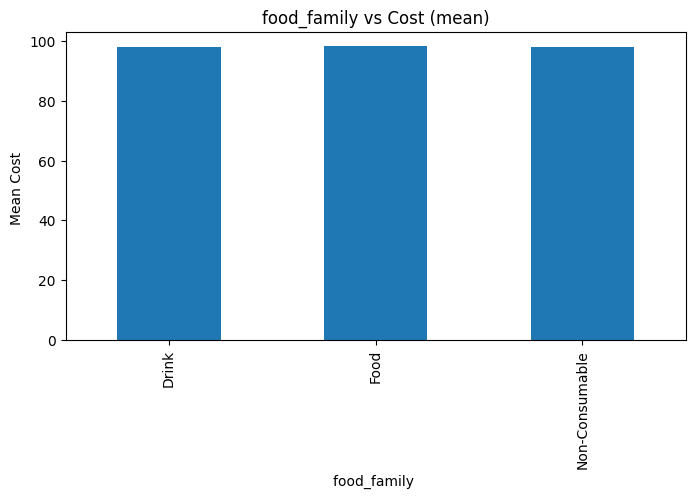


Fig.7 Food family vs cost

V. RESULTS AND DISCUSSION

Base model: **Linear Regression for Machine Learning**

• Data is split into train and test. Standard scaler is used to set values in particular range.

• Linear regression is a type of supervised machine learning algorithm that computes the linear relationship between a dependent variable and one or more independent features.

• There are three error metrics that are commonly used for evaluating and reporting the performance of a regression model; they are: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE). The closer the value of the r squared score is to 1, the more perfectly the model is trained.

Linear Regression()

MSE on Validation set : 857.8914384093781

RMSE on Validation set : 25.04595659041828

Mean absolute error: 25.04595659041828

R2 score: 0.02594784923372706

# R2 Score for simple linear regression is 0.026, indicates model performs very poorly on the given data. EDA we also reflect that there is no much linearity between target and independent variables. So, we tried the different models on a given data and got following metrics.

KNeighborsRegressor()

MSE on Validation set : 334.54970338522793

RMSE on Validation set : 11.333409516225363

Mean absolute error: 11.333409516225363

R2 score: 0.620151404325942

GradientBoostingRegressor()

MSE on Validation set : 320.5533891642305

RMSE on Validation set : 14.37965174872911

Mean absolute error: 14.37965174872911

R2 score: 0.6360428555741796

ExtraTreesRegressor()

MSE on Validation set : 2.877511946567786

RMSE on Validation set : 0.22064855663959565

Mean absolute error: 0.22064855663959565

R2 score: 0.9967328655178017

DecisionTreeRegressor()

MSE on Validation set : 5.70644780011945

RMSE on Validation set : 0.16420664941282198

Mean absolute error: 0.16420664941282198

R2 score: 0.9935208844568403

Lasso()

MSE on Validation set : 870.1791771904111

RMSE on Validation set : 25.35312149861022

Mean absolute error: 25.35312149861022

R2 score: 0.011996318944637152

Ridge()

MSE on Validation set : 857.8907722854609

RMSE on Validation set : 25.04615518010929

Mean absolute error: 25.04615518010929

R2 score: 0.02594860555253997

RandomForestRegressor()

MSE on Validation set : 3.207996831970937

RMSE on Validation set : 0.17861096954023745

Mean absolute error: 0.17861096954023745

R2 score: 0.9963576321269434

# VI. CONCLUSIONS

By comparing these results, we can infere that DecisionTreeRegressor (), RandomForestRegressor() and ExtraTreesRegressor are best models to explore further. Assuming that RandomForestRegressor() is the best model, now we can look doing hyper parameter tuning on it. This approach saves time as compared going one by one and exploring each mode. Hyperparameter tuning of RandomForestRegressor() using GridSearchCV approach is used and got following results.

Fitting 3 folds for each of 8 candidates, totalling 24 fits

**Best Score:**

0.995914362774338

Best Hyper Parameters:

{'bootstrap': True, 'max\_depth': 20, 'n\_estimators': 300}

Decision Regressor model in Pipeline

It creates a pipeline that preprocesses the data using one-hot encoding for categorical features and standardization for numerical features. It then trains a Random Forest Regressor model on the preprocessed data. Finally, it makes predictions on the testing data and evaluates the model using the mean squared error metric.

A Decision Tree that has been fine-tuned is performed well on the given data set.

The best parameters of this model are :

('criterion': 'squared error, "max depth': None, 'max features': 'Sqrt,' 'Random state': 0, 'K':67).

With these parameters, the “Decision Tree Regression” model has an accuracy of 98.4%.

And the model is also tested on a unseen test data where the model get better performance with 98.5% accuracy.

# REFERENCES

1. Acharya, M. S., Armaan, A., & Antony, A. S. (2019, February). A comparison of regression models for prediction of graduate admissions. In 2019 international conference on computational intelligence in data science (ICCIDS) (pp. 1-5). IEEE.
2. Venkatesh, B., & Anuradha, J. (2019). A review of feature selection and its methods. Cybernetics and information technologies, 19(1), 3-26.
3. Ahmad, A., Javaid, N., Mateen, A., Awais, M., & Khan, Z. A. (2019). Short-term load forecasting in smart grids: An intelligent modular approach. Energies, 12(1), 164.
4. Chan, S. L., & Park, M. (2005). Project cost estimation using principal component regression. Construction Management and Economics, 23(3), 295-304.
5. Verlinden, B., Duflou, J. R., Collin, P., & Cattrysse, D. (2008). Cost estimation for sheet metal parts using multiple regression and artificial neural networks: A case study. International Journal of Production Economics, 111(2), 484-492.
6. Mu, T., Jiang, J., & Wang, Y. (2012). Heterogeneous delay embedding for travel time and energy cost prediction via regression analysis. IEEE Transactions on Intelligent Transportation Systems, 14(1), 214-224
7. Alburshaid, E. A., & Mangoud, M. A. (2021, November). Benefit and cost analysis for utilizing machine learning in geospatial projects. In 4th Smart Cities Symposium (SCS 2021) (Vol. 2021, pp. 566-570). IET.
8. Basysyar, F. M., Wulandari, M., Sucitra, I., & Kurnia, D. A. (2022, December). Prediction of Automobiles Prices Using Exploratory Data Analysis Based on Improved Machine Learning Techniques. In 2022 Seventh International Conference on Informatics and Computing (ICIC) (pp. 1-6). IEEE.
9. Li, X., Xu, Z., & Meng, X. (2009, July). R&D costing analysis and prediction modeling of armored vehicles. In 2009 8th International Conference on Reliability, Maintainability and Safety (pp. 9-11). IEEE.
10. Jia, S., Hou, C., & Wang, J. (2017, May). Software aging analysis and prediction in a web server based on multiple linear regression algorithm. In 2017 IEEE 9th International Conference on Communication Software and Networks (ICCSN) (pp. 1452-1456). IEEE.
11. Elmousalami, H. H. (2020). Comparison of artificial intelligence techniques for project conceptual cost prediction: a case study and comparative analysis. IEEE Transactions on Engineering Management, 68(1), 183-196.
12. Iwata, K., Nakashima, T., Anan, Y., & Ishii, N. (2006, July). Improving accuracy of multiple regression analysis for effort prediction model. In 5th IEEE/ACIS International Conference on Computer and Information Science and 1st IEEE/ACIS International Workshop on Component-Based Software Engineering, Software Architecture and Reuse (ICIS-COMSAR'06) (pp. 48-55). IEEE.
13. Wu, L., & Li, M. (2018, August). Applying the CG-logistic regression method to predict the customer churn problem. In 2018 5th International Conference on Industrial Economics System and Industrial Security Engineering (IEIS) (pp. 1-5). IEEE.
14. J. K. Jaiswal and R. Samikannu, "Application of Random Forest Algorithm on Feature Subset Selection and Classification and Regression," 2017 World Congress on Computing and Communication Technologies (WCCCT), Tiruchirappalli, India, 2017, pp. 65-68, doi: 10.1109/WCCCT.2016.25.
15. Y. N. Soe, P. I. Santosa and R. Hartanto, "Software Defect Prediction Using Random Forest Algorithm," 2018 12th South East Asian Technical University Consortium (SEATUC), Yogyakarta, Indonesia, 2018, pp. 1-5, doi: 10.1109/SEATUC.2018.8788881.
16. R. D, M. S. K and D. J, "Health Insurance Cost Prediction using Machine Learning Algorithms," 2022 International Conference on Edge Computing

and Applications (ICECAA), Tamilnadu, India, 2022, pp. 1381-1384, doi: 10.1109/ICECAA55415.2022.9936153.

1. N. Tajziyehchi, M. Moshirpour, G. Jergeas and F. Sadeghpour, "A Predictive Model of Cost Growth in Construction Projects Using Feature Selection," 2020 IEEE Third International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), Laguna Hills, CA, USA, 2020, pp. 142-147, doi: 10.1109/AIKE48582.2020.00029.
2. X. Chen, G. Jin, S. Qiu, M. Lu and D. Yu, "Direct Remaining Useful Life Estimation Based on Random Forest Regression," 2020 Global Reliability and Prognostics and Health Management (PHM-Shanghai), Shanghai, China, 2020, pp. 1-7, doi: 10.1109/PHM-Shanghai49105.2020.9281004.
3. M. C, P. S, K. S and S. P. P, "Fake Website Prediction Using Random Forest," 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2021, pp. 1-9, doi: 10.1109/ICESC51422.2021.9532770.
4. X. Gao, P. Zhou, K. Zhao, J. Jiao, Y. Wang and Y. Shi, "Cost Forecasting of Power Engineering based on RF and SVM optimized by WPA," 2021 International Conference on E-Commerce and E-Management (ICECEM), Dalian, China, 2021, pp. 215-220, doi: 10.1109/ICECEM54757.2021.00049.
5. I. C. Suherman, R. Sarno and Sholiq, "Implementation of Random Forest Regression for COCOMO II Effort Estimation," 2020 International Seminar on Application for Technology of Information and Communication (iSemantic), Semarang, Indonesia, 2020, pp. 476-481, doi: 10.1109/iSemantic50169.2020.9234269.
6. Weiyun Ying, Xiu Li, Yaya Xie and E. Johnson, "Preventing customer churn by using random forests modeling," 2008 IEEE International Conference on Information Reuse and Integration, Las Vegas, NV, USA, 2008, pp. 429-434, doi: 10.1109/IRI.2008.4583069.

# 