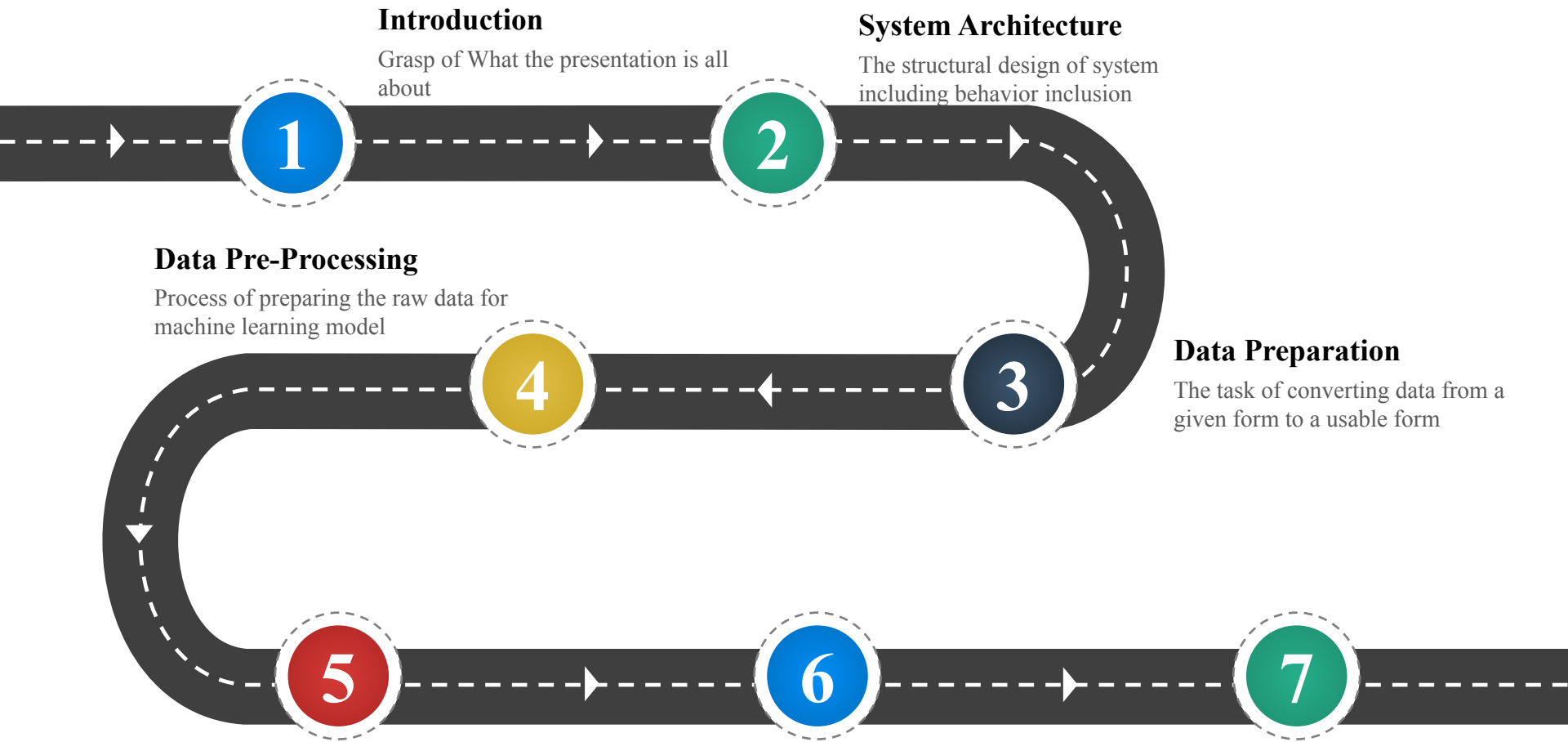
Customer Acquisition Cost Prediction Problem

By - Piyush Srivastava, Utkarsh Srivastava, Byrapu Shrisha Choudhary,



Introduction

Grasp of What the presentation is all about

System Architecture

The structural design of system including behavior inclusion

Data Pre-Processing

Process of preparing the raw data for machine learning model

Data Preparation

The task of converting data from a given form to a usable form

Models

Modeling relationship of one target var with several independent var

Closing Reflections

Looking back at the journey of the problems faced & solution

Acknowledgements

Readiness to show appreciation for & to return kindness

Introduction

Customer Acquisition Cost(CAC):
CAC is the cost related to acquiring new customers, it is calculated based on cost spent to get new customers

Scope of our Project

- **Predict Customer Acquisition Cost**
- **Analyze Multiple Machine Learning Models**
- **Determine how our model can be used to resolve Retail market problems**

About the Data Set

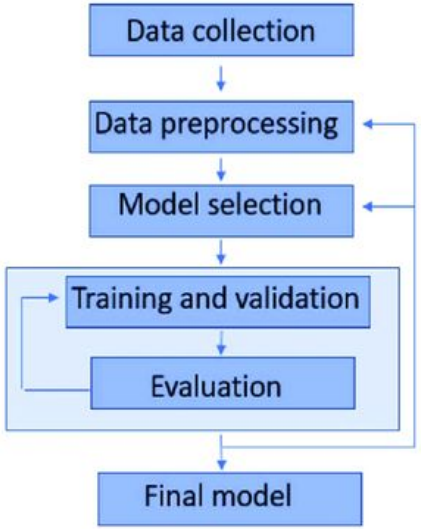
- Data Set we implemented on this project was divided into Test & Train
- There are **36256** samples in the "**Train.csv**" and **12086** samples in the "**Test.csv**".
- There are 17 Categorical and 23 Continuous attributes

	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA
1	store_sale	store_cost	unit_sales	promotion	sales_cou	marital_st	gender	total_child	education	member_c	occupatio	houseown	avg_cars_	avg. yearly	num_child	avg_cars_	brand_nan	SRP	gross_wei	net_weigh	recyclable	low_fat	units_per_
2	4.35	1.3485	3	Double Do	Canada	M	F	2	High Schoc	Bronze	Profession	Y	2	$110K - $1	1	2	Steady	1.45	9.65	7.64	1	0	14
3	5.58	1.8414	3	Shelf Clear	USA	S	F	4	High Schoc	Bronze	Profession	Y	2	$130K - $1	0	2	Ebony	1.86	18.4	17.3	0	1	9
4	5.1	2.142	3	Two Day S	Mexico	S	F	2	Graduate I	Bronze	Profession	Y	2	$70K - $90	0	2	Portsmout	1.7	11.2	9.19	1	0	28
5	4.2	2.058	3	Cash Regis	USA	S	F	2	Partial Hig	Normal	Skilled Ma	Y	1	$10K - $30	0	1	Toucan	1.4	17.9	15.8	1	1	32

AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO
units_per_	store_type	store_city	store_stat	store_sqft	grocery_sc	frozen_sqf	meat_sqft	coffee_ba	video_stor	salad_bar	prepared_	florist	media_typ	cost
14	Deluxe Sup	Vancouver	BC	23112	16418	4016	2678	1	1	1	1	1	Cash Regis	142.71
9	Supermark	Seattle	WA	21215	13305	4746	3164	1	0	0	0	0	Radio	99.77
28	Deluxe Sup	Merida	Yucatan	30797	20141	6393	4262	1	1	1	1	1	Product At	76.88
32	Deluxe Sup	Salem	OR	27694	18670	5415	3610	1	1	1	1	1	Daily Pape	55.59

Target

System Architecture

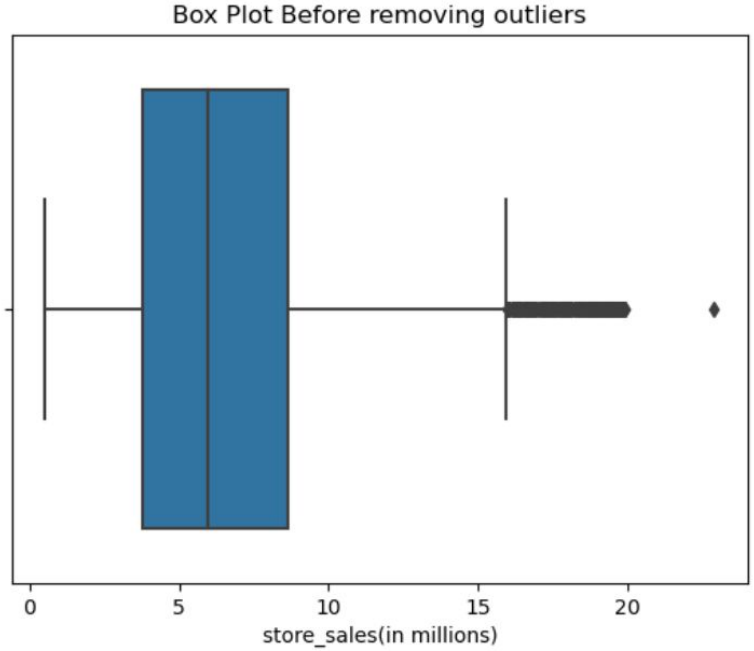


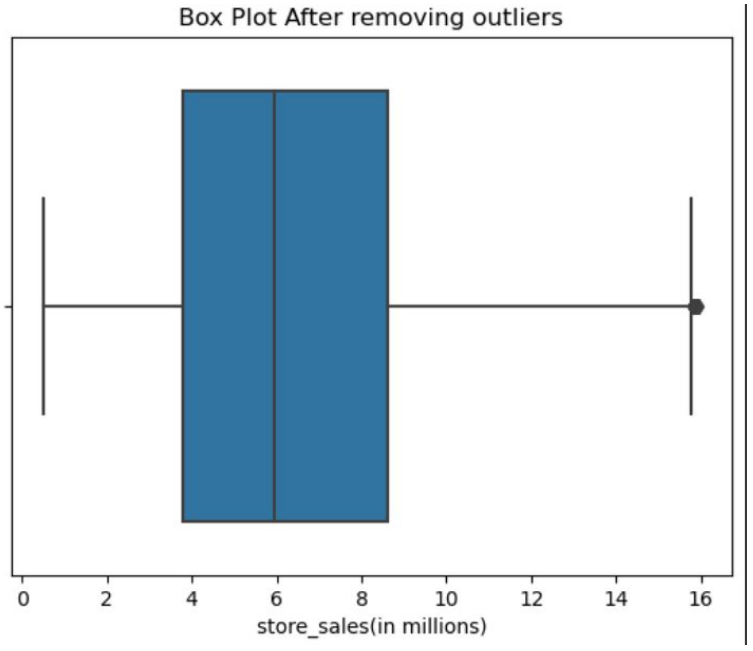
We chose Linear Regression as our base model and implemented couple of other models and performed evaluation based on below parameters

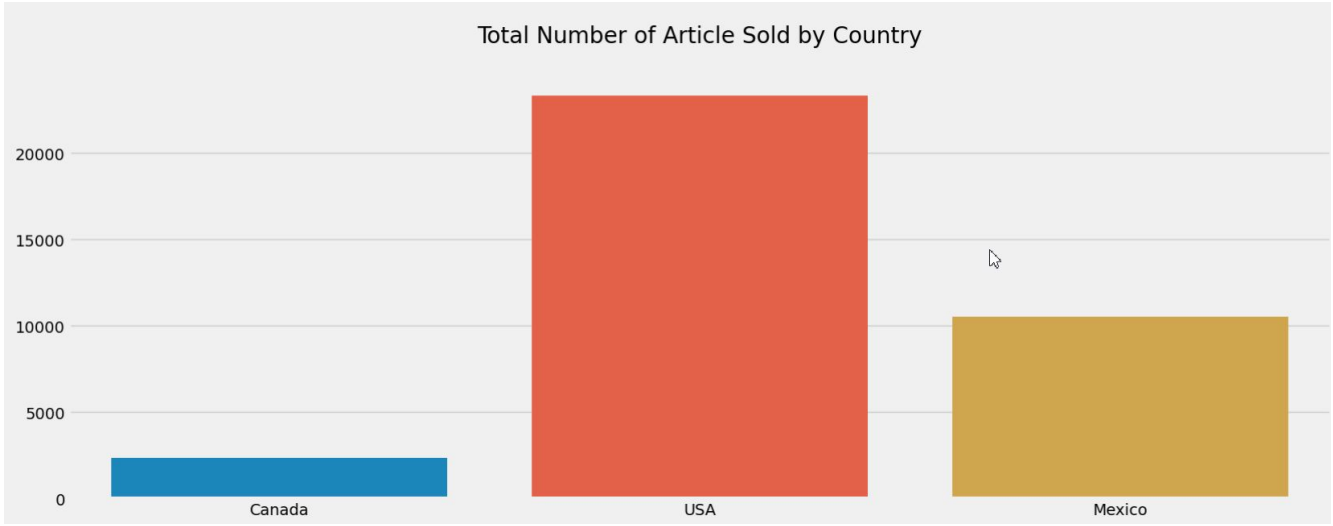
- **Mean Squared Error (MSE)**
- **Root Mean Squared Error (RMSE)**
- **Mean Absolute Error (MAE)**
- **Mean absolute percentage error(MAPE)**

Data Pre-Processing: Challenges

- Data pre-processing is a process of preparing the raw data & making it suitable for a machine learning model.
- Exploratory data analysis will include imputation of null values, missing values, duplicate

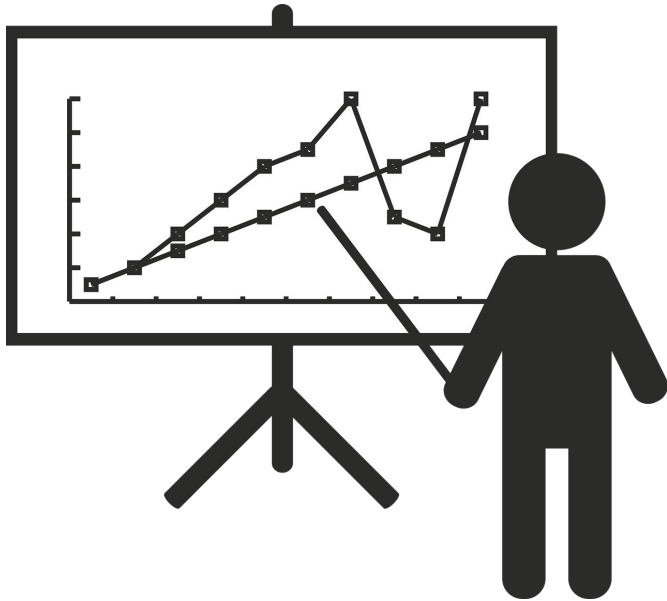





Total Number of Article Sold by Country
20000
15000
10000
5000
0
Canada
USA
Mexico

Data Preparation: Challenges

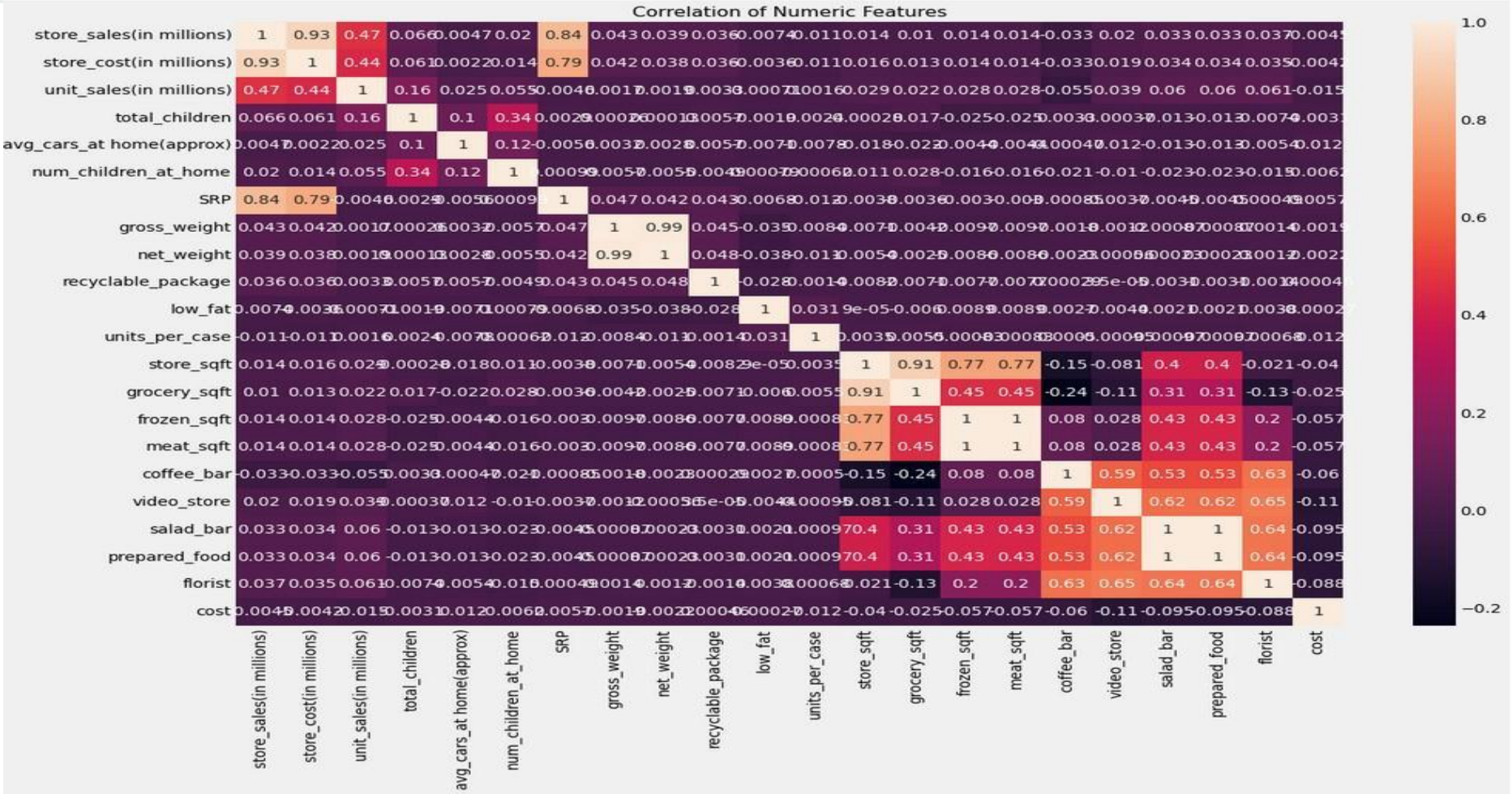
- Data Processing is the task of converting data from a given form to much more usable & desired form i.e making it more meaningful & informative.
- The dataset we have chosen is “PREDICT COST ON MEDIA CAMPAIGNS IN FOOD MART OF USA”.
- Various methods like EDA & Statistical techniques were used to treat the outliers, any wrong markings & rectified them and made sure of the least possible errors.



4 different patterns for the values in the column:

- X to Y – an example is $30K to 50K. We can convert this to a numeric value by extracting the numbers and averaging them.
- NaN We'll preserve NaN values, and not convert them at all.
- X+ – an example is $150K +. We can convert this to a numeric value by extracting the number.

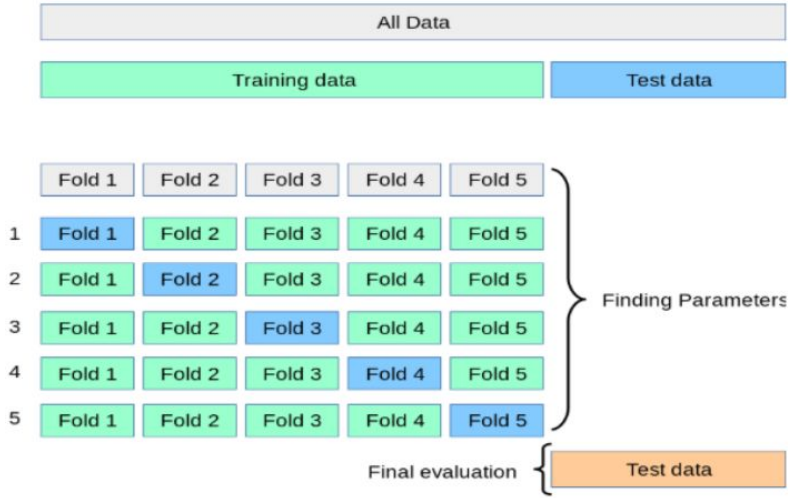
- As we can clearly see from the unique values of the media type column that some of the values are repeating and have mixed up data into a single vlaue. Like (TV, Radio, Daily Paper, Sunday Paper) multiple individual values have mixed up together to form a single value of media type.
- So we will be segregating all unique values and creating a new column in the dataset


Correlation of Numeric Features
store_sales(in millions)
store_cost(in millions)
unit_sales(in millions)
total_children
avg_cars_at home(approx)
num_children_at_home
SRP
gross_weight
net_weight
recyclable_package
low_fat
units_per_case
store_sqft
grocery_sqft
frozen_sqft
meat_sqft
coffee_bar
video_store
salad_bar
prepared_food
florist
cost
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

HyperParameter Tuning Parameters chosen

- Hyperparameter Tuning with Grid search

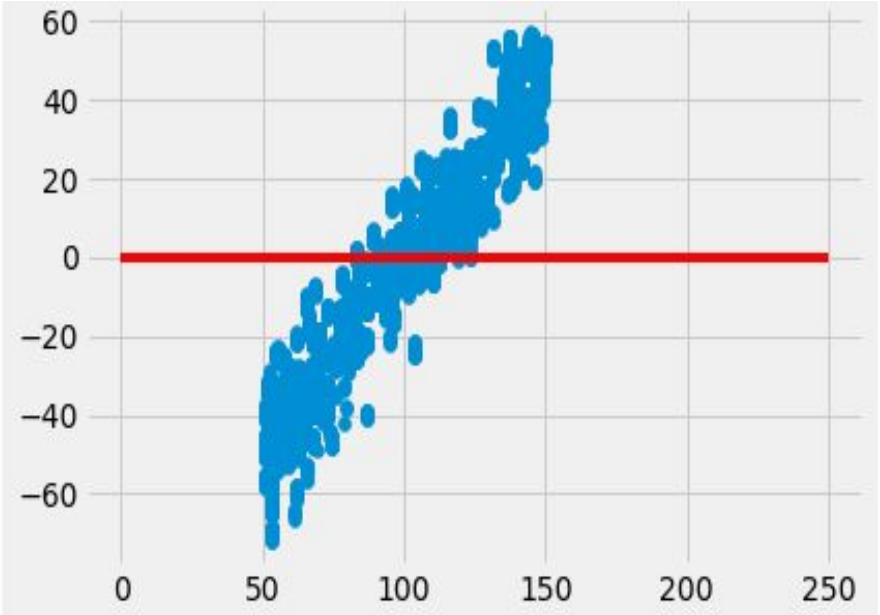
  Grid Search starts with defining a search space **grid**. The grid consists of selected hyperparameter names and values, and **grid search** exhaustively searches the best combination of these given values. One of the popular hyperparameter optimization techniques is grid search, which can further help improve the performance of a model by finding the optimal combination of hyperparameter values.



Base Model for evaluation:

- **Linear Regression:**

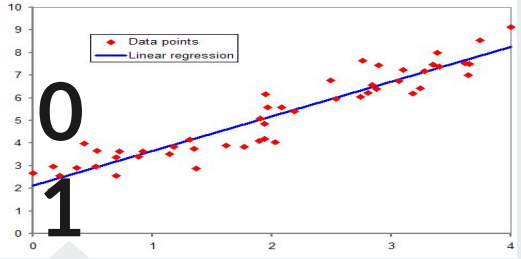
The method for modeling the relationship of one target (dependent) variable y with one or several independent variables (predictors) x, is called linear regression (LR).Scatter graph we can clearly see the linear relation between x="y_test", y="y_test-y_pred_test"



- Hyperparameter tuning of Linear Regression Model using GridSearchCV approach

```
R2 Score Of Linear Regression:  0.09183104999360854
MSE: 817.86
```

Models used and optimised



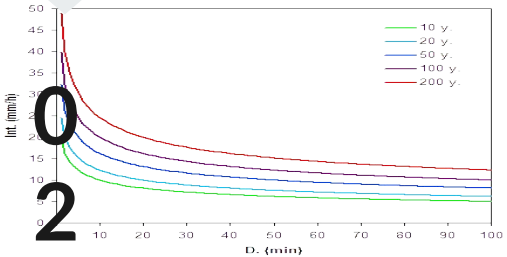
01

Linear Regression

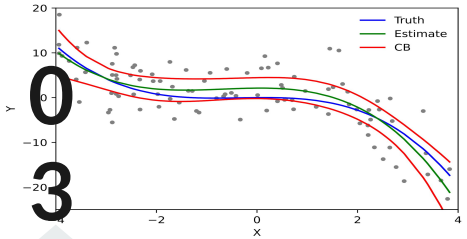
When there is only one predictor variable it is called simple linear regression & when there are several predictors, it is called multiple linear regression.

Lasso Regression

The “lasso” minimizes the residual sum of squares subject to the sum of the absolute value of the coefficients being less than a constant



02

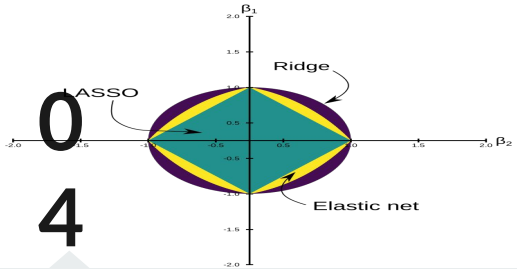


03

Ridge Regression

The ridge regression is the sum of squared regression coefficients, giving rise to the ridge regression.

Model



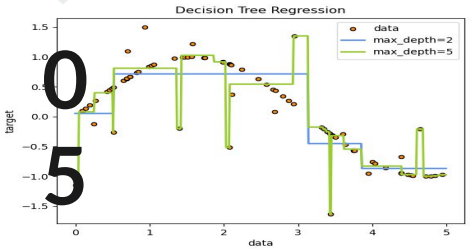
04

Elastic Net Regression

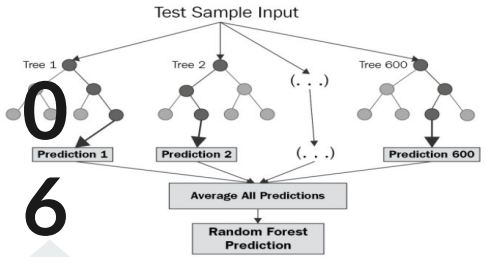
We propose the elastic net, a new regularization & variable selection method. The elastic net encourages a grouping effect, where strongly correlated predictions tend to be in or out.

Decision Tree Regression

A decision tree is a data structure that is built using nodes & edges. Each node in the tree can be either a decision node that makes decisions or a leaf node that gives the outcome.



05



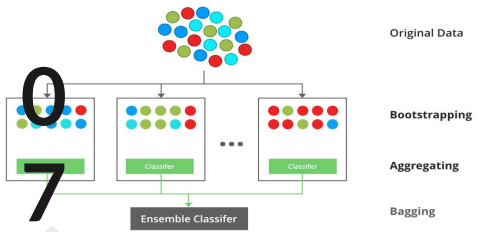
06

Random Forest

Random forest is a data mining algorithm that uses decision trees to draw conclusions based on randomly sampled data.

Model

07

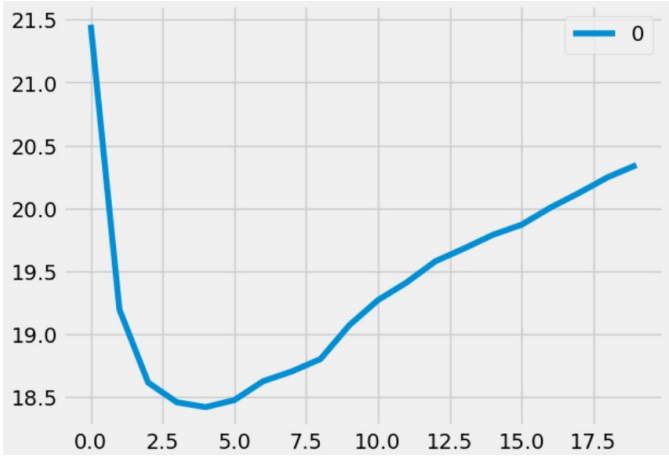


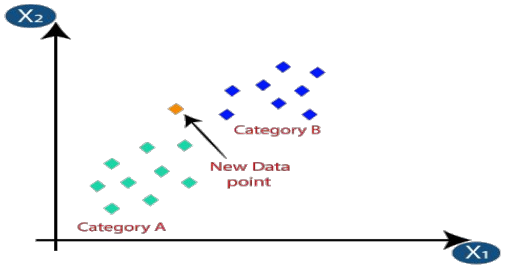
XGBoost Regression

The prediction performance of the XGBoost method is evaluated by comparing observed & predicted PM2.5 concentration using three measures of forecast accuracy.

KNN (K-Nearest Neighbor)

KNN algorithm assumes the similarity between the new case/data & available cases & put the new case into the categorythat is most similar to the available categories.





Model Metrics

- It can be noticed that the **Mean Absolute percentage error** of **Random Forest Regressor** (0.0006441) & **Decision Tree Regressor** (0.0002626) is almost equivalent to each other, since MAPE is less for decision tree so it considered best.
- **Best performing model uptil now is Random Forest Regressor.**

Model	R2 Score	MAPE	MSE	RMSE	MAE
Linear Regression	0.09700030561843187	0.28410133648341507	817.8603343304433	4.918606516643921	24.192690065572044
Lasso Regression	0.09696626236502959	0.2841634528705908	817.8911678144217	4.919209130292609	24.198618467554166
Ridge Regression	0.09700029855952941	0.28410134680534177	817.8603407237969	4.918606618636711	24.192691068896863
ElsaticNet Regression	0.09683070271187555	0.28410134680534177	817.8603407237969	4.918606618636711	24.192691068896863
Decision Tree	0.9985228153889094	0.0002626614463205253	1.337908204633205	0.18956870373673051	0.03593629343642431
Random Forest	0.998897519257417	0.0006441324071288373	0.9985333044208483	0.28194379860356245	0.07949230557100617
XgBoost	0.9077826959267506	0.0006441324071288373	0.9985333044208483	0.28194379860356245	0.07949230557100617
KNN	0.49166984813177905	0.19493527822602488	460.4022244347664	4.083704154847434	16.676639624318195

HyperParameter Tuning

- There are two types of parameters in machine learning. The parameters that are learned from the training algorithm and the learning algorithm parameters that are optimized separately. Those are tuning parameters also called hyperparameter models such as C value, Sequential Forward, Backward and Bi-Directional.
- **Hyperparameter tuning** also known as hyperparameter optimization is an important step in any machine learning model training that directly affects model performance.
- **Hyperparameters** are parameters that are **defined before training** to specify how we want model training to happen. We have full control over hyperparameter settings and by doing that we control the learning process. **Parameters**, on the other hand, are **found during the training.**

Design, develop and train the Pipeline

ML pipeline is **a means of automating the machine learning workflow by enabling data to be transformed and correlated into a model that can then be analyzed to achieve outputs**. This type ML pipeline makes the process of inputting data into the ML model fully automated.

1. **Creation of ML pipeline**

- Make_pipeline()" function that will create the pipeline
  From sklearn.pipeline import Pipeline, make_pipeline

2. **Random Forest Base Model as was best model use as reference**

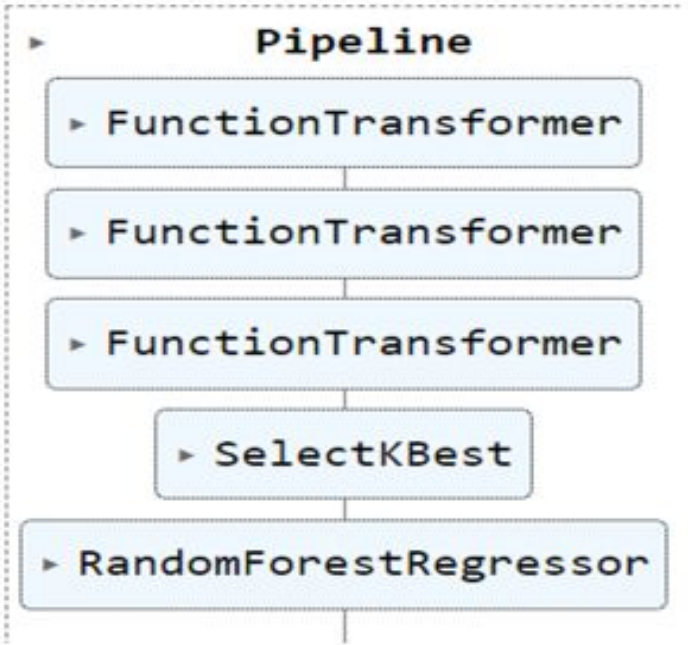
3. Filtered sample from train data as test data is already processed.
   Typically when our code run to , data is not processed.

4. **Creating activities to put in the pipeline used (Function transformers)**

5. **Creating pipeline using the activities.**

6. Call the fit() function on the pipeline object

7. **Saving pipeline for deployment to production**
   import pickle



Implementation of Pipelines

Design, develop and train the Pipeline

Pipeline:

1. we need to tie together many different processes that we use to prepare data for machine learning based model
2. It is paramount that the stage of transformation of data represented by these processes are standardized
3. Pipeline, along with the GridsearchCV helps search over the hyperparameter space applicable at each stage

Hyper parameters are like handles available to the modeler to control the behavior of the algorithm used for modeling Hyper parameters are supplied as arguments to the model algorithms while initializing them.

Linear Regression Base Model

Train R2 Score Of Linear Regression: 0.0954127994241738 Test R2 Score Linear Regression: 0.09700030561843187

Random Forest Base Model

Train R2 Score Of Random Forest: 0.9998232139847923 Test R2 Score Random Forest: 0.998913645321768

Random Forest Model score after hyperparameter tuning

Best R2 score: .9971852037905068 Best Hyper Parameters: 'bootstrap': True, 'max_depth': 17, 'n_estimators': 30}

Based on the conducted study, key findings are:

1. There is considerable improvement in the r2 score from the base regression model when random forest model is used.

2. Best performance w.r.t r-square score is obtained using random forest model.

3. There is no considerable impact on performance after Hyperparameter tuning in the random forest model

FUTURE IMPROVEMENTS

1. As a next steps, we can explore more features which can potentially impact the media cost.

2. Further we can try applying PCA(principle component analysis) to reduce multicollinearity.

3. Also, we can explore state of art deep neural network to capture variations in the data.

Acknowledgement

- We owe an immense amount of Gratitude to **Mr. Reddy** for helping us with the paper submissions.
- Our thanks to IITD, Delhi Professors and senior members of IIITD Delhi for giving guidance in Computer science and Machine learning Skills.

Thank you.