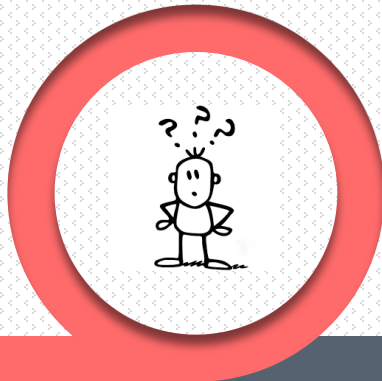# EXL Excellence Quotient 2021 Business Case Study

Team DIVERGENT

CHALLENGE

WHAT ?

To develop a tool which predicts and forecast the number of daily covid-19 cases at a county level in USA.

HOW ?

Performing exploratory Data analysis on the given train dataset containing relevant information updated up to January 2021 and training a model to predict on validation dataset.
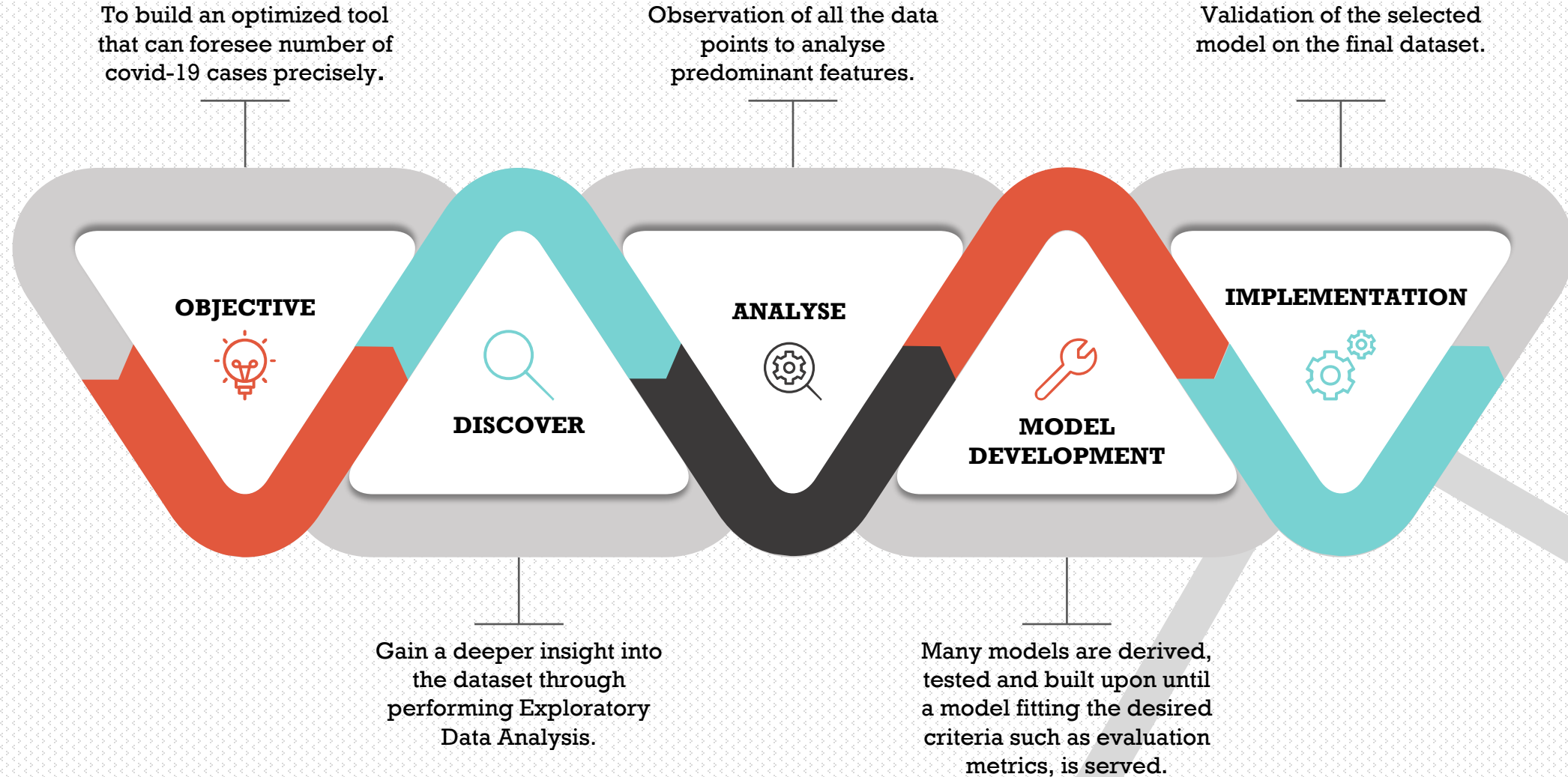
WHY ?

To design a vaccination strategy to optimize the total distribution cost to minimize the number of covid-19 cases forecasted by the model on a daily basis.

TARGET

A model/tool is to be designed for a US Public Health Client

SOLUTION DESIGN

To build an optimized tool that can foresee number of covid-19 cases precisely.

Observation of all the data points to analyse predominant features.

Validation of the selected model on the final dataset.

OBJECTIVE

DISCOVER

ANALYSE

MODEL DEVELOPMENT

IMPLEMENTATION

Gain a deeper insight into the dataset through performing Exploratory Data Analysis.

Many models are derived, tested and built upon until a model fitting the desired criteria such as evaluation metrics, is served.
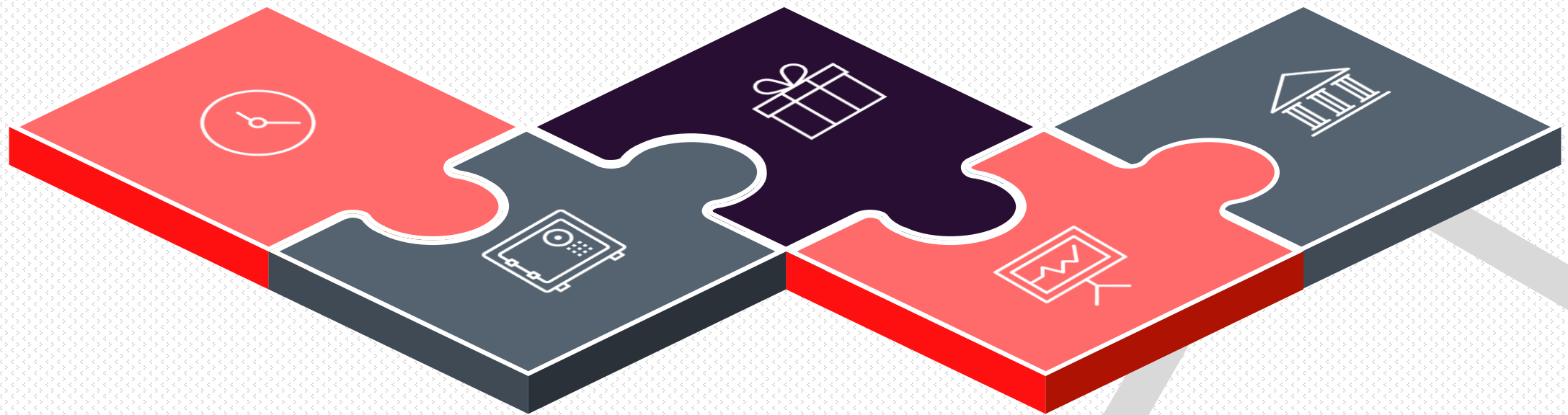
B rainstorming and scaling the features to discover, fill or drop missing values from various columns.

D evelopment of model by training and testing several Machine Learning algorithms on the dataset viz. Decision Tree, Random forest, AdaBoost.

A fter reading the training dataset, EDA is performed to visualize all the relevant information
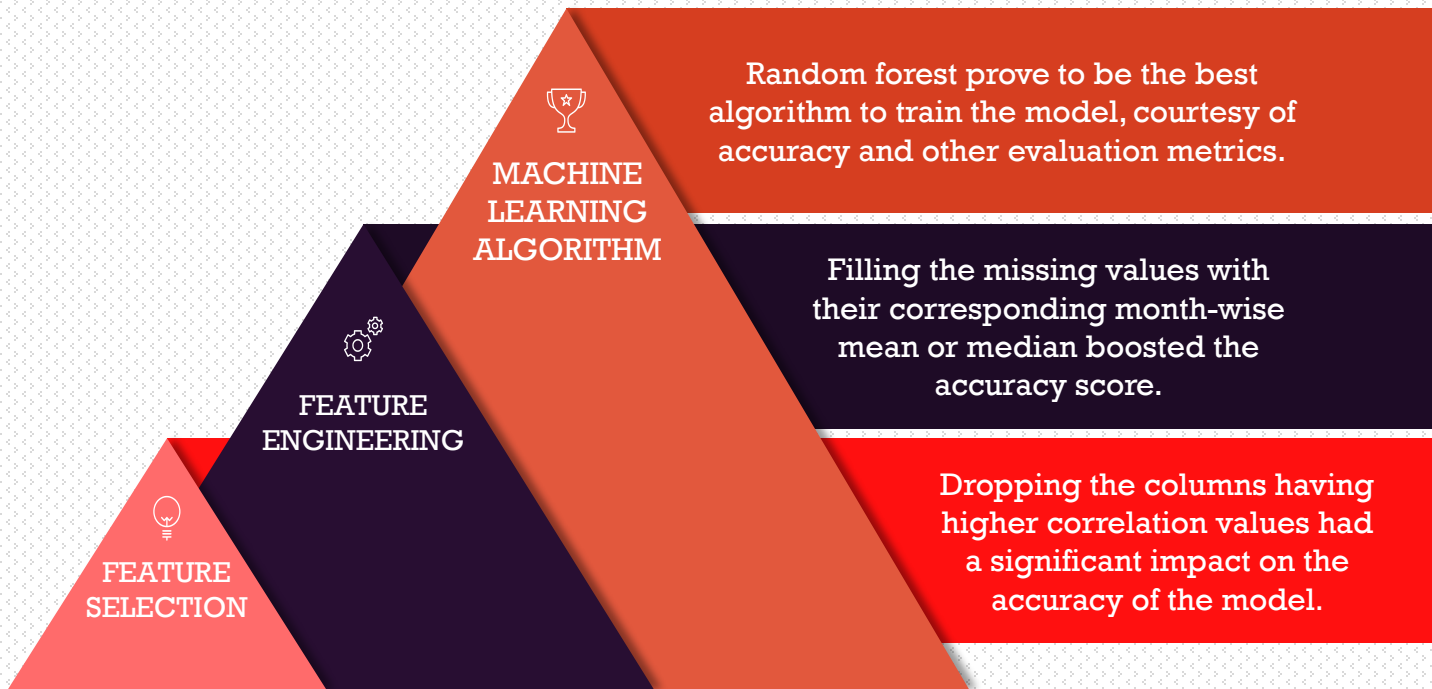
C hoosing dominant features with the help of feature selection technique like Correlation matrix to examine dependency between columns

E valuation and prediction on the validation dataset using our final model.

REVIEW

TOP DRIVERS OF THE MODEL

MACHINE LEARNING ALGORITHM

Random forest prove to be the best algorithm to train the model, courtesy of accuracy and other evaluation metrics.

FEATURE ENGINEERING

Filling the missing values with their corresponding month-wise mean or median boosted the accuracy score.

FEATURE SELECTION

Dropping the columns having higher correlation values had a significant impact on the accuracy of the model.

ADDITIONAL VARIABLES TO EXPLORE

Finding an optimal combination of hyper-parameters of a machine learning model can minimize a predefined loss function to give better result

Hyper Parameter Tuning

Exploring important features obtained from extra tree regressor, could have yielded better accuracy.

ExtraTree Regressor

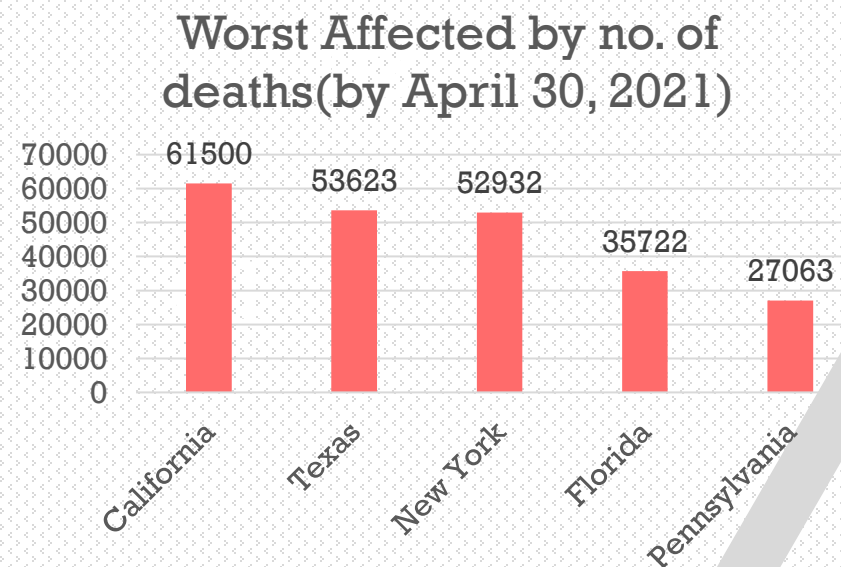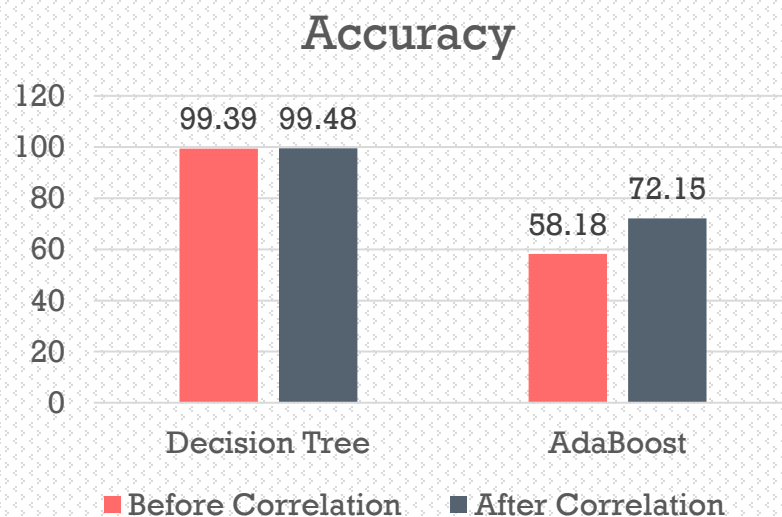| ALGORITHM	Mean Squared Error (MSE)	Mean Absolute Error (MAE)	Accuracy	Root Mean Squared Error (RMSE)
Random Forest	150828.66	36.55	99.89%	388.36
Decision Tree	900271.54	56.15	99.39%	948.82
AdaBoost	58200265.99	5774.65	58.18%	7628.91

The results shown here was obtained by training the model on trial dataset. As observed Random forest performed best among all.

The graph conveys top 5 worst affected states of USA, according to the total projected number of deaths by April 30,2021.

Accuracy

	Before Correlation	After Correlation
Decision Tree	99.39	99.48
AdaBoost	58.18	72.15

Worst Affected by no. of deaths(by April 30, 2021)

California	61500
Texas	53623
New York	52932
Florida	35722
Pennsylvania	27063

SOURCE - worldmeters.info

6

KEY POINTS

- Supply chain readiness is key to efficiently deploying COVID-19 vaccines to the target population in line with defined vaccination strategies.
- We will need to ensure adequate human resources capacity for the management of the vaccine cold chain and supply chain.
- Procedure of reverse logistics need to be strengthened to allow tracking of vaccines during the campaign.
- We also have to ensure a safe and effective health care waste management plan and budget for training, are in place, including the option of outsourcing to the private sector

DISTRIBUTION

- Assessment of current state,: identifying where there is greater vaccine demand based on case rates and locations of those most in need.

- Establishing management protocols to ensure quality and integrity of the COVID-19 throughout the supply chain

- Building of an optimization model to identify where distribution centers should be located and thus ideal transportation network.

DEPLOYMENT OPERATIONS

- Reception and storage of vaccines and ancillary products.
- Production or purchase of coolant packs
- Transportation of vaccines
- Reverse Logistics and managing recalls
- Management of supply chain information.