A PROJECT REPORT ON,

"HOUSE PRICE PREDICTION USING LINEAR REGRESSION"

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE

SUBMITTED BY

THOTA SAIKUMAR

DEPARTMENT OF COMPUTER ENGINEERING

NBN SINHGAD SCHOOL OF ENGINEERING, PUNE-41

SAVITRIBAI PHULE PUNE UNIVERSITY

2021 - 22

Sinhgad Institutes

CERTIFICATE

This is to certify that the project report entitles

"House Price Prediction Using Linear Regression"

Submitted by

Thota Saikumar

is a bonafide work carried out by him under the supervision of **Prof. Shital Kamat** and it is approved for the partial fulfillment of the requirement of University of Pune as a part of Laboratory Practice III work syllabus (Final year Computer Engineering).

(Prof. Ms. Snehal Rathod)

Subject In-charge

(Prof. Mr. Amol Dhumane)

Head, Department of Computer Engineering

ACKNOWLEDGEMENT

I would like to acknowledge the importance of collaboration and knowledge-sharing among professionals in this field, as well as the advancements in technology and data science that have made this possible. Additionally, I would like to thank the individuals who have contributed to the development and implementation of linear regression models for house price prediction, and those who continue to advance this field through their research and innovations Building a house price prediction model using linear regression involves the collective effort of various professionals, including data scientists, statisticians, software developers, and real estate experts. Their contributions to this field have enabled the creation of accurate and reliable models that help individuals and organizations make informed decisions in the real estate market..Our sincere thanks and deep gratitude to Head of Department, Prof. A. V. Dhumane and other faculty member; but also, to all those individuals involved both directly and indirectly for their help in all aspect of the project.

At last, but not least we express our sincere thanks to our Institute's Principal **Dr. Y. P. Reddy**, for providing us infrastructure and technical environment.

Thota Saikumar

ABSTRACT

Linear regression is a widely used statistical technique for predicting a continuous target variable based on one or more predictor variables. In the context of house price prediction, linear regression can be used to model the relationship between various features of a house, such as its size, number of bedrooms, location, and age, and its sale price. In this project, we aim to build a linear regression model to predict the sale price of houses in a particular city based on a set of input features. We begin by collecting a dataset of house prices and their associated features, which we split into training and testing sets. We then pre-process the data by performing feature scaling and handling missing values, if any.

Next, we train the linear regression model using the training data and evaluate its performance using various metrics such as mean squared error and R-squared. We also perform feature selection and regularization techniques to improve the model's accuracy and reduce overfitting. Finally, we use the trained model to predict the sale prices of houses in the testing set and compare the predicted prices with the actual prices using various evaluation metrics. The results of the evaluation can help us identify the strengths and weaknesses of the model and guide us in further improving its performance.

CONTENTS

1. Problem Statement Definition

Given a dataset of housing prices with different features such as the number of bedrooms, bathrooms, square footage, etc., the goal is to build a linear regression model that can accurately predict the sale price of a house based on its features..

2. Software Requirements and Specification

2.1 Introduction

Linear Regression is a popular machine learning algorithm that is used to predict a continuous output variable based on one or more input variables. In the case of house price prediction, the input variables may include factors such as the size of the house, the number of bedrooms and bathrooms, the location of the house, and the age of the house.

2.2 Scope

Linear regression is a commonly used technique in machine learning to model the relationship between a dependent variable and one or more independent variables.

2.3 Requirements

A Dataset.

Data cleaning and preparation tools.

Programming language and libraries.

Training and testing datasets.

Evaluation metrics.

Model deployment.

3. System Implementation

3.1 Feature Engineering

Feature engineering is the process of transforming raw data into meaningful features that can be used to improve the performance of machine learning algorithms. It involves selecting, combining, and transforming the features in the input data to create new features that are more informative or easier for the model to understand. Feature engineering is a crucial step in machine learning, as the quality of the features can have a significant impact on the accuracy and effectiveness of the model. Good feature engineering can lead to more accurate models, faster training times, and better generalization to new data.

Feature engineering can involve a variety of techniques, such as selecting the most important variables, transforming the data through scaling, normalization, or encoding categorical variables, creating new features through feature extraction, and reducing the dimensionality of the data through feature selection or extraction techniques. Effective feature engineering can lead to improved model performance, faster training times, and better generalization to new data. It requires a deep understanding of the data and the problem at hand, as well as creativity and experimentation to identify the most relevant features.

.

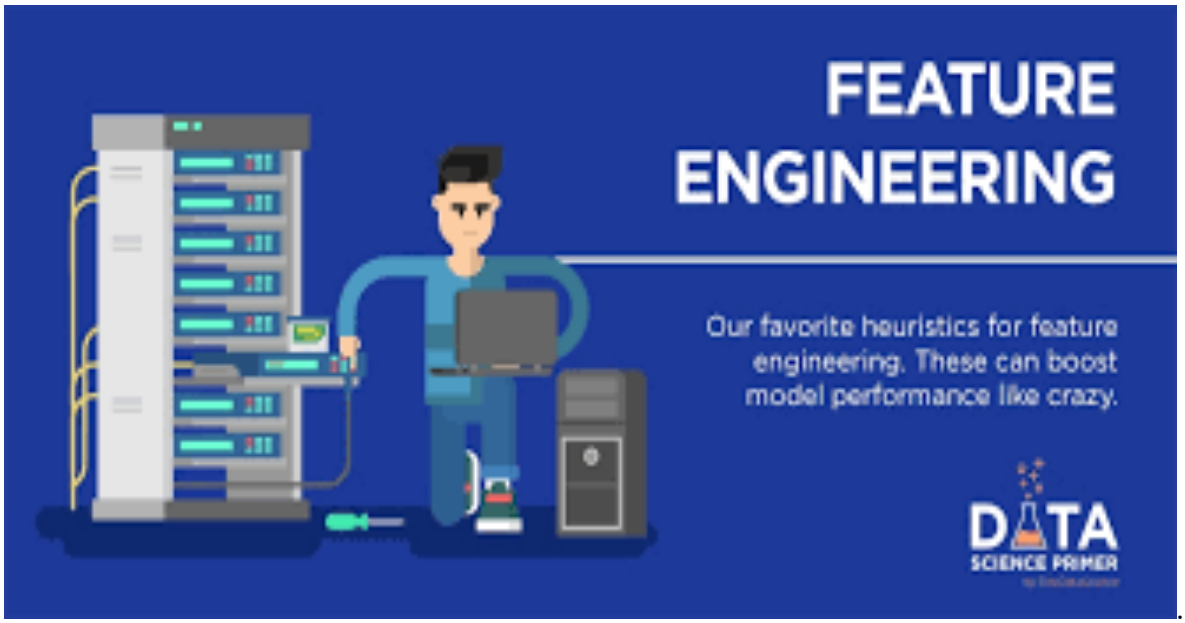
3.2 Model Deployment

Linear regression is a popular machine learning algorithm used for predicting numerical values. In the context of house price prediction, linear regression can be used to build a model that predicts the price of a house based on its features such as number of bedrooms, bathrooms, square footage, location, etc. Once the model has been trained and evaluated, it can be deployed to make predictions on new data. There are several ways to deploy a machine learning model, but one common approach is to create an API (application programming interface) that can be used to serve predictions over the internet. Next, an API can be created using a web framework such as Flask or Django. The API should define an endpoint that accepts input data in the form of a JSON object containing the features of the house (e.g., number of bedrooms, square footage, etc.). The API should then use the loaded model to make a prediction based on the input data and return the predicted house price as a JSON object. Finally, the API can be deployed to a web server or cloud platform such as AWS Lambda or Google Cloud Functions. The API can then be accessed by clients over the internet using HTTP requests.

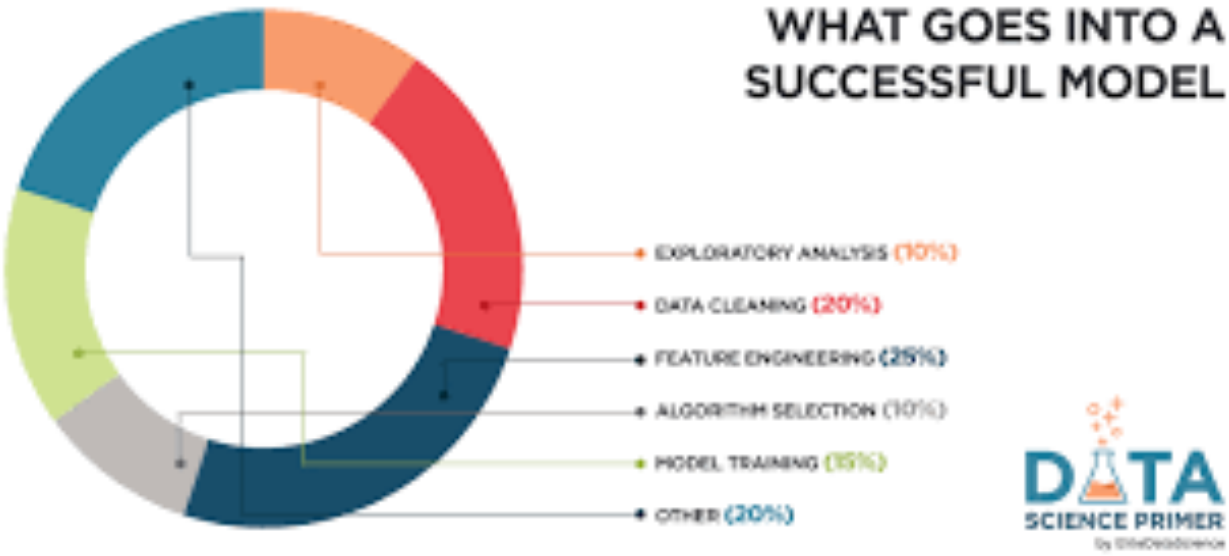
4. GUI/Snapshots

4.1 Feature Engineering

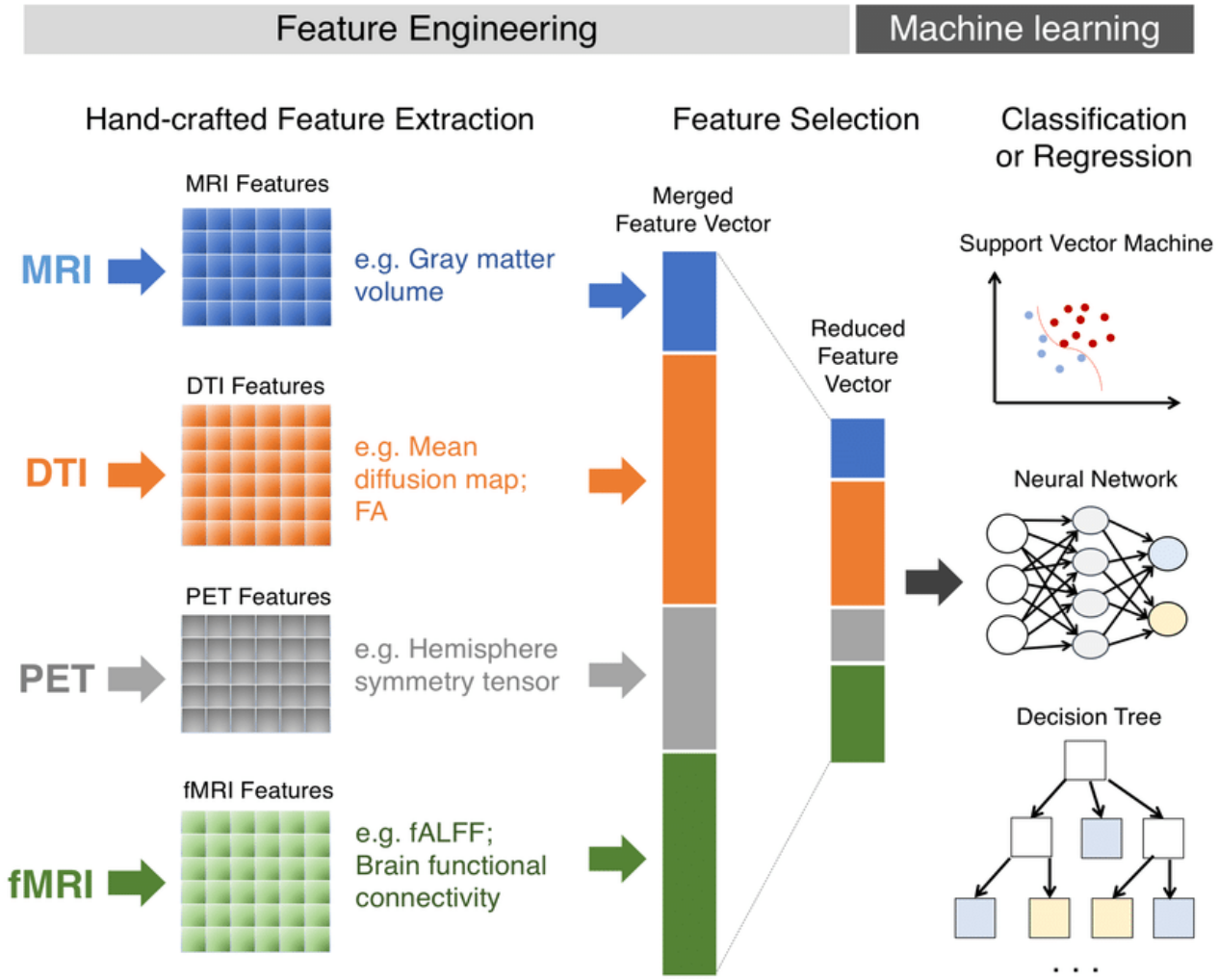
1.



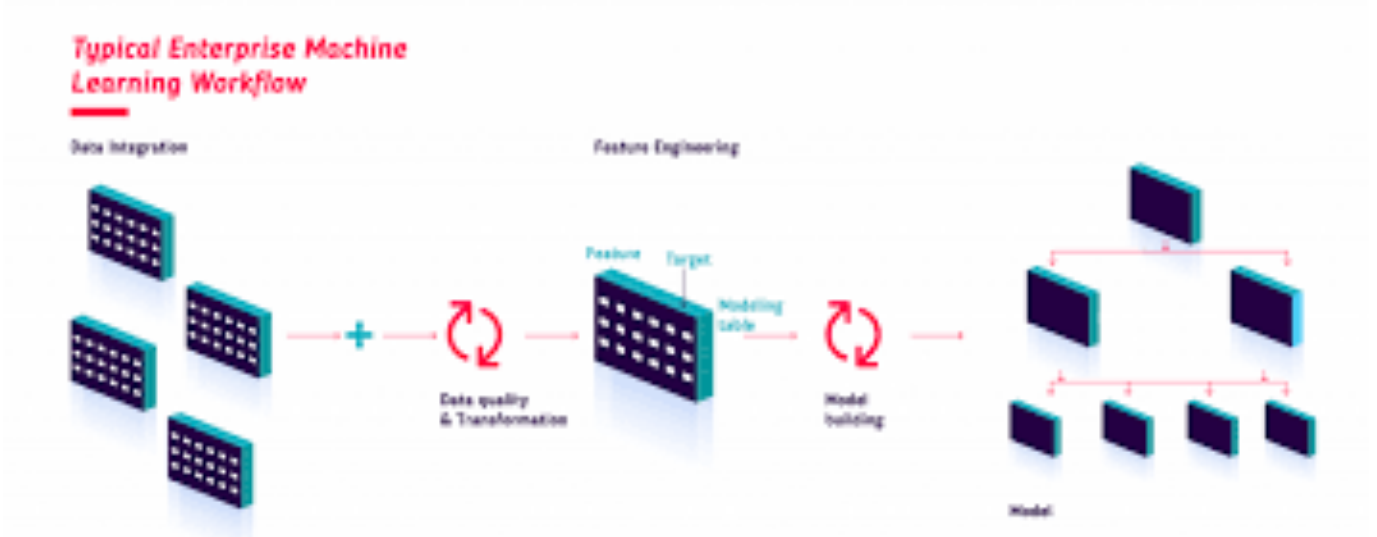
2.



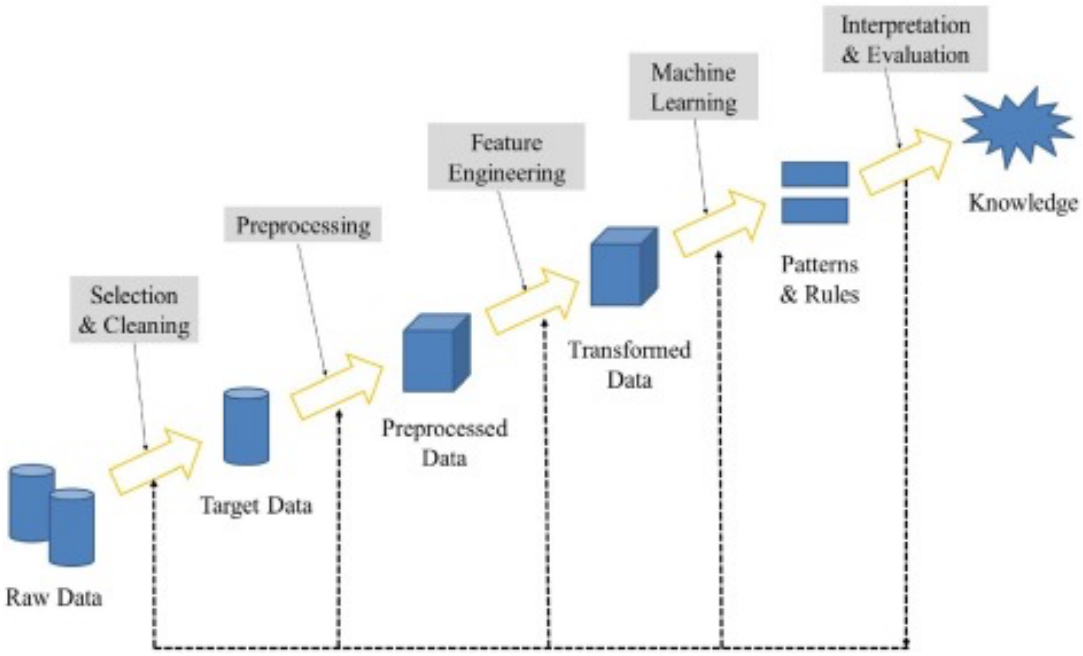
3.


Feature Engineering
Machine learning
Hand-crafted Feature Extraction
Feature Selection
Classification or Regression
MRI Features
MRI
e.g. Gray matter volume
DTI Features
DTI
e.g. Mean diffusion map; FA
PET Features
PET
e.g. Hemisphere symmetry tensor
fMRI Features
fMRI
e.g. fALFF; Brain functional connectivity
Merged Feature Vector
Reduced Feature Vector
Support Vector Machine
Neural Network
Decision Tree

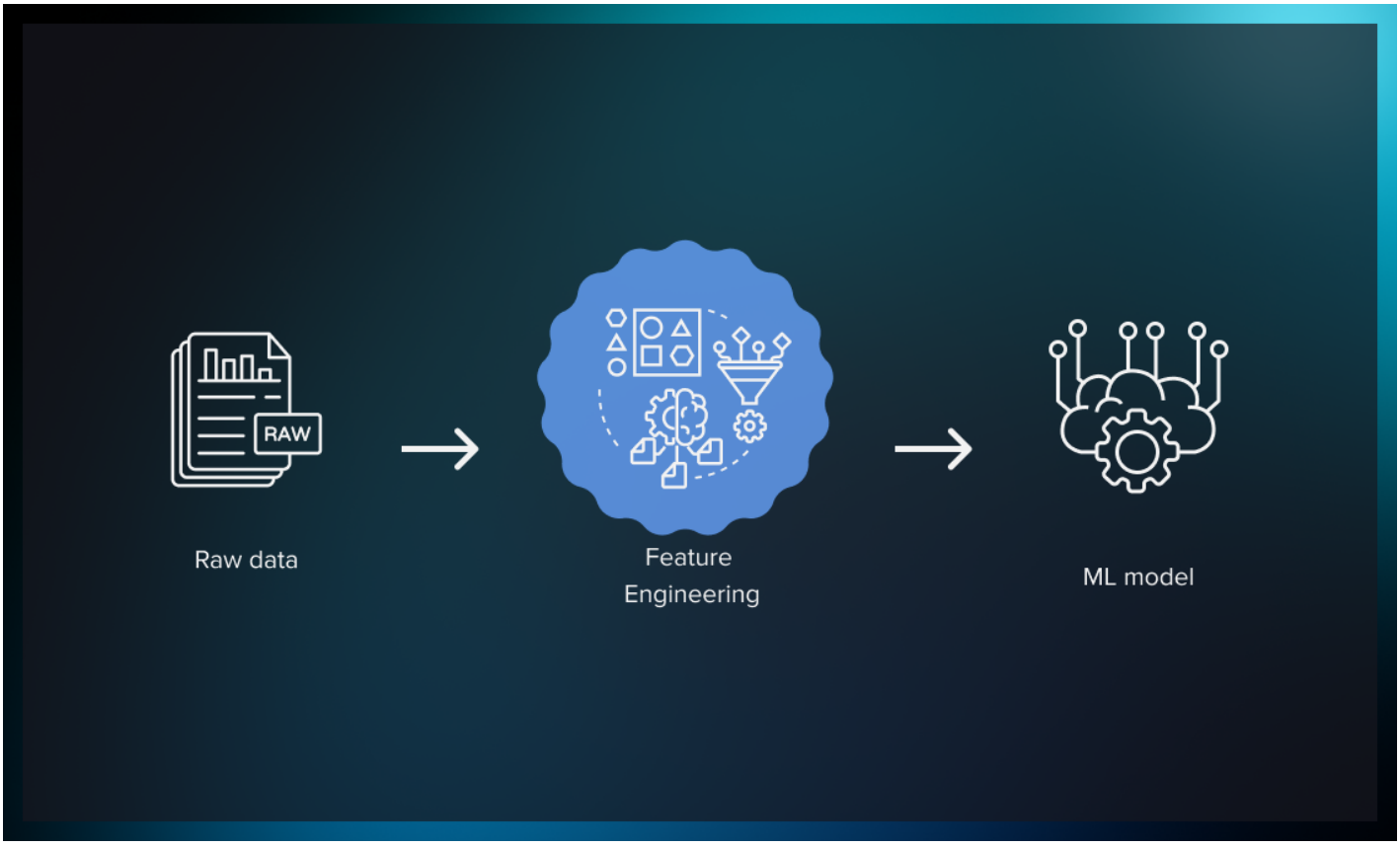
4.


Typical Enterprise Machine Learning Workflow

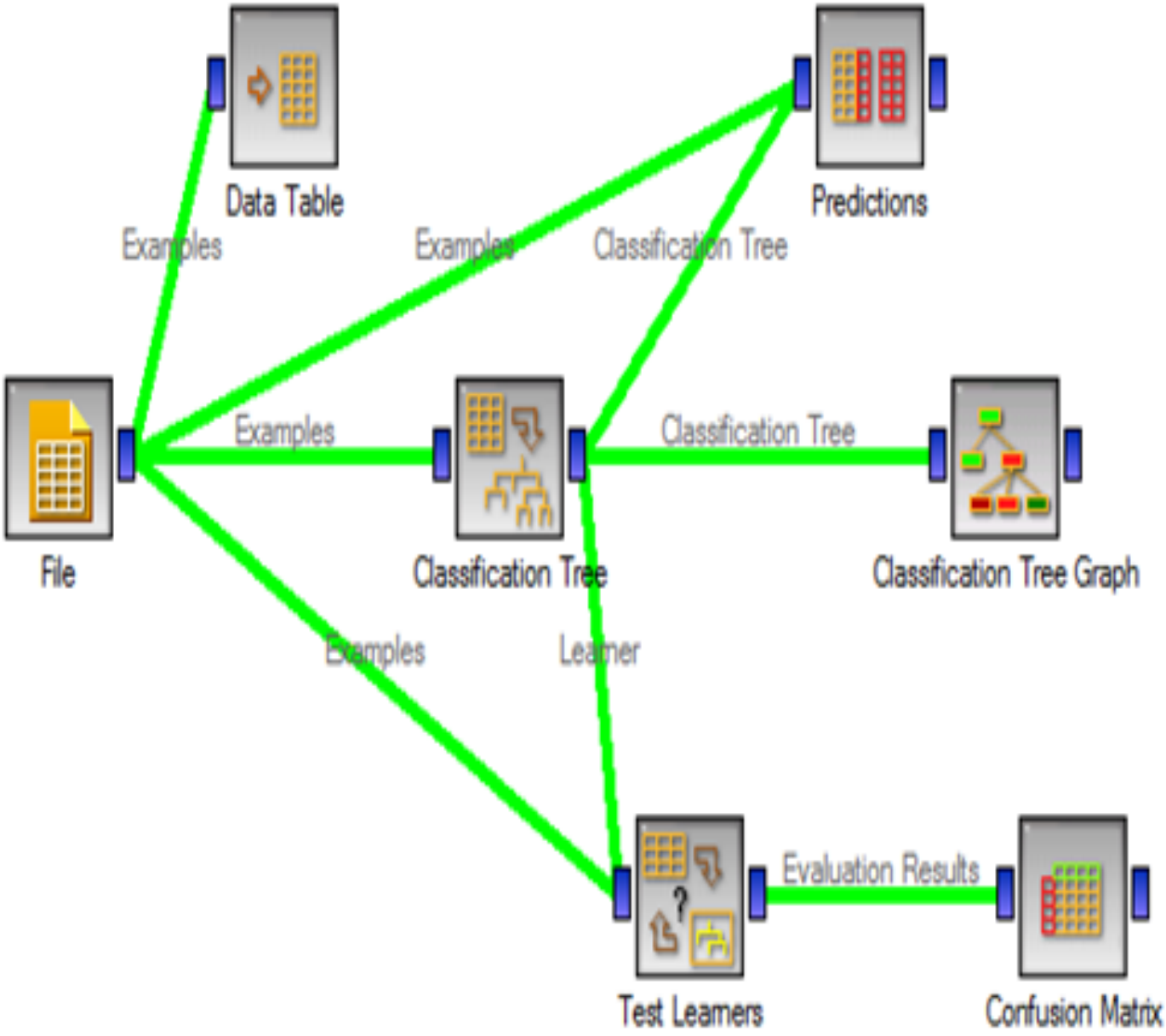
5

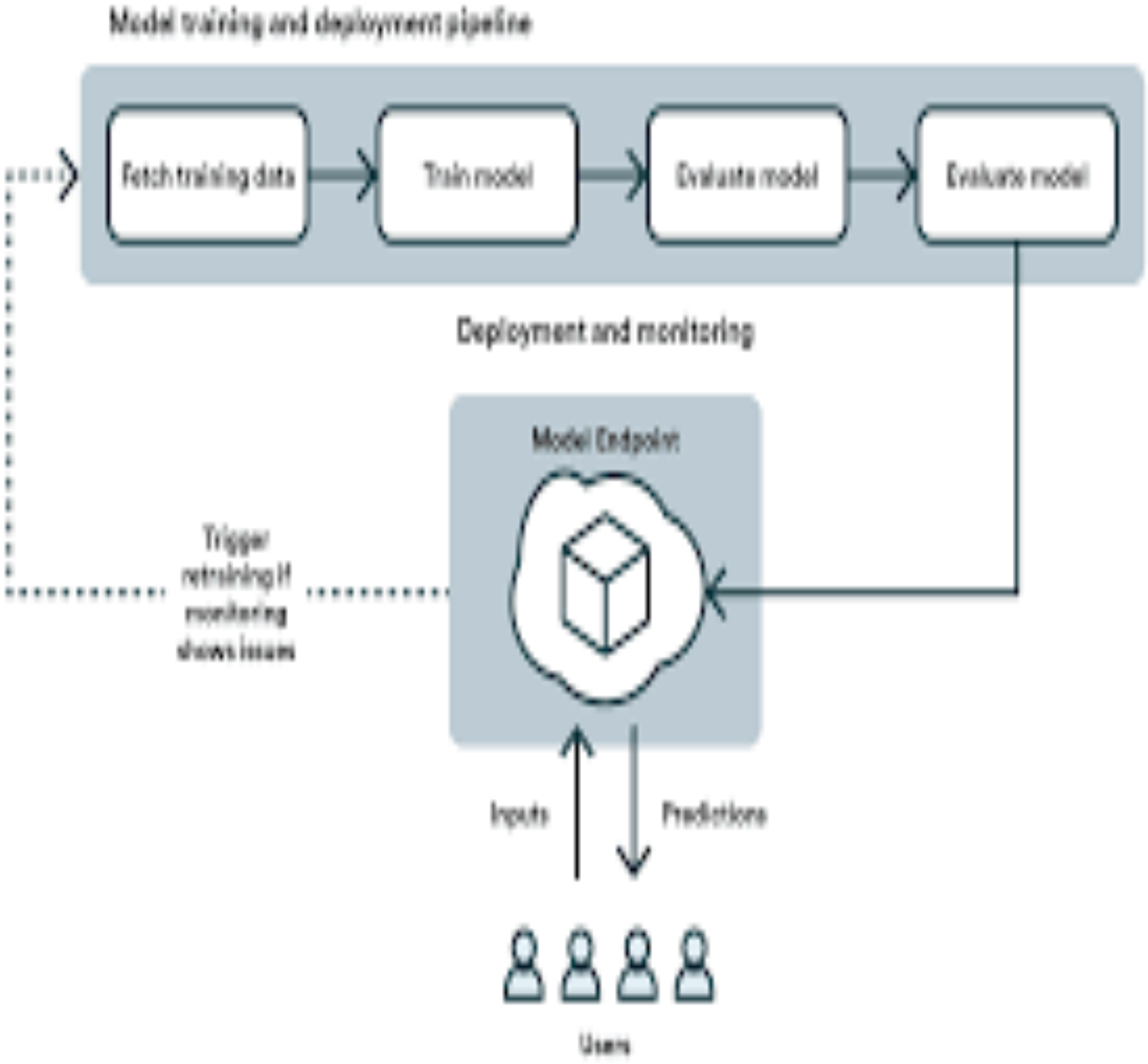

Interpretation
& Evaluation
Machine
Learning
Feature
Engineering
Preprocessing
Selection
& Cleaning
Knowledge
Patterns
& Rules
Transformed
Data
Preprocessed
Data
Target Data
Raw Data

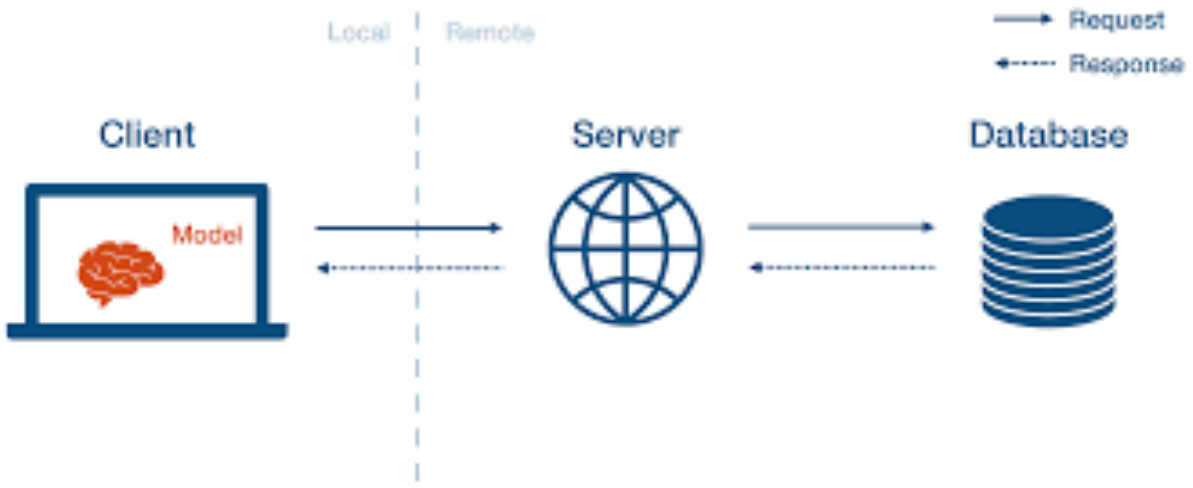
6.

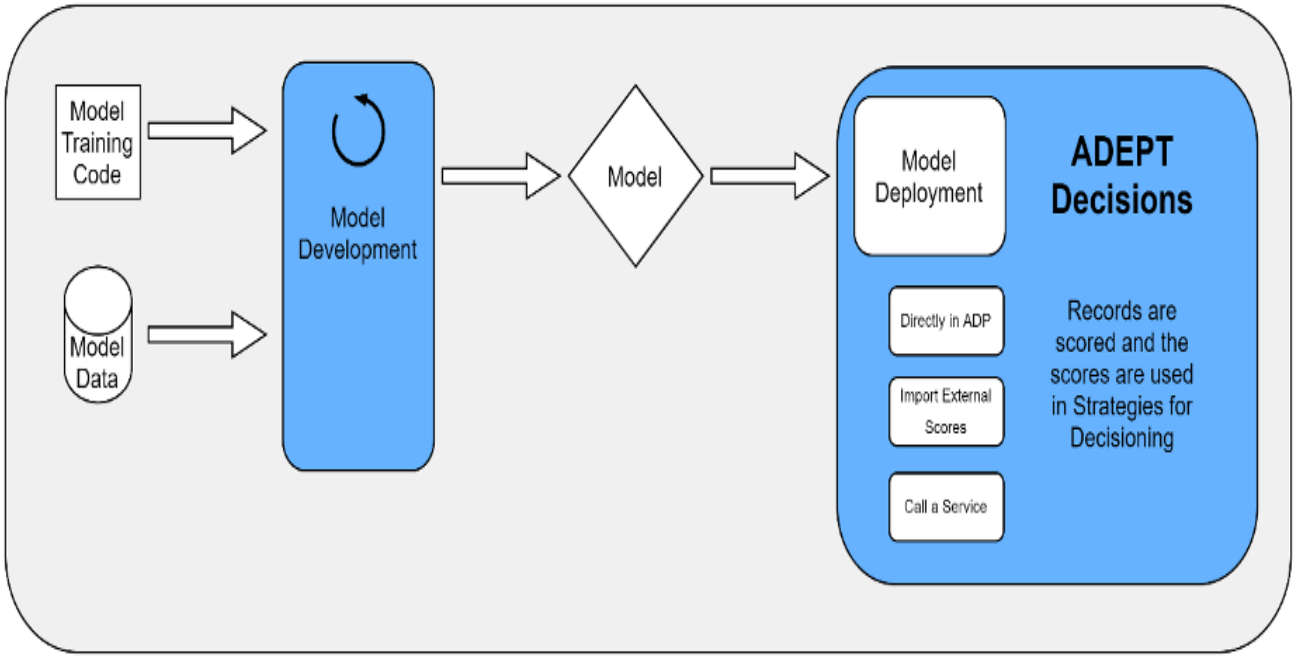

RAW
Raw data
Feature
Engineering
ML model

4.2 Model Deployment

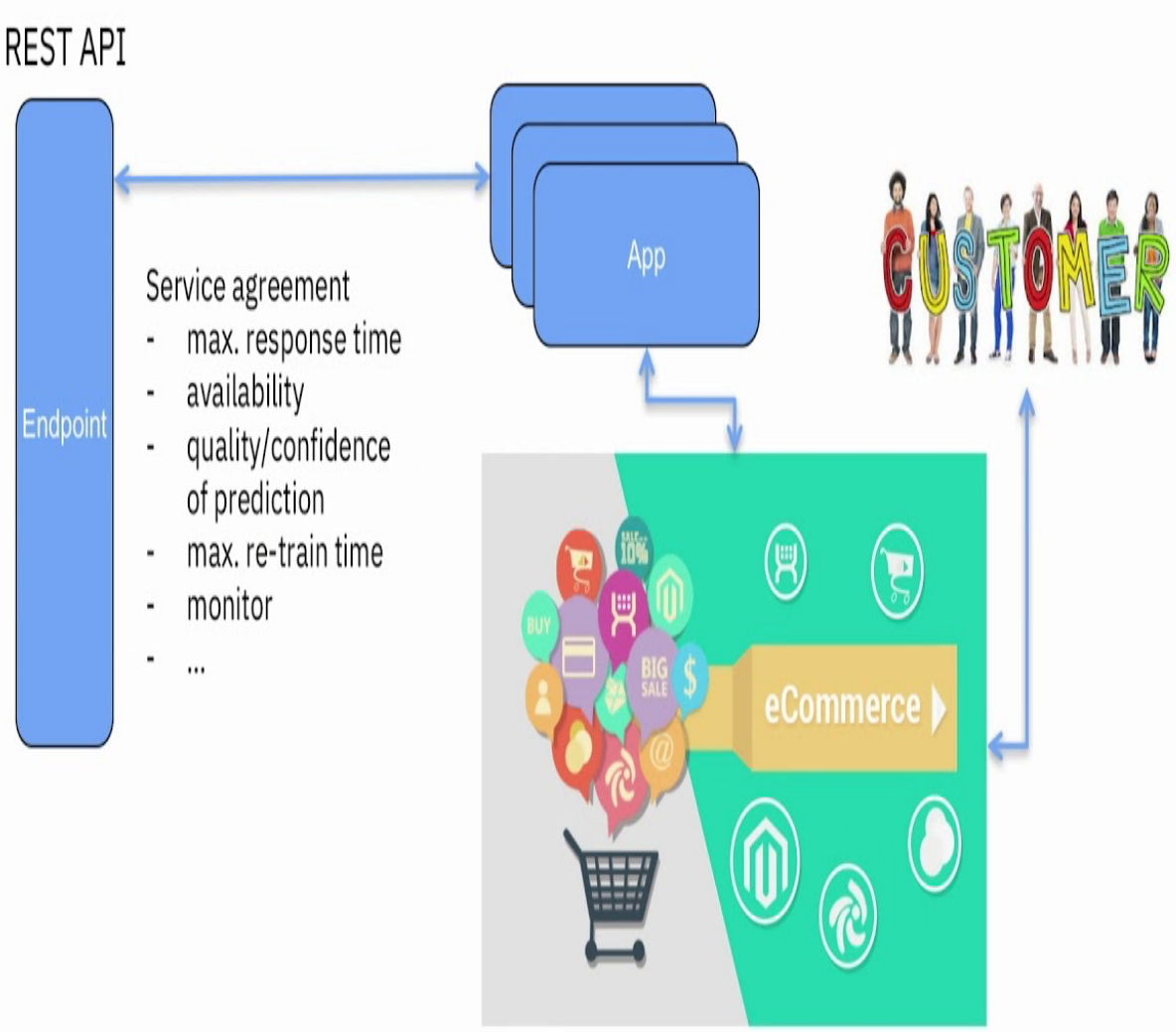



Model training and deployment pipeline
Fetch training data
Train model
Evaluate model
Evaluate model
Deployment and monitoring
Model Endpoint
Trigger retraining if monitoring shows issues
Inputs
Predictions
Users

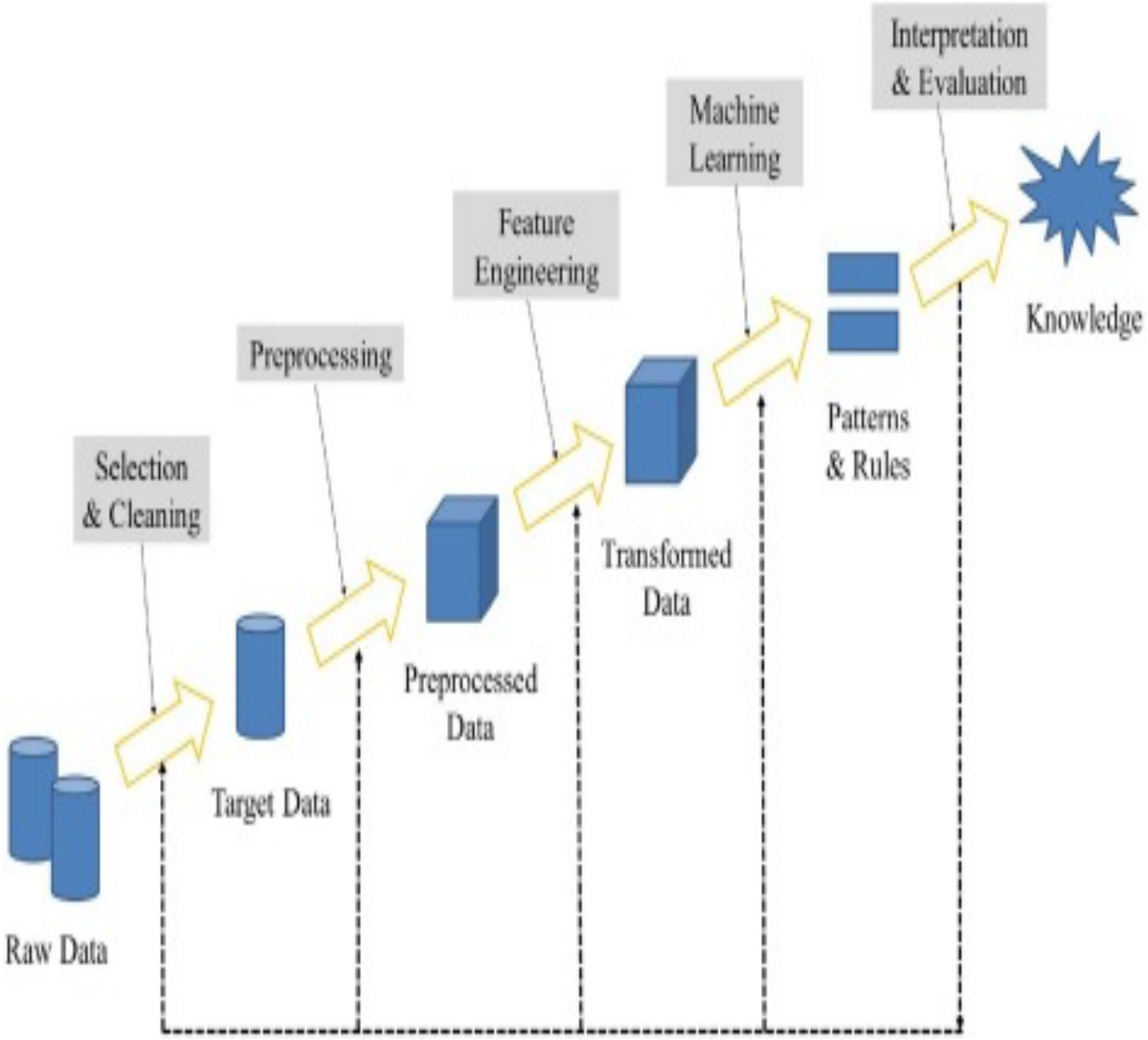

Local
Remote
Request
Response
Client
Server
Database
Model


Model Training Code
Model Data
Model Development
Model
Model Deployment
ADEPT Decisions
Directly in ADP
Import External Scores
Call a Service
Records are scored and the scores are used in Strategies for Decisioning

Production Environments – eCommerce



https://blog.paymill.com/en/keep-e-commerce-system-functioning-optimum-capacity/,
http://www.atlascorps.org/blog/customers-and-clients-the-devils-of-business-growth/


Interpretation
& Evaluation
Machine
Learning
Feature
Engineering
Preprocessing
Selection
& Cleaning
Knowledge
Patterns
& Rules
Transformed
Data
Preprocessed
Data
Target Data
Raw Data

5. Conclusion

House price prediction using linear regression is a commonly used method in the field of real estate and finance. The model uses historical data to predict future house prices based on a variety of factors such as location, size, number of bedrooms and bathrooms, and other amenities.

6. References

1. "Predicting House Prices with Linear Regression" by John Witten Auer: https://towardsdatascience.com/predicting-house-prices-with-linear-regression-5c7bfe8f7095

2. "House Prices: Advanced Regression Techniques" on Kaggle: https://www.veracode.com/security/sql-injection

3. "Regression Models for House Price Prediction" by Tim Trundle and Thorold Klein: https://www.sciencedirect.com/science/article/abs/pii/S037722171731070X

4. "A Comparative Study of Machine Learning Techniques for House Price Prediction" by Alaa Tharwat et al.: https://www.sciencedirect.com/science/article/abs/pii/S1568494620303934.