

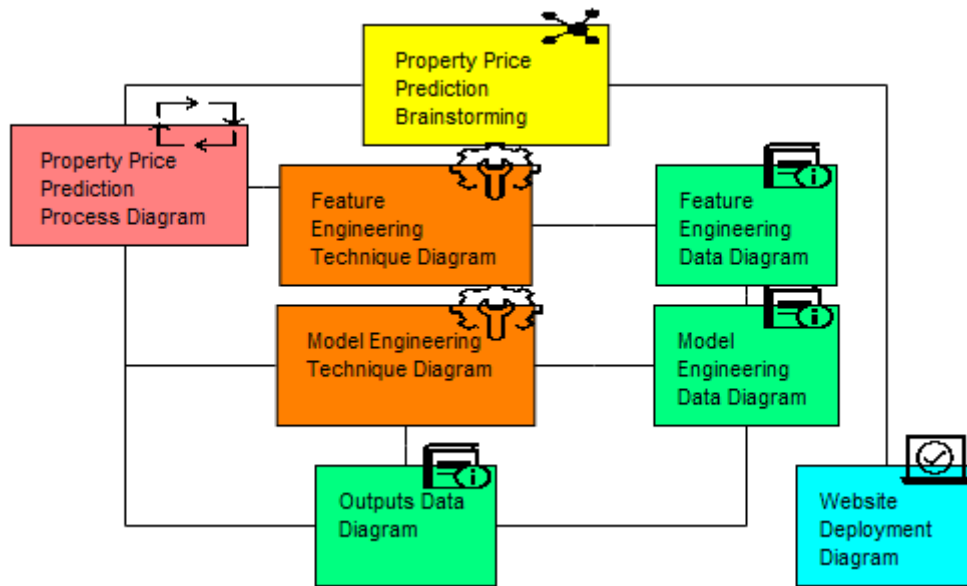
Property Price Prediction Documentation Report

Filename:	Property Price Prediction.doc
Status:	
Date:	30.9.2019
Author:	Hourieh

Table of contents

Property Price Prediction: Diagrams	3
Property Price Prediction: Brainstorming.....	4
Property Price Prediction Feature Engineering: Data	7
Property Price Prediction Model Engineering: Data	8
Property Price Prediction Report: Data	9
Property Price Prediction Website Deployment: Deployment	10
Property Price Prediction: Process	12
Property Price Prediction Feature Engineering: Technique	15
Property Price Prediction Model Engineering: Technique	16

Property Price Prediction: Diagrams



Brainstorming diagram name: Property Price Prediction Brainstorming Diagram

Diagram definition: This diagram shows high level problem, tasks, comments, operations, etc.

Data diagram name: Feature Engineering Data Diagram

Diagram definition: This diagram shows data items used or generated in feature engineering.

Data diagram name: Model Engineering Data Diagram

Diagram definition: This diagram shows data items used or generated in model engineering.

Data diagram name: Outputs Data Diagram

Diagram definition: This diagram shows all the reports expected to be generated.

Deployment diagram name: Model Training Deployment Diagram

Diagram definition: This diagram shows how to deploy the model training task.

Deployment diagram name: Website Deployment Diagram

Diagram definition: This diagram shows how to deploy the website.

Process diagram name: Property Price Prediction Process Diagram

Diagram definition: This process diagrams shows different organizations, users, order of the tasks and how tasks are related.

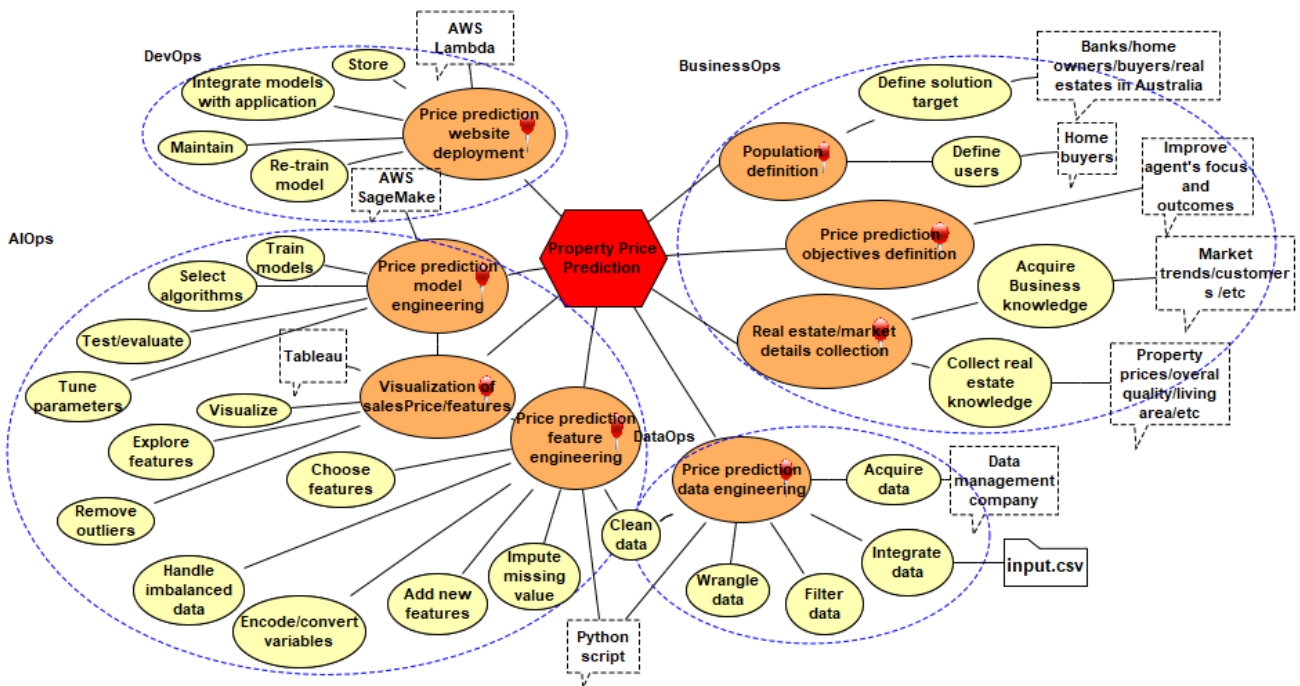
Technique diagram name: Feature Engineering Technique Diagram

Diagram definition: This diagram shows techniques used for feature engineering.

Technique diagram name: Model Engineering Technique Diagram

Diagram definition: This diagram shows techniques used for model engineering.

Property Price Prediction: Brainstorming



Comment: AWS Lambda

Comment: AWS SageMaker

Comment: Banks/home owners/buyers/real estates in Australia

Comment: Data management company

Comment: Home buyers

Comment: Improve agent's focus and outcomes

Comment: Market trends/customers/etc

Comment: Property prices/overall quality/living area/etc

Comment: Python script

Comment: Tableau

Input file name: input.csv

Operation: AIOps

Operation: BusinessOps

Operation: DataOps

Operation: DevOps

Problem: Property Price Prediction

Problem Definition: This problem involves predicting property prices using machine learning algorithms

Task Name: Acquire Business knowledge

Task definition:

Task Name: Acquire data

Task definition:

Task Name: Add new features

Task definition:

Task Name: Choose features

Task definition:

Task Name: Clean data

Task definition:

Task Name: Collect real estate knowledge

Task definition:

Task Name: Define solution target

Task definition:

Task Name: Define users

Task definition:

Task Name: Encode/convert variables

Task definition:

Task Name: Explore features

Task definition:

Task Name: Filter data

Task definition:

Task Name: Handle imbalanced data

Task definition:

Task Name: Impute missing value

Task definition:

Task Name: Integrate data

Task definition:

Task Name: Integrate models with application

Task definition:

Task Name: Maintain

Task definition:

Task Name: Population definition

Task definition:

Task Name: Price prediction data engineering
Task definition:

Task Name: Price prediction feature engineering
Task definition:

Task Name: Price prediction model engineering
Task definition:

Task Name: Price prediction objectives definition
Task definition:

Task Name: Price prediction website deployment
Task definition: In this task software engineers integrate ML models with the property price prediction website

Task Name: Re-train model
Task definition:

Task Name: Real estate/market details collection
Task definition:

Task Name: Remove outliers
Task definition:

Task Name: Select algorithms
Task definition:

Task Name: Store
Task definition:

Task Name: Test/evaluate
Task definition:

Task Name: Train models
Task definition:

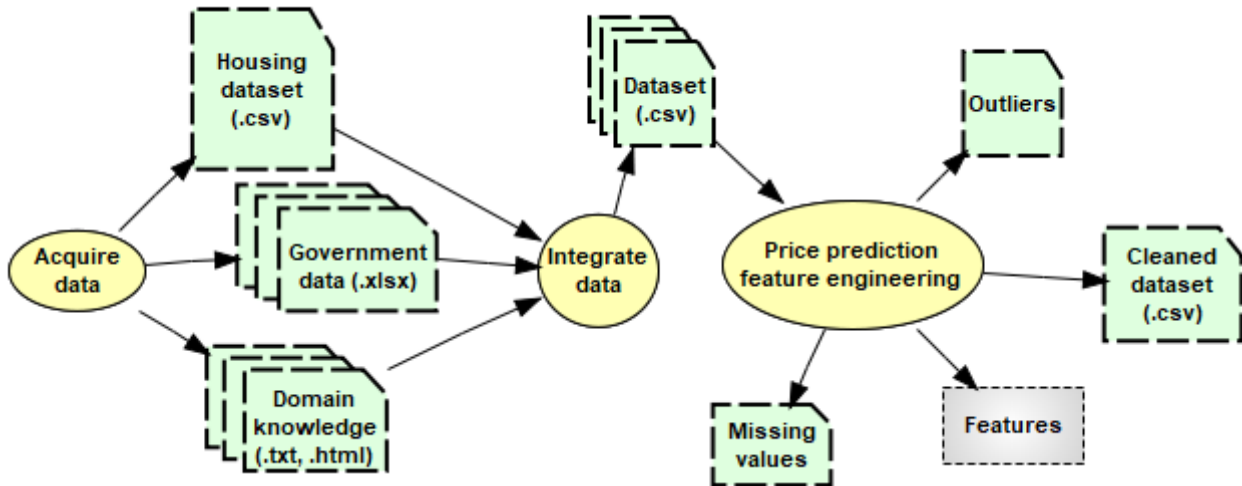
Task Name: Tune parameters
Task definition:

Task Name: Visualization of salesPrice/features
Task definition:

Task Name: Visualize
Task definition:

Task Name: Wrangle data
Task definition:

Property Price Prediction Feature Engineering: Data



Data name: Cleaned dataset (.csv)

Data name: Housing dataset (.csv)

Data name: Missing values

Data name: Outliers

Information Name: Features

Multiple data name: Dataset (.csv)

Multiple data name: Domain knowledge (.txt, .html)

Multiple data name: Government data (.xlsx)

Task Name: Acquire data

Task definition:

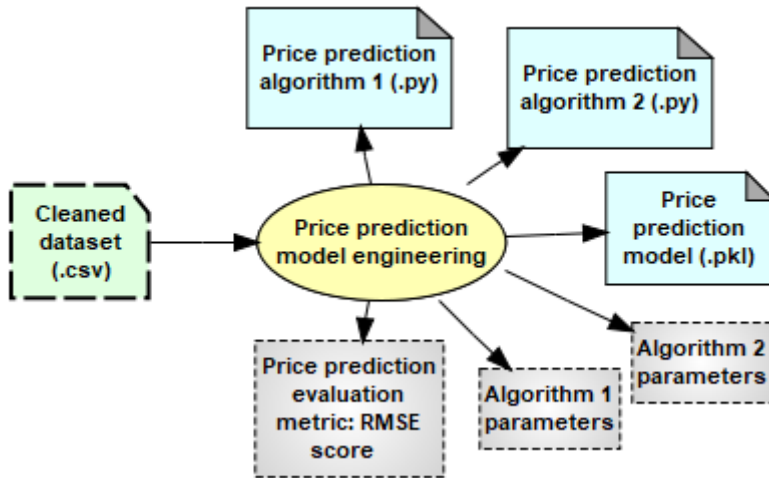
Task Name: Integrate data

Task definition:

Task Name: Price prediction feature engineering

Task definition:

Property Price Prediction Model Engineering: Data



Code file/Model name: Price prediction algorithm 1 (.py)

Code file/Model name: Price prediction algorithm 2 (.py)

Code file/Model name: Price prediction model (.pkl)

Data name: Cleaned dataset (.csv)

Information Name: Algorithm 1 parameters

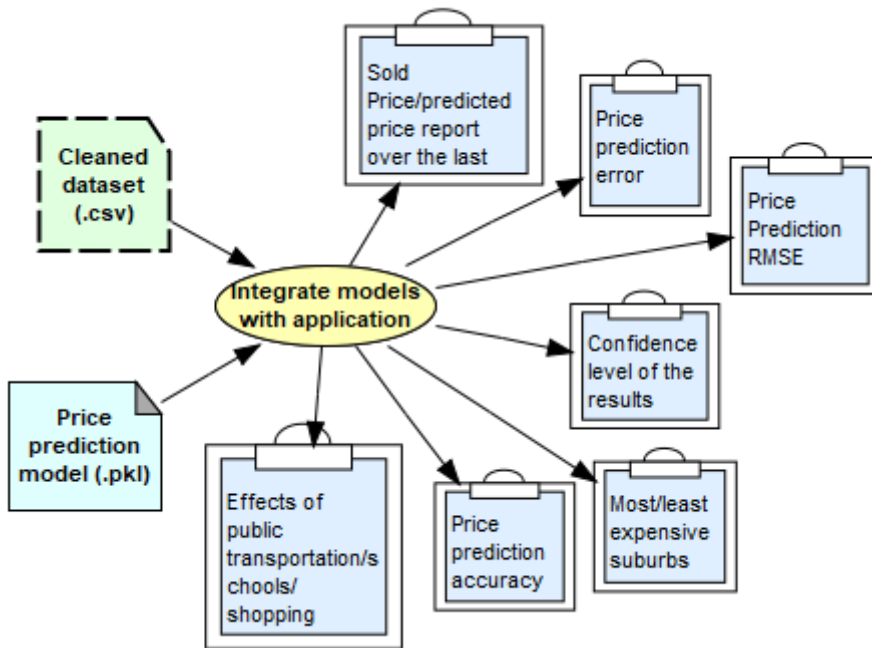
Information Name: Algorithm 2 parameters

Information Name: Price prediction evaluation metric: RMSE score

Task Name: Price prediction model engineering

Task definition:

Property Price Prediction Report: Data



Code file/Model name: Price prediction model (.pkl)

Data name: Cleaned dataset (.csv)

Report name: Confidence level of the results

Report name: Effects of public transportation/schools/ shopping centers/etc on salePrice

Report name: Most/least expensive suburbs

Report name: Price prediction accuracy

Report name: Price prediction error

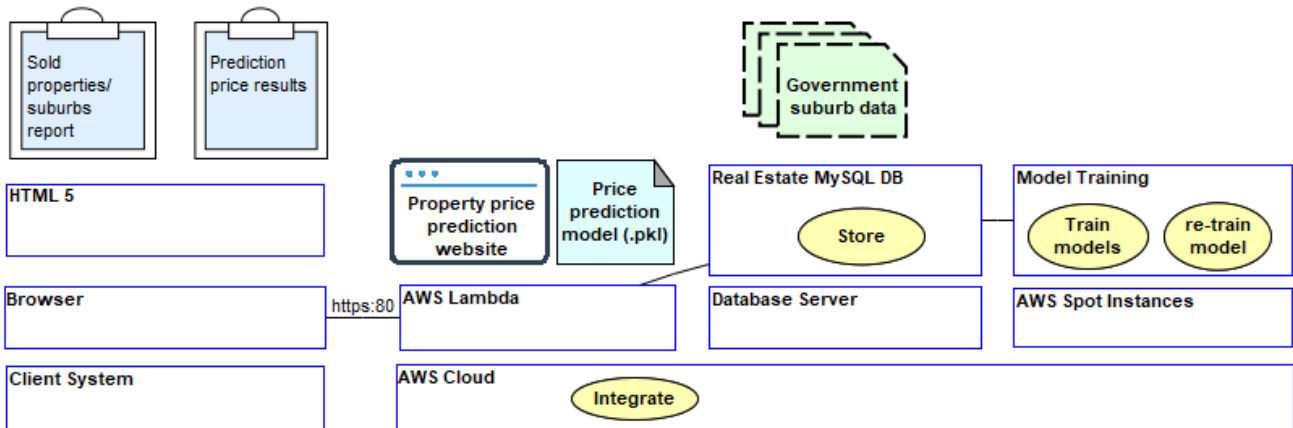
Report name: Price Prediction RMSE score

Report name: Sold Price/predicted price report over the last year

Task Name: Integrate models with application

Task definition:

Property Price Prediction Website Deployment: Deployment



Application Name: Property price prediction website

Code file/Model name: Price prediction model (.pkl)

Component Name: AWS Cloud

Component Name: AWS Lambda

Component Name: AWS Spot Instances

Component Name: Browser

Component Name: Client System

Component Name: Database Server

Component Name: HTML 5

Component Name: Model Training

Component Name: Real Estate MySQL DB

Multiple data name: Government suburb data

Report name: Prediction price results

Report name: Sold properties/ suburbs report

Task Name: Integrate

Task definition:

Task Name: re-train model

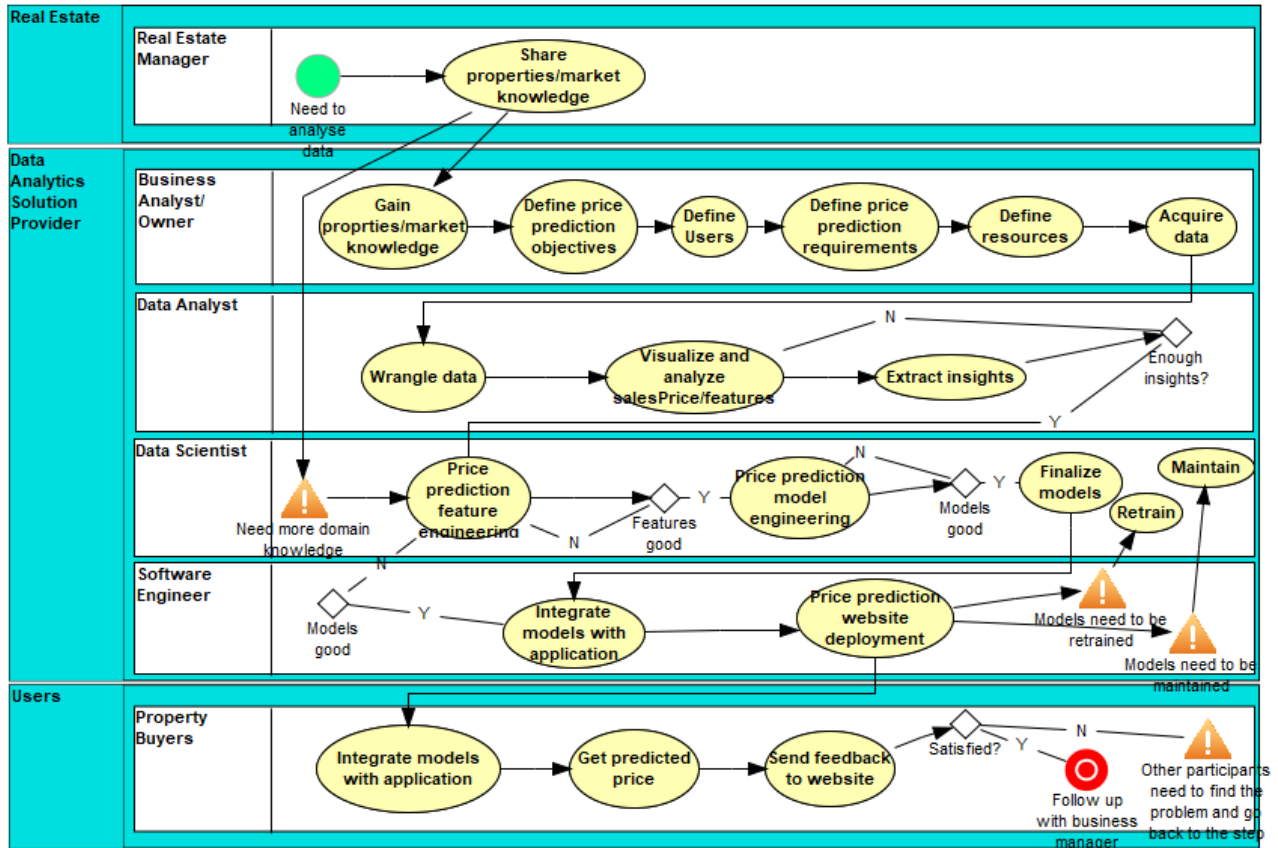
Task definition:

Task Name: Store

Task definition:

Task Name: Train models
Task definition:

Property Price Prediction: Process



Alert: Models need to be maintained

Alert: Models need to be retrained

Alert: Need more domain knowledge

Alert: Other participants need to find the problem and go back to the step

Condition name: Enough insights?

Condition definition:

Condition name: Features good enough?

Condition definition:

Condition name: Models good enough?

Condition definition:

Condition name: Models good enough?

Condition definition:

Condition name: Satisfied?

Condition definition:

End: Follow up with business manager

Organization name: Data Analytics Solution Provider
Organization definition:

Organization name: Real Estate
Organization definition:

Organization name: Users
Organization definition:

Participant name: Business Analyst/ Owner
Participant definition:

Participant name: Data Analyst
Participant definition:

Participant name: Data Scientist
Participant definition:

Participant name: Property Buyers
Participant definition:

Participant name: Real Estate Manager
Participant definition:

Participant name: Software Engineer
Participant definition:

Start: Need to analyse data

Task Name: Acquire data
Task definition:

Task Name: Define price prediction objectives
Task definition:

Task Name: Define price prediction requirements
Task definition:

Task Name: Define resources
Task definition:

Task Name: Define Users
Task definition:

Task Name: Extract insights
Task definition:

Task Name: Finalize models
Task definition:

Task Name: Gain properties/market knowledge
Task definition:

Task Name: Get predicted price
Task definition:

Task Name: Integrate models with application
Task definition:

Task Name: Integrate models with application
Task definition:

Task Name: Maintain
Task definition:

Task Name: Price prediction feature engineering
Task definition:

Task Name: Price prediction model engineering
Task definition:

Task Name: Price prediction website deployment
Task definition:

Task Name: Retrain
Task definition:

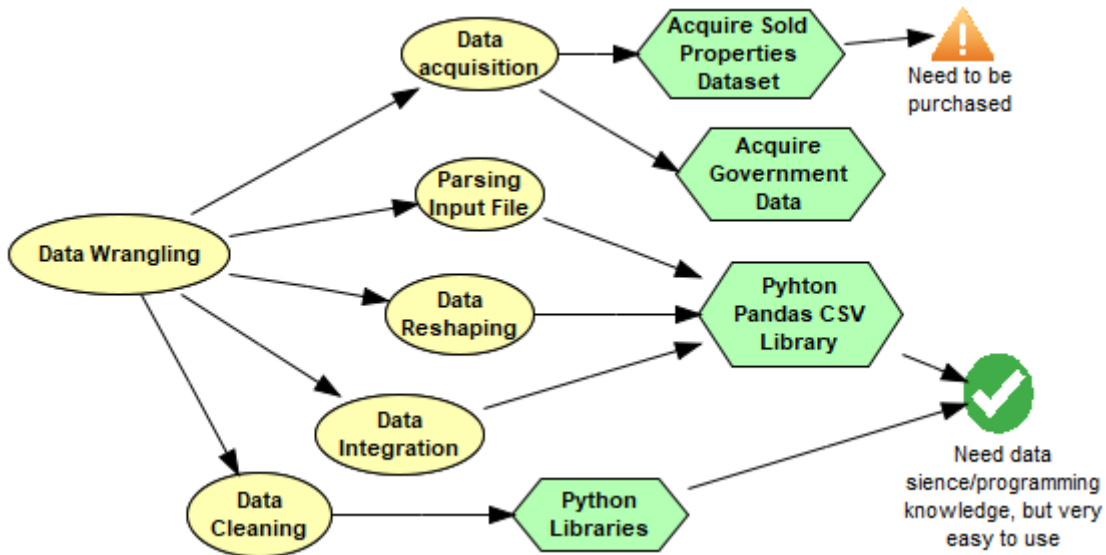
Task Name: Send feedback to website
Task definition:

Task Name: Share properties/market knowledge
Task definition: Domain expert share domain knowledge with the business analyst/manager

Task Name: Visualize and analyze salesPrice/features
Task definition:

Task Name: Wrangle data
Task definition:

Property Price Prediction Feature Engineering: Technique



Alert: Need to be purchased

Success: Need data sience/programming knowledge, but very easy to use

Task Name: Data acquisition

Task definition:

Task Name: Data Cleaning

Task definition:

Task Name: Data Integration

Task definition:

Task Name: Data Reshaping

Task definition:

Task Name: Data Wrangling

Task definition:

Task Name: Parsing Input File

Task definition:

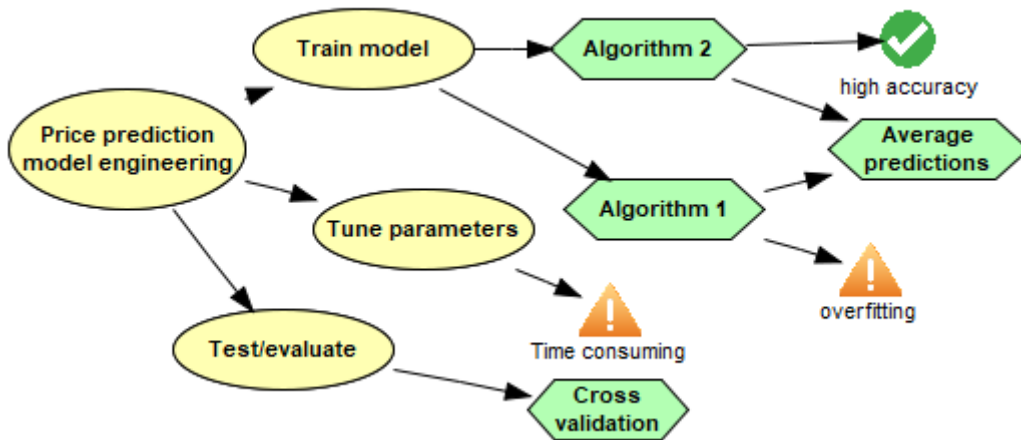
Technique name: Acquire Government Data

Technique name: Acquire Sold Properties Dataset

Technique name: Pyhton Pandas CSV Library

Technique name: Python Libraries

Property Price Prediction Model Engineering: Technique



Alert: overfitting

Alert: Time consuming

Success: high accuracy

Task Name: Price prediction model engineering

Task definition:

Task Name: Test/evaluate

Task definition:

Task Name: Train model

Task definition:

Task Name: Tune parameters

Task definition:

Technique name: Algorithm 1

Technique name: Algorithm 2

Technique name: Average predictions

Technique name: Cross validation