

Capstone Project

DS Bootcamp, 2024

Task Overview

Intro:

- Notebook is becoming one of the most essential items in our daily lives. The overwhelming prices may vary according to the notebooks' brand, specification, and others. In case a new rising factory produces new notebooks to the market, it would be nice to estimate the base price for the newly produced notebook in order to help the company set the prices for each of their products and indicate the factors which affect the price.

Task:

- Build a model to predict notebook prices based on given data.

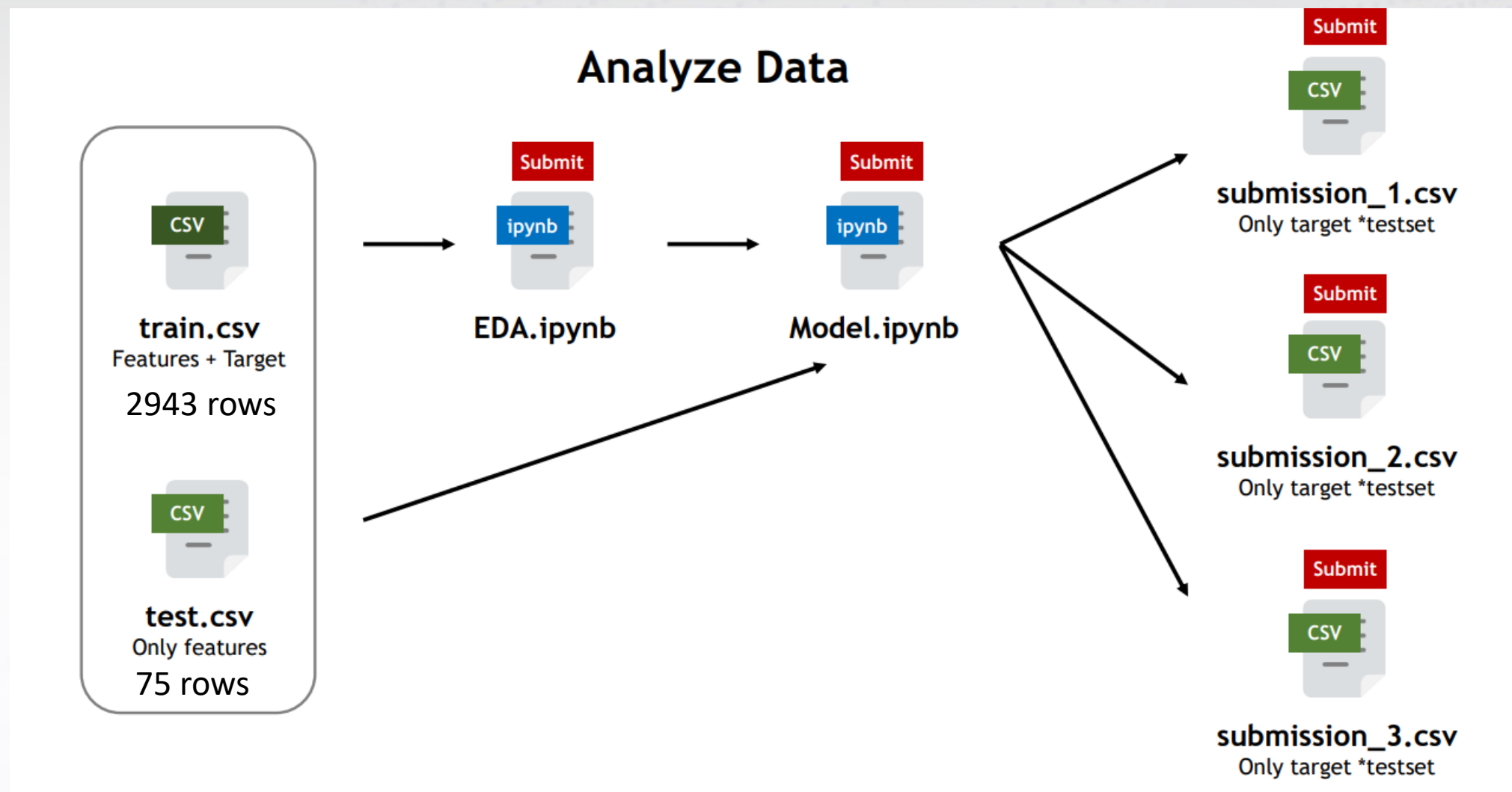


Data Overview

The current dataset consists of 10 columns and 2,943 observations. The columns' name and description are as follows:

	feature	meaning
1	productID	ID of each notebook
2	model	Notebook model
3	color	Color of each notebook
4	gpu	Graphics Processing Unit
5	cpu	Central Processing Unit
6	ram	Random Access Memory
7	harddisk	capacity of an internal storage
8	OS	Operating System
9	special_features	Special unique feature of each notebook model
10	price	Notebook price

Workflow



Submission

ผู้เข้าเรียนต้องทำการวิเคราะห์ข้อมูล และสร้าง machine learning model เพื่อหา notebook price โดยต้องจัดส่งไฟล์ดังต่อไปนี้ ภายใน **21 ตุลาคม 67** ไปที่ e-mail: visai.ai.academy@gmail.com

1. Python notebook จำนวน 2 ไฟล์

- 1) Clean data and EDA – เขียนคำอธิบายและ insights ไว้ใน notebook
- 2) Modeling – เขียนคำอธิบายและ สรุปผล ไว้ใน notebook

2. Notebook price ของ test set โดยส่งคำตอบใน column “price” ตามรูปแบบของ submission form [submission_template.csv](#) ไม่เกินจำนวน 3 ไฟล์ (อาจมาจากหลาย model) โดยตั้งชื่อไฟล์ดังต่อไปนี้

- 1) submission_1.csv
- 2) submission_2.csv
- 3) submission_3.csv

Submission

Capstone Project เป็นงานรายบุคคล โดยมีคะแนนเต็ม 100 ซึ่งคิดเป็น 40% (ผู้เรียนเดี่ยว) และ 10% (ผู้เรียนกลุ่ม) ของคะแนนสำหรับการผ่านหลักสูตรทั้งหมด หากทีมงานตรวจสอบว่าเกิดการลอกกัน หรือการโกงคำตอบ ทีมงานจะปรับให้ผู้เรียนไม่ผ่านหลักสูตร (และแจ้งผู้ประสานงานกรณีบริษัท)

- **Score ในส่วน python notebook**

- 1) EDA: 20 points
- 2) Clean and manipulation: 20 points
- 3) Modeling: 20 points

- **Score ในส่วนของ test set (ใช้คะแนนที่สูงสุดจาก submission.csv) *ต้องสามารถรันคำตอบได้จาก notebook ที่ส่งมา**

- 1) 10 points: RMSE is lower than 450
- 2) 20 Points: RMSE is lower than 400
- 3) 30 Points: RMSE is lower than 350
- 4) 40 points: RMSE is lower than 300