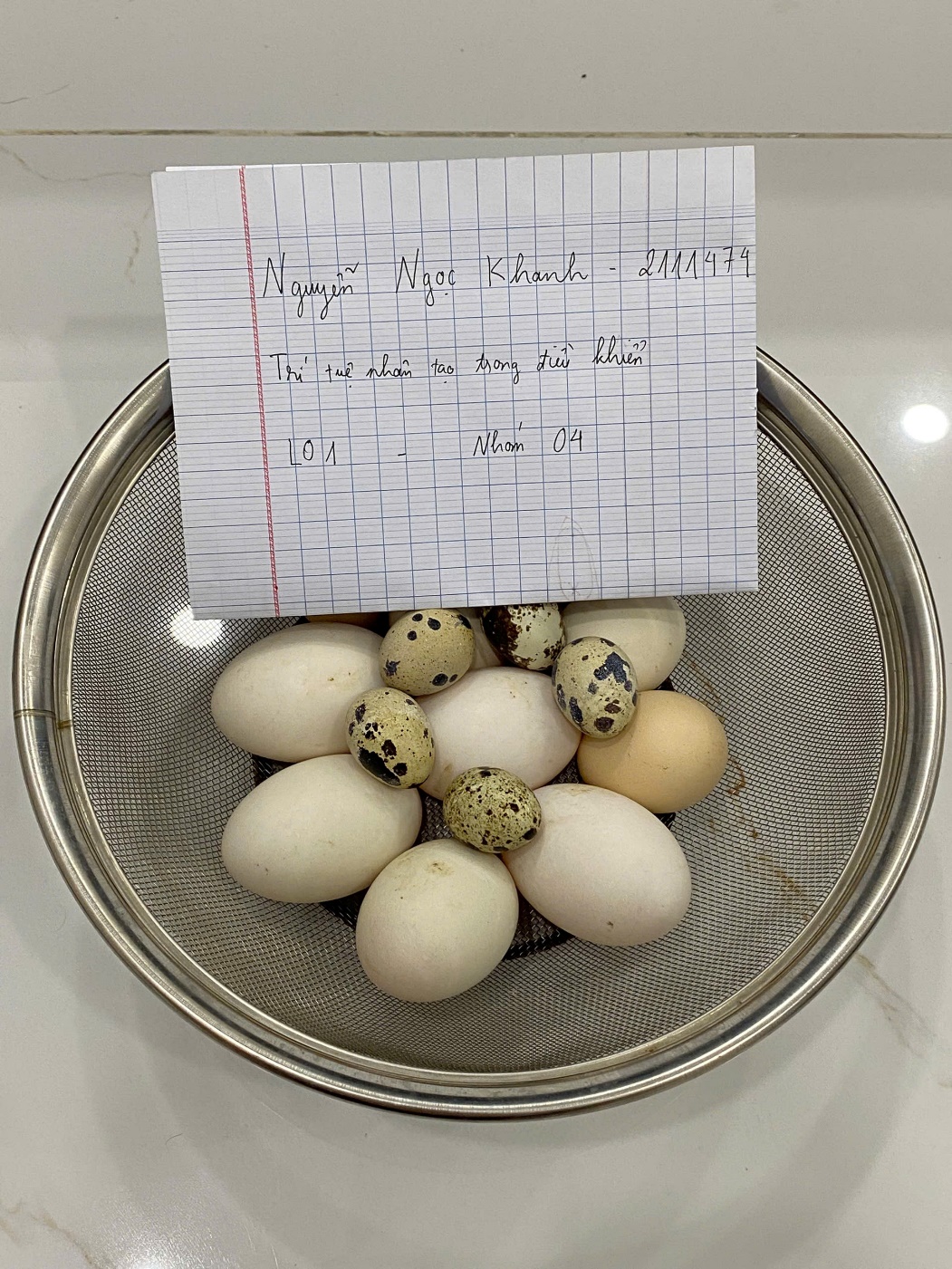
**Báo cáo Homework 3**

**Nguyễn Ngọc Khanh – 2111474**

1. **Tổng quan**

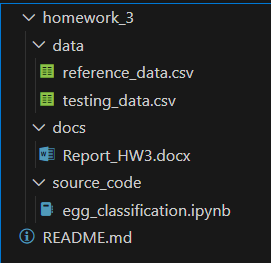
Đề bài Homework 3: Thu thập dữ liệu thực tế: tối thiểu 3 class cho 3 thuộc tính, sử dụng classification bằng phương pháp đại số, bằng K-NN.

Em lựa chọn 3 class lần lượt là các loại trứng: trứng gà ri (Chicken Egg), trứng vịt (Duck Egg) và trứng chim cút (Quail Egg). Ba thuộc tính lần lượt là chiều dài (Length, đơn vị cm), chiều rộng (Width, đơn vị cm) và khối lượng (Weight, đơn vị gam).



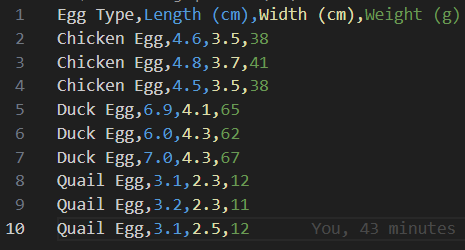
Em chọn 3 bộ thuộc tính tương ứng với 3 quả trứng mỗi loại để làm dữ liệu tham chiếu (đã được gán nhãn) cho phương pháp đại số và K-NN. Với dữ liệu dùng để kiểm thử, phân loại, em chọn 2 bộ thuộc tính tương ứng với 2 quả trứng mỗi loại.

Cấu trúc thư mục gồm các folder:

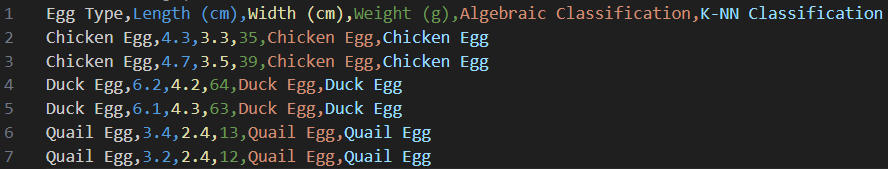


- data: chứa dữ liệu tham chiếu và dữ liệu kiểm thử

+ Dữ liệu tham chiếu được lưu ở file “reference\_data.csv”, gồm các nội dung như loại trứng (class) và các thuộc tính chiều dài, chiều rộng, khối lượng:



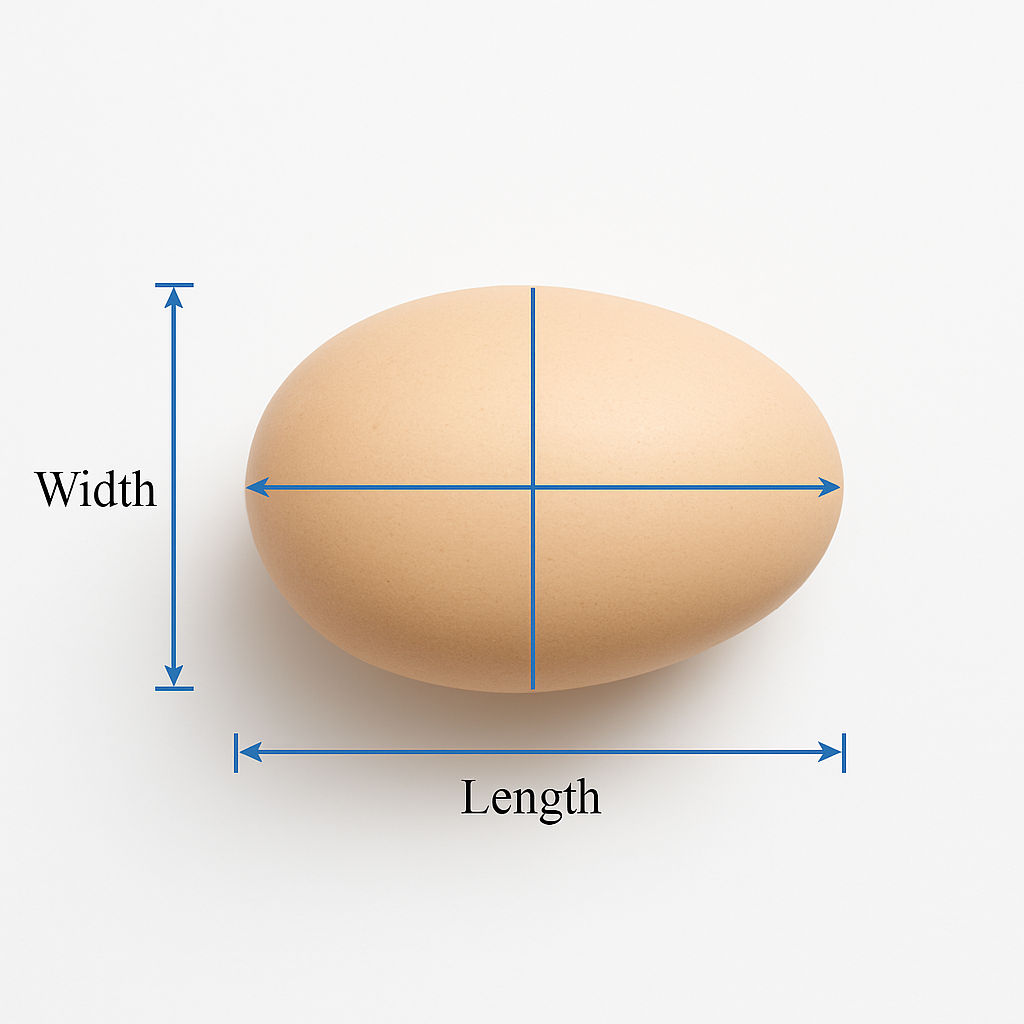
+ Dữ liệu kiểm thử được lưu ở file “testing\_data.csv”, gồm các nội dung như nhãn của loại trứng, các thuộc tính và kết quả phân loại của phương pháp đại số và K-NN:



- docs: chứa báo cáo, kết quả phân loại.

- source code: chứa mã nguồn của các phương pháp phân loại đại số và K-NN.

Em sẽ đo đạc 3 thuộc tính và ghi kết quả vào các file trên. Thuộc tính khối lượng sẽ được đo bằng cân điện tử, thuộc tính chiều dài và chiều rộng được đo như sau:



1. **Phân loại**
2. **Phương pháp đại số**

Em phân loại bằng phương pháp đại số sử dụng hàm LogisticRegression trong module sklearn.linear\_model với 1000 vòng lặp tối đa để hội tụ.

Kết quả classification bằng phương pháp đại số như sau:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Data point | Class Label | Algebraic Classification Result | Length (cm) | Width (cm) | Weight (g) |
| 1 | Chicken Egg | Chicken Egg | 4.3 | 3.3 | 35 |
| 2 | Chicken Egg | Chicken Egg | 4.7 | 3.5 | 39 |
| 3 | Duck Egg | Duck Egg | 6.2 | 4.2 | 64 |
| 4 | Duck Egg | Duck Egg | 6.1 | 4.3 | 63 |
| 5 | Quail Egg | Quail Egg | 3.4 | 2.4 | 13 |
| 6 | Quail Egg | Quail Egg | 3.2 | 2.4 | 12 |

Như vậy, phương pháp đại số đã phân loại đúng tất cả 6 điểm dữ liệu.

1. **Phương pháp K-NN**

Em phân loại bằng phương pháp K-NN sử dụng hàm KneighborsClassifier trong module sklearn.neighbors với 3 nearest neighbors.

Kết quả classification bằng phương pháp K-NN như sau:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Data point | Class Label | K-NN Classification Result | Length (cm) | Width (cm) | Weight (g) |
| 1 | Chicken Egg | Chicken Egg | 4.3 | 3.3 | 35 |
| 2 | Chicken Egg | Chicken Egg | 4.7 | 3.5 | 39 |
| 3 | Duck Egg | Duck Egg | 6.2 | 4.2 | 64 |
| 4 | Duck Egg | Duck Egg | 6.1 | 4.3 | 63 |
| 5 | Quail Egg | Quail Egg | 3.4 | 2.4 | 13 |
| 6 | Quail Egg | Quail Egg | 3.2 | 2.4 | 12 |

Như vậy, phương pháp K-NN đã phân loại đúng tất cả 6 điểm dữ liệu.