

Bài toán: Cho tập dữ liệu với các thông tin của khách hàng và họ có mua hàng hay không.

Input:

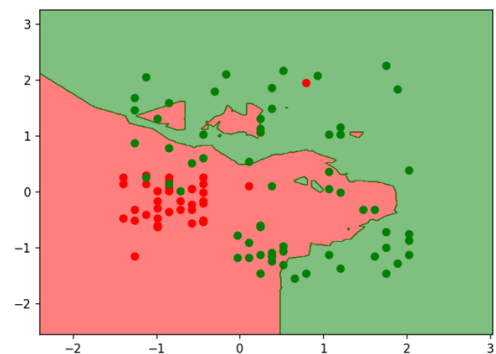
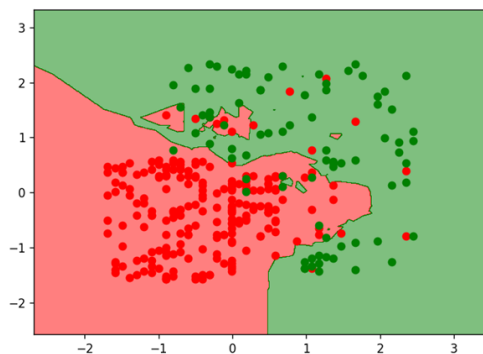
+ Độ tuổi (Age)

+ Mức lương ước đoán (Estimated Salary)

Output: Dự đoán khách hàng sẽ mua hàng hay không?

Phương pháp: Sử dụng thuật toán K Nearest Neighbors

Visualization Test và Confusion Matrix



Training Set vs Testing Set (k=3)

Confusion Matrix:

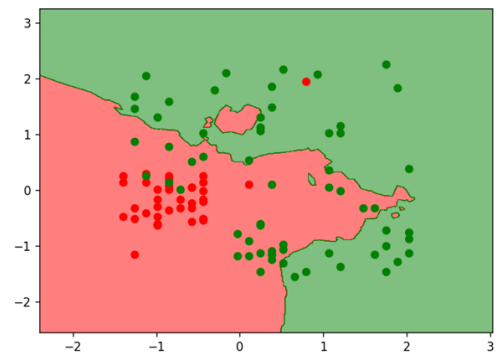
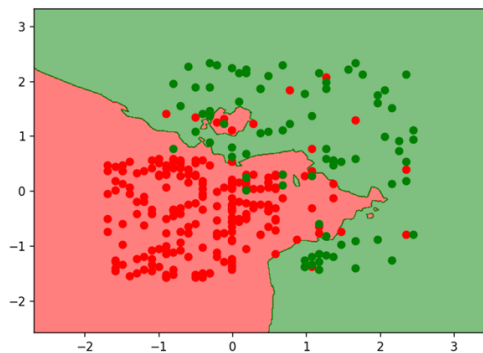
Training Set vs Testing Set:

[[207 12]

[10 71]]

[[37 1]

[30 32]]

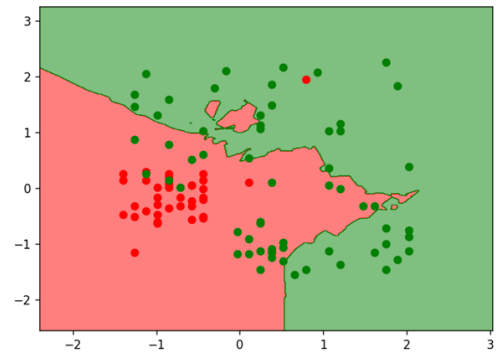
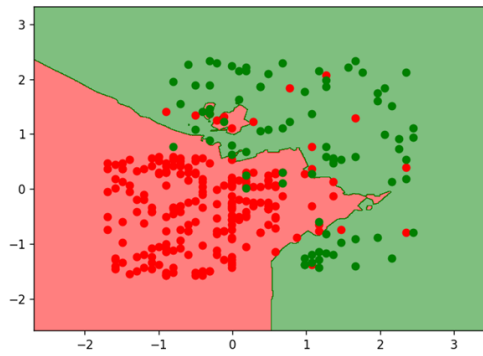


Training Set vs Testing Set (k=5)

Confusion Matrix:

Training Set vs Testing Set:

```
[[205 14]
 [ 9 72]]
[[37 1]
 [27 35]]
```

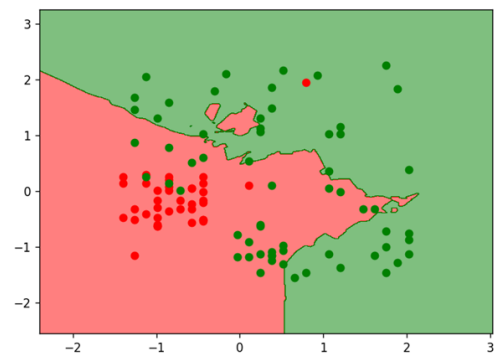
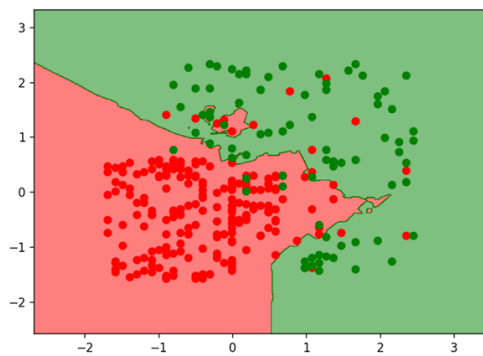


Training Set vs Testing Set (k=7)

Confusion Matrix:

Training Set vs Testing Set:

```
[[205 14]
 [11 70]]
[[37 1]
 [27 35]]
```



Training Set vs Testing Set (k=9)

Confusion Matrix:

Training Set vs Testing Set:

```
[[203 16]
 [ 8 73]]
```

[[37 1]
[28 34]]

Tỉ lệ điểm dữ liệu phân loại sai:

	model	train	test
0	3.0	0.073333	0.31
1	5.0	0.076667	0.28
2	7.0	0.083333	0.28
3	9.0	0.080000	0.29