Họ tên: Hà Huy Sơn

Lớp: 59K2-CNTT

**BÀI TẬP THỰC HÀNH 3: K-Nearest Neighbors**

**Câu 1:** Hãy sử dụng giải thuật KNN để xác định thể loại cho các phim. Với K=1, 3, 5.

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên phim** | **Số lượng cú đá** | **Số lượng nụ hôn** | **Loại phim** |
| Californiaman | 3 | 104 | Tình cảm |
| He isn’t really into dudes | 2 | 100 | Tình cảm |
| Beautiful Woman | 1 | 81 | Tình cảm |
| Kevin Longblade | 101 | 10 | Hành động |
| Robo Slayer 3000 | 99 | 5 | Hành động |
| Amped II | 98 | 2 | Hành động |
| Kiss kick | 18 | 90 | ? |
| Last blood | 23 | 6 | ? |
| Lucky men | 8 | 10 | ? |

Gọi **d**là euclidean distance

Kí hiệu:

A1 = Californiaman

A2 = He isn’t really into dudes

A3 = Beautiful Woman

A4 = Kevin Longblade

A5 = Robo Slayer 3000

A6 = Amped II

* **Kiss kick**

d(Kiss kick, A1) = = 20.518

d(Kiss kick, A2) = = 18.867

d(Kiss kick, A3) = = 19.235

d(Kiss kick, A4) = = 115.277

d(Kiss kick, A5) = = 117.413

d(Kiss kick, A6) = = 118.928

+ Giả sử K = 1 thì sẽ có một láng giềng gần nhất là He isn’t really into dudes (A2). Theo thuật toán KNN sẽ lấy loại phim nào trong các láng giềng gần nhất để làm loại phim cho phim cần được xác định loại (d(Kiss kick, A2) nhỏ nhất) => Phim cần xác định thuộc thể loại phim Tình cảm.

+ Giả sử K = 3 thì sẽ có 3 láng giềng gần nhất là He isn’t really into dudes (A2), Beautiful Woman (A3) và Californiaman (A1). Vì trong 3 láng giềng gần nhất đều có 3 phim đều thuộc thể loại phim Tình cảm => Phim cần xác định thuộc thể loại phim Tình cảm.

+ Giả sử K = 5 thì sẽ có 5 láng giềng gần nhất là He isn’t really into dudes (A2), Beautiful Woman (A3), Californiaman (A1), Kevin Longblade (A4) và Robo Slayer 3000 (A5). Vì trong 5 láng giềng gần nhất có 3 láng giềng thuộc thể loại phim Tình cảm và có 2 láng giềng thuộc thể loại phim Hành động => Phim cần xác định thuộc thể loại phim Tình cảm.

* **Last blood**

d(Last blood, A1) = = 100.019

d(Last blood, A2) = = 96.317

d(Last blood, A3) = = 78.160

d(Last blood, A4) = = 78.102

d (Last blood, A5) = = 76.006

d(Last blood, A6) = = 75.106

+ Giả sử K = 1 thì sẽ có một láng giềng gần nhất là Amped II (A6). Theo thuật toán KNN sẽ lấy loại phim nào trong các láng giềng gần nhất để làm loại phim cho phim cần được xác định loại (d(Last blood, A6) nhỏ nhất) => Phim cần xác định thuộc thể loại phim Hành động.

+ Giả sử K = 3 thì sẽ có 3 láng giềng gần nhất là Amped II (A6), Robo Slayer 3000 (A5) và Kevin Longblade (A4). Vì trong 3 láng giềng gần nhất đều có 3 phim đều thuộc thể loại phim Hành động => Phim cần xác định thuộc thể loại phim Hành động.

+ Giả sử K = 5 thì sẽ có 5 láng giềng gần nhất là Amped II (A6), Robo Slayer 3000 (A5) và Kevin Longblade (A4), Beautiful Woman (A3) và He isn’t really into dudes (A2). Vì trong 5 láng giềng gần nhất có 3 láng giềng thuộc thể loại phim Hành động và có 2 láng giềng thuộc thể loại phim Tình cảm => Phim cần xác định thuộc thể loại phim Hành động.

* **Lucky men**

d(Lucky men, A1) = = 94.132

d(Lucky men, A2) = = 90.199

d(Lucky men, A3) = = 71.344

d(Lucky men, A4) = = 93

d(Lucky men, A5) = = 91.137

d(Lucky men, A6) = = 90.354

+ Giả sử K = 1 thì sẽ có một láng giềng gần nhất là Beautiful Woman (A3). Theo thuật toán KNN sẽ lấy loại phim nào trong các láng giềng gần nhất để làm loại phim cho phim cần được xác định loại (d(Lucky men, A3) nhỏ nhất) => Phim cần xác định thuộc thể loại phim Tình cảm.

+ Giả sử K = 3 thì sẽ có 3 láng giềng gần nhất là Beautiful Woman (A3), He isn’t really into dudes (A2) và Amped II (A6). Vì trong 3 láng giềng gần nhất có 2 láng giềng thuộc thể loại phim Tình cảm và 1 láng giềng thuộc thể loại phim Hành động => Phim cần xác định thuộc thể loại phim Tình cảm.

+ Giả sử K = 5 thì sẽ có 5 láng giềng gần nhất là Beautiful Woman (A3), He isn’t really into dudes (A2) và Amped II (A6), Robo Slayer 3000 (A5) và Kevin Longblade (A4). Vì trong 5 láng giềng gần nhất có 3 láng giềng thuộc thể loại phim Hành động và có 2 láng giềng thuộc thể loại phim Tình cảm => Phim cần xác định thuộc thể loại phim Hành động.

Kết luận:

TH: K = 1 và K = 3

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên phim** | **Số lượng cú đá** | **Số lượng nụ hôn** | **Loại phim** |
| Californiaman | 3 | 104 | Tình cảm |
| He isn’t really into dudes | 2 | 100 | Tình cảm |
| Beautiful Woman | 1 | 81 | Tình cảm |
| Kevin Longblade | 101 | 10 | Hành động |
| Robo Slayer 3000 | 99 | 5 | Hành động |
| Amped II | 98 | 2 | Hành động |
| Kiss kick | 18 | 90 | Tình cảm |
| Last blood | 23 | 6 | Hành động |
| Lucky men | 8 | 10 | Tình cảm |

TH: K = 5

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên phim** | **Số lượng cú đá** | **Số lượng nụ hôn** | **Loại phim** |
| Californiaman | 3 | 104 | Tình cảm |
| He isn’t really into dudes | 2 | 100 | Tình cảm |
| Beautiful Woman | 1 | 81 | Tình cảm |
| Kevin Longblade | 101 | 10 | Hành động |
| Robo Slayer 3000 | 99 | 5 | Hành động |
| Amped II | 98 | 2 | Hành động |
| Kiss kick | 18 | 90 | Tình cảm |
| Last blood | 23 | 6 | Hành động |
| Lucky men | 8 | 10 | Hành động |