The following are the review documents of a movie which are used as training sample for naïve Bayes classifier. Classify the document with text “I Hated poor Acting “into class + or -.

|  |  |  |
| --- | --- | --- |
| Doc | Text | Class |
| 1 | I loved the movie | + |
| 2 | I hated the movie | - |
| 3 | A great movie. Good movie | + |
| 4 | Poor acting | - |
| 5 | Great acting. A good movie | + |

**Key:**

Words: I, loved, the, movie, hated, a, great, good, poor, acting

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Doc | I | Loved | The | Movie | Hated | A | Great | Good | Poor | Acting | class |
| 1 | 1 | 1 | 1 | 1 |  |  |  |  |  |  | + |
| 2 | 1 |  | 1 | 1 | 1 |  |  |  |  |  | - |
| 3 |  |  |  | 2 |  | 1 | 1 | 1 |  |  | + |
| 4 |  |  |  |  |  |  |  |  | 1 | 1 | - |
| 5 |  |  |  | 1 |  | 1 | 1 | 1 |  | 1 | + |

Compute p(+|doc)= p(I|+)p(Hated|+)p(poor|+)p(Acting|+)p(+)

P(+)=3/5=0.6 p(wk|+)=

p(I|+)=(1+1)/(14+10)=0.0833

p(Hated|+)=(0+1)/ (14+10)=0.0417

p(poor|+)=(0+1)/ (14+10)= 0.0417

p(Acting|+)=(1+1)/ (14+10)= 0.0833

Compute p(-|doc)= p(I|-)p(Hated|-)p(poor|-)p(Acting|-)p(-)

P(-)=2/5=0.4 p(wk|+)=

p(I|-)=(1+1)/(6+10)=0.125

p(Hated|-)=(1+1)/ (6+10)= 0.125

p(poor|-)=(1+1)/ (6+10)= 0.125

p(Acting|-)=(1+1)/ (6+10)= 0.125

p(+|Doc)=(0.0833)(0.0417)(0.0417)(0.0833)(0.6)=0.0000072

p(-|Doc)=(0.125) (0.125) (0.125) (0.125) (0.4) = 0.00009765

**p(-|Doc)** > p(+|Doc)

From the TFDF values of the documents, construct Naïve Bayes model to classify the document “comedy, adventure, thriller, fiction“ into Class A or Class B. [note: Use multinomial distribution]

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| TFIFD | Action | Comedy | Romantic | Adventure | thriller | Fiction | Class |
| D1 | 0 | 2 | 4 | 0 | 0 | 0 | A |
| D2 | 3 | 1 | 0 | 4 | 2 | 0 | B |
| D3 | 0 | 0 | 3 | 0 | 1 | 2 | A |
| D4 | 3 | 0 | 4 | 3 | 2 | 0 | B |
| D5 | 4 | 2 | 0 | 1 | 0 | 1 | B |
| D6 |  | 3 |  | 2 | 2 | 3 | **?** |

**Key:**

P(comedy|A)= (2+1)/(12+2)= 0.2143

P(Adventure|A)= (0+1)/(12+2)=0.0714

P(Thriller|A) )= (1+1)/(12+2)=0.1429

P(Fiction|A) )= (2+1)/(12+2)=0.2143

P(comedy|B)= (3+1)/(30+2)= 0.125

P(Adventure|B)= (8+1)/(30+2)=0.2812

P(Thriller|B) = (4+1)/(30+2)=0.1562

P(Fiction|B)= (1+1)/(30+2)=0.059

P(D6|A)=10!X (0.2143)3/3!X(0.0714)2/2!X(0.1429)2/2!X(0.2143)3/3!=0.000254

P(D6|B)= 10!X (0.125)3/3!X(0.2812)2/2!X(0.1562)2/2!X(0.059)3/3!=0.0000195