**Prediksi Cacat Software Menggunakan Naïve Bayes**

Berdasarkan tabel berikut klasifiksikan (bc = 10, loc = 35) menggunakan Naïve Bayes

|  |  |  |
| --- | --- | --- |
| Branch Count | Line of Code | Defective? |
| 5 | 15 | No |
| 3 | 5 | No |
| 9 | 20 | No |
| 15 | 40 | Yes |
| 16 | 35 | Yes |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **BC** | | | Mean | StDev |
| Defective | Yes | 15 | 16 |  | 15.5 | 0.71 |
| No | 5 | 3 | 9 | 5.67 | 3.06 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | **LoC** | | | Mean | StDev |
| Defective | Yes | 40 | 35 |  | 37.5 | 3.54 |
| No | 15 | 5 | 20 | 13.3 | 7.64 |

P(defective = No) = 3/5 = 0.6

P(defective = Yes) = 2/5 = 0.4

Langkah 1:

**P(bc = 10 | No)**

Hitung density Gauss:

P(bc = 10 | No) = Gauss(x = 10, mean = 5.67, stdev = 3.06)

,

Langkah 2:

**P(LoC = 35 | No)**

Hitung density Gauss:

P(loc = 35 | No) = Gauss(x = 35, mean = 13.3, stdev = 7.64)

,

Dengan demikian:

P(Bc = 10, Loc = 35, | Defective = No)

={P(P(bc = 10 | No) . P(loc = 35 | no) . P(defective = No))}

= 0.048 . 0.00093 . 0.6

= **0.000027**

**Langkah 3**

P(defective = Yes) = 2/5 = 0.4

**P(bc = 10 | Yes)**

P(bc = 10 | Yes) = Gauss(x = 10, mean = 15.5, stdev = 0.71)

,

Langkah 4

**P(LoC = 35 | Yes)**

P(loc = 35 | Yes) = Gauss(x = 35, mean = 37.5, stdev = 3.54)

,

Dengan demikian:

P(Bc = 10, Loc = 35, | Defective = Yes)

={P(P(bc = 10 | Yes) . P(loc = 35 | Yes) . P(defective = Yes))}

= 0.00 . 0.088 . 0.4

= **0.000**

Oleh karena itu:

P(Bc = 10, Loc = 35, | Defective = No) = 0.000027

P(Bc = 10, Loc = 35, | Defective = Yes) = 0.000

Karena 0.000027 > 0.000, maka

Pilih hMAP dan diklasifikasikan sebagai terkena Defective (**no**).

**Text Sentiment Analysis Menggunakan Naïve Bayes**

Ada data latih sebanyak 5 dokumen teks.

|  |  |  |
| --- | --- | --- |
| Dokumen | Teks | Kelas |
| 1 | kuliahnya sangat membosankan | - |
| 2 | suara kurang jelas | - |
| 3 | ruang kuliah sangat panas | - |
| 4 | sangat menguasai materi | + |
| 5 | banyak studi kasus | + |

Hasil preprocessing

Kuliah, bosan, suara, kurang, jelas, ruang, panas, sangat, kuasa, materi, banyak, studi, kasus

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Dok. | kuliah | bosan | suara | kurang | jelas | ruang | panas | sangat | kuasa | materi | banyak | studi | kasus | Kelas |
| 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | - |
| 2 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | - |
| 3 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | - |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | + |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | + |

Dokumen dengan kelas positif

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Dok. | kuliah | bosan | suara | kurang | jelas | ruang | panas | sangat | kuasa | materi | banyak | studi | kasus | Kelas |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | + |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | + |
| **SUm** | **0** | **0** | **0** | **0** | **0** | **0** | **0** | **1** | **1** | **1** | **1** | **1** | **1** |  |

Hitung:

Misalnya *n* adalah jumlah kata kasus (*case*)positif adalah sebanyak **6**. nk adalah berapa kali kata *k* muncul dalam kasus positif (+).

Maka, , sehingga

*Vocabulary* adalah jumlah kata yang ada dalam dokumen yang dalam hal ini *vocabularynya* adalah Kuliah, bosan, suara, kurang, jelas, ruang, panas, sangat, kuasa, materi, banyak, studi, kasus. Berarti ada sebanyak **13** kata *vocabulary*.

Dokumen dengan kelas negatif (-)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Dok. | kuliah | bosan | suara | kurang | jelas | ruang | panas | sangat | kuasa | materi | banyak | studi | kasus | Kelas |
| 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | - |
| 2 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | - |
| 3 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | - |
| **Sum** | **2** | **1** | **1** | **1** | **1** | **1** | **1** | **2** | **0** | **0** | **0** | **0** | **0** |  |

Misalnya ada data uji sebuah dokumen kalimat *S* = “*materi* *disampaikan kurang jelas”*. Setelah prapemrosesan menjadi “*materi sampai kurang jelas”*. Karena kata *“sampai”* tidak dikenal dalam data latih, maka kata tersebut dieliminasi. Sehingga sekarang katanya adalah “materi”, “kurang”, “jelas”.

di mana *V* merupakan “value” atau “class”.

Jika *vj* = +; *p(+)p(materi|+) p(kurang|+)p(jelas|+) =*

Jika *vj* = -; *p(-)p(materi|-) p(kurang|-)p(jelas|-) =*

Untuk kalimat uji *S = “materi disampaikan kurang jelas”*, setelah membuang kata “*sampai”*, maka:

atau

atau

Karena **0.000117 < 0.000657**, maka kalimat *“materi disampaikan kurang jelas”* diprediksi memiliki sentiment ***Negatif***.

Bahan Bacaan

https://www.youtube.com/watch?v=EGKeC2S44Rs

https://web.stanford.edu/~jurafsky/slp3/6.pdf