1. **Word Sense Disambiguation**

Kalimat:

• k1: Rumah dinas kepala cabang kantor pemasaran yang baru berhalaman luas.  
• k2: Kami harus menyediakan uang puluhan juta untuk operasi tumor di kepala kakek.  
• k3: Setelah menangkap belut dan menyiangi rumput halaman belakang, baru kami menyantap gulai  
kepala kakap.  
• k4: Kepalaku sudah panas dipenuhi puluhan halaman materi ujian NLP.  
• k5: Kabarnya tiga bulan lagi buku terbaru setebal delapan puluh halaman karya penulis ternama itu  
akan diluncurkan.  
• k6: Tiga puluh kepala keluarga menerima santunan uang di halaman Dinas Sosial tadi siang.  
• k7: Halaman utama surat kabar nasional dipenuhi berita kepala rumah sakit ternama terkena operasi  
tangkap tangan uang panas.

1. Misalkan s1 adalah makna kata kepala sebagai organ tubuh dan s2 adalah makna kepala sebagai unti tertinggi dalam suatu hirarki. Makna kepala pada setiap kalimat adalah sebagai berikut.

|  |  |
| --- | --- |
| Kalimat | Makna |
| k1 | s2 |
| k2 | s1 |
| k3 | s1 |
| k4 | s1 |
| k5 | - |
| k6 | s2 |
| k7 | s2 |

1. Co-occurance vector berdasarkan kalimat training

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Kalimat | tiga | puluh | uang | halaman | dinas | siang | kabar | rumah | nama | kena | panas | operasi | Makna |
| k1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | s2 |
| k2 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | s1 |
| k3 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | s1 |
| k4 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | s1 |
| k5 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | - |

1. Pada penghitungan naïve bayes, **kalimat 5 tidak diperhitungkan** sebagai training karena tidak mengandung kata kepala, sehingga tidak relevan pada kasus ini.

* **Kalimat 6:**

**“Tiga** **puluh** kepala keluarga menerima santunan **uang** di **halaman** **Dinas** Sosial tadi **siang**.”

Mengandung fitur: tiga, puluh, uang, halaman, dinas, siang.

Dengan menggunakan add-one smoothing diperoleh:

Dari prior probability dan MLE probability tersebut diperoleh,

Berdasarkan penghitungang tersebut, dapat disimpulkan bahwa posterior probability untuk s1 lebih besar dari s2, sehingga kata kepala pada kalimat 6 diprediksi memiliki makna s1 (organ tubuh).

* **Kalimat 7:**

**“Halaman** utama surat **kabar** nasional dipenuhi berita kepala **rumah** sakit **ternama** **terkena** **operasi**  
tangkap tangan **uang** **panas**.”

Mengandung fitur: halaman, kabar, rumah, nama, kena, operasi, uang, panas.

Dengan menggunakan add-one smoothing diperoleh:

Dari prior probability dan MLE probability tersebut diperoleh,

Berdasarkan penghitungang tersebut, dapat disimpulkan bahwa posterior probability untuk s2 lebih besar dari s1, sehingga kata kepala pada kalimat 7 diprediksi memiliki makna s2 (unit tertinggi hirarki).

1. Menerapkan dua kasus untuk fitur kata dasar
2. **Menggunakan 7 kata dasar terbaik**

Untuk menentukan 7 kata dasar yang dipilih, akan digunakan information gain

Berdasarkan information gain tersebut, diambil 7 fitur berikut:

**panas, dinas, siang, rumah, halaman, puluh, uang**

Co-occurance vector:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Kalimat | panas | puluh | uang | rumah | dinas | siang | halaman | Makna |
| k1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | s2 |
| k2 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | s1 |
| k3 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | s1 |
| k4 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | s1 |
| k5 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | - |

Kalimat 5 tidak digunakan dalam penghitungan naïve bayes

**Prediksi untuk kalimat 6 dan 7**

* **Kalimat 6:**

**“**Tiga **puluh** kepala keluarga menerima santunan **uang** di **halaman Dinas** Sosial tadi **siang**.”

Mengandung fitur: puluh, uang, halaman, dinas, siang.

Dengan menggunakan add-one smoothing diperoleh:

Dari prior probability dan MLE probability tersebut diperoleh,

Berdasarkan penghitungang tersebut, dapat disimpulkan bahwa posterior probability untuk s1 lebih besar dari s2, sehingga kata kepala pada kalimat 6 diprediksi memiliki makna s1 (organ tubuh).

* **Kalimat 7:**

**“Halaman** utama surat kabar nasional dipenuhi berita kepala **rumah** sakit ternama terkena operasi  
tangkap tangan **uang panas**.”

Mengandung fitur: halaman, rumah, uang, panas.

Dengan menggunakan add-one smoothing diperoleh:

Dari prior probability dan MLE probability tersebut diperoleh,

Berdasarkan penghitungang tersebut, dapat disimpulkan bahwa posterior probability untuk s1 lebih besar dari s2, sehingga kata kepala pada kalimat 7 diprediksi memiliki makna s1 (organ tubuh).

1. **Menggunakan seluruh kata dasar pada training data**

Asumsi: kata dasar pada kalimat 5 tetap diperhitungkan sebagai fitur, namun saat penghitungan naïve bayes tidak digunakan seperti soal sebelumnya

**Kata dasar:** tiga, puluh, uang, halaman, dinas, siang, kabar, rumah, nama, panas, operasi, kepala, cabang, kantor, pasar, yang, baru, luas, kami, harus, sedia, juta, untuk, tumor, di, kakek, setelah, tangkap, belut, dan, rumput, belakang, santap, gulai, kakap, sudah, penuh, materi, ujian, NLP, bulan, lagi, buku, tebal, delapan, karya, tulis, itu, akan, luncur

**Co-occurance vectors:**

**TERLAMPIR**

**Prediksi untuk kalimat 6 dan 7**

* **Kalimat 6:**

**“Tiga** **puluh** **kepala** keluarga menerima santunan **uang** **di** **halaman** **Dinas** Sosial tadi **siang**.”

Mengandung fitur: tiga, puluh, kepala, uang, di, halaman, dinas, siang.

Dengan menggunakan add-one smoothing diperoleh:

Dari prior probability dan MLE probability tersebut diperoleh,

Berdasarkan penghitungang tersebut, dapat disimpulkan bahwa posterior probability untuk s1 lebih besar dari s2, sehingga kata kepala pada kalimat 6 diprediksi memiliki makna s1 (organ tubuh).

* **Kalimat 7:**

**“Halaman** utama surat **kabar** nasional **dipenuhi** berita kepala **rumah** sakit **ternama** terkena **operasi**  
**tangkap** tangan **uang** **panas**.”

Mengandung fitur: halaman, kabar, penuh, kepala, rumah, nama, operasi, tangkap, uang, panas.

Dengan menggunakan add-one smoothing diperoleh:

Dari prior probability dan MLE probability tersebut diperoleh,

Berdasarkan penghitungang tersebut, dapat disimpulkan bahwa posterior probability untuk s2 lebih besar dari s1, sehingga kata kepala pada kalimat 7 diprediksi memiliki makna s2 (unit tertinggi hirarki).