



Nama : Muhammad Azhar Rasyad
NIM : 0110217029
Mata Kuliah : Natural Language Processing

Ujian Tengah Semester

1. Jelaskan salah satu terapan NLP!

Jawaban :

Salah satu terapan NLP yang membuat saya tertarik yaitu Sentimen Analisis, berdasarkan yang saya ketahui bahwa sentimen analisis merupakan proses dalam menganalisis review yang diberikan oleh user terhadap sesuatu. Hasil analisis review user tersebut dapat digunakan sebagai informasi dalam memperbaiki atau mengembangkan sesuatu.

Contoh pada suatu aplikasi e-commerce terdapat review yang diberikan user berupa rating 1-5 disertai komentarnya, dari review tersebut banyak yang memberikan rating 4-5 dengan memberikan saran supaya aplikasi e-commerce tersebut dapat lebih baik lagi semisal agar ditambahkan lagi untuk promo-promonya, akan tetapi disisi lain yaitu rating 1-3 memberikan kritik dari sisi aplikasinya semisal tampilannya kurang bagus, aplikasinya lambat, dan sebagainya. Kedua hal tersebut bukanlah sebagai pajangan semata, akan tetapi dapat digunakan untuk membuat aplikasi e-commerce tersebut berkembang lebih baik lagi.

Dari contoh di atas bisa saja aplikasi tersebut memiliki review yang sangat banyak dan sulit untuk dicek satu persatu sehingga penerapan NLP ini dapat digunakan sebagai alternatif yang baik. Mulai dari menganalisis review user hingga memberikan sebuah informasi berdasarkan sudut pandang user tersebut dari pengalamannya dan akhirnya dapat memberikan manfaat bagi yang direview berupa saran bahkan kritik.

2. Jelaskan sesuai pemahaman anda apa yang dimaksud dengan sintaksis, semantik, dan pragmatik ! Berikan contohnya (contoh selain yang diberikan di slide)

Jawaban :

Menurut saya, sintaksis yaitu urutan suatu kata dalam kalimat dapat memiliki arti yang berbeda, contoh “aku makan nasi” akan tetapi jika dirubah urutannya menjadi “nasi makan aku” maka kedua kalimat tersebut memiliki arti yang berbeda.

Semantik menurut saya yaitu pengartian konteks secara menyeluruh berdasarkan sudut pandang yang berbeda. Contohnya kata “Tahu” jika dilihat dari sudut pandang makanan maka kata “Tahu” itu adalah suatu makanan yang terbuat dari kedelai, lain halnya jika kata “Tahu” jika dilihat dari sudut pandang berbeda semisal “Saya tahu jawaban nomor 1”, kata tahu tersebut bukanlah makanan yang sebelumnya disebutkan akan tetapi mengacu pada informasi yang dimiliki.

Berdasarkan yang saya ketahui Pragmatik yaitu mengartikan konteks berdasarkan kondisi sosial yang ada dimasyarakat. Contoh “Jalanan di jakarta pada senin pagi sudah penuh”, hal tersebut jika dilihat dari kondisi sosialnya bahwa rata-rata jalanan di jakarta pada hari senin dipagi hari itu sering terjadi kemacetan karena banyaknya masyarakat yang bekerja di jakarta.

3. Tuliskan sebuah kalimat yang ambigu, dan beritahukan ambiguitas kalimat tersebut tergolong ambigu dari sisi sintaksis, semantik, atau pragmatik!

Jawaban :

A : Kamu udah makan?

B : Udah tapi belum kenyang.

A : Makan apa emangnya?

B : MAKAN HATI!

Contoh di atas tergolong dari sisi semantik, karena makan hati yang dimaksud bukanlah makanan hati yang biasa dimasak, akan tapi makna makan hati disana termasuk suasana hatinya yang sedang buruk akibat perbuatan orang lain.

(untuk soal nomor 4, 5 dan 6 menggunakan korpus sederhana berikut)

Diberikan 5 buah kalimat pendek :

- (a) Buah ada beragam warna
- (b) Apel adalah buah berwarna hijau
- (c) Apel ada yang berwarna merah juga
- (d) Cabai ada yang berwarna hijau dan merah
- (e) Buah yang berwarna lain juga ada

4. Tokenisasikan kalimat tersebut, dan buat daftar kata unigram dan bigramnya (menggunakan padding awal dan akhir kalimat).

Jawaban :

Tokenisasi Unigram

Kalimat 1 : buah ada beragam warna

Unigram Sementara :

- buah = 1
- ada = 1
- beragam = 1
- warna = 1

Kalimat 2 : apel adalah buah berwarna hijau

Unigram Sementara :

- buah = 1 + 1
- ada = 1
- beragam = 1
- warna = 1
- apel = 1
- adalah = 1
- berwarna = 1
- hijau = 1

Kalimat 3 : apel ada yang berwarna merah juga

Unigram Sementara :

- buah = 1 + 1
- ada = 1 + 1
- beragam = 1
- warna = 1
- apel = 1 + 1
- adalah = 1
- berwarna = 1 + 1
- hijau = 1
- yang = 1
- merah = 1
- juga = 1

Kalimat 4 : cabai ada yang berwarna hijau dan merah

Unigram Sementara :

- buah = 1 + 1
- ada = 1 + 1 + 1
- beragam = 1
- warna = 1
- apel = 1 + 1
- adalah = 1
- berwarna = 1 + 1 + 1
- hijau = 1 + 1
- yang = 1 + 1
- merah = 1 + 1
- juga = 1
- cabai = 1
- dan = 1

Kalimat 5 : buah yang berwarna lain juga ada

Unigram Sementara :

- buah = 1 + 1 + 1
- ada = 1 + 1 + 1 + 1
- beragam = 1
- warna = 1

- apel = 1 + 1
- adalah = 1
- berwarna = 1 + 1 + 1 + 1
- hijau = 1 + 1
- yang = 1 + 1 + 1
- merah = 1 + 1
- juga = 1 + 1
- cabai = 1
- dan = 1
- lain = 1

Total Tokenisasi Unigram

- buah = 3
- ada = 4
- beragam = 1
- warna = 1
- apel = 2
- adalah = 1
- berwarna = 4
- hijau = 2
- yang = 3
- merah = 2
- juga = 2
- cabai = 1
- dan = 1
- lain = 1

Tokenisasi Bigram

Kalimat 1 : <s> buah ada beragam warna </s>

Bigram Sementara :

- <s>, buah = 1
- buah, ada = 1
- ada, beragam = 1

- beragam, warna = 1
- warna, </s> = 1

Kalimat 2 : <s> apel adalah buah berwarna hijau </s>

Bigram Sementara :

- <s>, buah = 1
- buah, ada = 1
- ada, beragam = 1
- beragam, warna = 1
- warna, </s> = 1
- <s>, apel = 1
- apel, adalah = 1
- adalah, buah = 1
- buah, berwarna = 1
- berwarna, hijau = 1
- hijau, </s> = 1

Kalimat 3 : <s> apel ada yang berwarna merah juga </s>

Bigram Sementara :

- <s>, buah = 1
- buah, ada = 1
- ada, beragam = 1
- beragam, warna = 1
- warna, </s> = 1
- <s>, apel = 1 + 1
- apel, adalah = 1
- adalah, buah = 1
- buah, berwarna = 1
- berwarna, hijau = 1
- hijau, </s> = 1
- apel, ada = 1
- ada, yang = 1
- yang, berwarna = 1
- berwarna, merah = 1

- merah, juga = 1
- juga, </s> = 1

Kalimat 4 : <s> cabai ada yang berwarna hijau dan merah </s>

Bigram Sementara :

- <s>, buah = 1
- buah, ada = 1
- ada, beragam = 1
- beragam, warna = 1
- warna, </s> = 1
- <s>, apel = 1 + 1
- apel, adalah = 1
- adalah, buah = 1
- buah, berwarna = 1
- berwarna, hijau = 1 + 1
- hijau, </s> = 1
- apel, ada = 1
- ada, yang = 1 + 1
- yang, berwarna = 1 + 1
- berwarna, merah = 1
- merah, juga = 1
- juga, </s> = 1
- <s>, cabai = 1
- cabai, ada = 1
- hijau, dan = 1
- dan, merah = 1
- merah, </s> = 1

Kalimat 5 : <s> buah yang berwarna lain juga ada </s>

Bigram Sementara :

- <s>, buah = 1 + 1
- buah, ada = 1
- ada, beragam = 1
- beragam, warna = 1

- warna, </s> = 1
- <s>, apel = 1 + 1
- apel, adalah = 1
- adalah, buah = 1
- buah, berwarna = 1
- berwarna, hijau = 1 + 1
- hijau, </s> = 1
- apel, ada = 1
- ada, yang = 1 + 1
- yang, berwarna = 1 + 1 + 1
- berwarna, merah = 1
- merah, juga = 1
- juga, </s> = 1
- <s>, cabai = 1
- cabai, ada = 1
- hijau, dan = 1
- dan, merah = 1
- merah, </s> = 1
- buah, yang = 1
- berwarna, lain = 1
- lain, juga = 1
- juga, ada = 1
- ada, </s> = 1

Total Tokenisasi Bigram

- <s>, buah = 2
- buah, ada = 1
- ada, beragam = 1
- beragam, warna = 1
- warna, </s> = 1
- <s>, apel = 2
- apel, adalah = 1
- adalah, buah = 1
- buah, berwarna = 1

- berwarna, hijau = 2
- hijau, </s> = 1
- apel, ada = 1
- ada, yang = 2
- yang, berwarna = 3
- berwarna, merah = 1
- merah, juga = 1
- juga, </s> = 1
- <s>, cabai = 1
- cabai, ada = 1
- hijau, dan = 1
- dan, merah = 1
- merah, </s> = 1
- buah, yang = 1
- berwarna, lain = 1
- lain, juga = 1
- juga, ada = 1
- ada, </s> = 1

5. Berapa peluang kalimat "Buah berwarna hijau dan merah" muncul dari korpus tersebut menggunakan model bahasa bigram?

Jawaban :

Kalimat : $\langle s \rangle$ buah berwarna hijau dan merah $\langle /s \rangle$

- $P(\text{buah} \mid \langle s \rangle) = 2 / 5 = 0.4$
- $P(\text{berwarna} \mid \text{buah}) = 1 / 3 = 0.33$
- $P(\text{hijau} \mid \text{berwarna}) = 2 / 4 = 0.5$
- $P(\text{dan} \mid \text{hijau}) = 1 / 2 = 0.5$
- $P(\text{merah} \mid \text{dan}) = 1 / 1 = 1$
- $P(\langle /s \rangle \mid \text{merah}) = 1 / 2 = 0.5$

$P(\langle s \rangle \text{ buah berwarna hijau dan merah} \langle /s \rangle) = 0.4 \times 0.33 \times 0.5 \times 0.5 \times 1 \times 0.5 = \mathbf{0.0165}$

6. Apa yang harus dilakukan jika ternyata ada token bigram yang tidak pernah muncul di korpus training? Misalnya kalimat "Buah apel berwarna hijau dan merah sekaligus" . Berapa peluang kalimat tersebut sebelum dan setelah diberikan penanganan?

Jawaban :

Ada beberapa yang dapat dilakukan untuk mengatasinya salah satunya dengan cara membuat model training yang termasuk simbol kata untuk *Unknown Words* (<UNK>) sehingga jika token bigram yang tidak ada dalam korpus training muncul maka token bigram tersebut menjadi <UNK> saat kemunculan pertama. Cara lain yang dapat dilakukan dengan smoothing, yaitu menganggap kata yang tidak pernah muncul dianggap muncul sekali. Pada kasus kali ini akan dicoba menggunakan metode smoothing yaitu laplace smoothing.

Korpus Training

<s> Buah ada beragam warna </s>

<s> Apel adalah buah berwarna hijau </s>

<s> Apel ada yang berwarna merah juga </s>

<s> Cabai ada yang berwarna hijau dan merah </s>

<s> Buah yang berwarna lain juga ada </s>

Sebelum penanganan Kalimat : <s> buah apel berwarna hijau dan merah sekaligus </s>

- $P(\text{buah} \mid \text{<s>}) = 2 / 5 = 0.4$
- **$P(\text{apel} \mid \text{buah}) = 0 / 3 = 0$**
- **$P(\text{berwarna} \mid \text{apel}) = 0 / 2 = 0$**
- $P(\text{hijau} \mid \text{berwarna}) = 2 / 4 = 0.5$
- $P(\text{dan} \mid \text{hijau}) = 1 / 2 = 0.5$
- $P(\text{merah} \mid \text{dan}) = 1 / 1 = 1$
- **$P(\text{sekaliigus} \mid \text{merah}) = 0 / 2 = 0$**
- **$P(\text{</s>} \mid \text{sekaliigus}) = 0 / 0 = 0$**

$P(\text{<s> buah apel berwarna hijau dan merah sekaligus </s>})$

$$= 0.4 \times 0 \times 0 \times 0.5 \times 0.5 \times 1 \times 0 \times 0$$

$$= 0$$

Setelah penanganan menggunakan metode laplace smoothing, dengan $V = 16$

Kalimat : $\langle s \rangle$ buah apel berwarna hijau dan merah sekaligus $\langle /s \rangle$

- $P(\text{buah} \mid \langle s \rangle) = 2 / 5 = (2 + 1) / (5 + 16) = 3 / 21 = \mathbf{0.14}$
- $P(\text{apel} \mid \text{buah}) = 0 / 3 = (0 + 1) / (3 + 16) = 1 / 19 = \mathbf{0.053}$
- $P(\text{berwarna} \mid \text{apel}) = 0 / 2 = (0 + 1) / (2 + 16) = 1 / 18 = \mathbf{0.056}$
- $P(\text{hijau} \mid \text{berwarna}) = 2 / 4 = (2 + 1) / (4 + 16) = 3 / 20 = \mathbf{0.15}$
- $P(\text{dan} \mid \text{hijau}) = 1 / 2 = (1 + 1) / (2 + 16) = 2 / 18 = \mathbf{0.11}$
- $P(\text{merah} \mid \text{dan}) = 1 / 1 = (1 + 1) / (1 + 16) = 2 / 17 = \mathbf{0.12}$
- $P(\text{sekali} \mid \text{merah}) = 0 / 2 = (0 + 1) / (2 + 16) = 1 / 18 = \mathbf{0.056}$
- $P(\langle /s \rangle \mid \text{sekali}) = 0 / 0 = (0 + 1) / (0 + 16) = 1 / 16 = \mathbf{0.063}$

$P(\langle s \rangle \text{ buah apel berwarna hijau dan merah sekaligus } \langle /s \rangle) = \mathbf{2.90259003e-9}$