

Sistem

Dokumen HTML

*Preprocessing*

KNN *Clustering*

*Naïve Bayes*



Hasil Klasifikasi

|  |  |
| --- | --- |
| **Problem** | |
| Banyaknya data-data yang tersebar dalam dunia maya, khususnya artikel *online*. Artikel tersebut belum terklasifikasi sesuai dengan topik yang dibahasnya. | Perlunya aplikasi yang dapat mengelompokkan artikel dari beberapa situs *website* secara otomatis |

|  |  |
| --- | --- |
| **Opportunity** | |
| Keterbatasan SDM dan waktu dalam menginputkan artikel yang berasal dari situs resmi | Belum ada aplikasi yang mengelompokan berita dari beberapa situs secara otomatis |

|  |  |
| --- | --- |
| **Approach** | |
| Aplikasi *online* berbasis *web* | Metode *Naive Bayes* dan *K-Nearest Neighbour* |

|  |  |
| --- | --- |
| **Software Development** | |
| Metode Pengembangan: *Sequential Linear* | Metode Perancangan: UML *(Unified Modeling Language)* |

|  |
| --- |
| **Software Implementation** |
| Aplikasi yang dapat mengelompokkan artikel berita yang berasal dari beberapa situs *website* secara otomatis |

|  |
| --- |
| **Result** |
| Implementasi Metode  *Naive Bayes* dan *K-Nearest Neighbour* untuk Kategorisasi Artikel Berita di Internet |

Definisi Aktor

|  |  |  |
| --- | --- | --- |
| **No.** | **Aktor** | **Deskripsi** |
| 1. | Website | Aktor ini merupakan halaman *website* berita *online* yang akan diambil isinya *(scraping).* *Website* yang akan di-*scraping* diantaranya detik.com, kompas.com, liputan6.com, viva.co.id, dan tribunnews.com. |
| 2. | Website Visitor | Aktor ini merupakan orang yang akan menggunakan aplikasi ini. Aktor ini bisa melihat berita-berita yang sudah dikategorisan sesuai dengan kategori masing-masing topik dan bisa melihat proses dari kategorisasi dokumen berita. |

Use Case

1. *Usecase Scraping*

|  |  |
| --- | --- |
| *Usecase id* | UC-1 |
| *Usecase name* | Melakukan *Scraping* |
| *Actors* | *Website* |
| *Description* | *Usecase* ini menggambarkan proses pengambilan data dari halaman website seperti *link, image,* dan isi artikel. |
| *Precondition* | Aplikasi telah berjalan pada *web server* dan tersambung dengan jaringan internet. |
| *Normal flow* | 1. Sistem melakukan *load* dengan *link* beritayang sudah ditentukan. 2. Sistem akan melakukan proses *scraping* *website.* 3. Sistem menyimpan data hasil *scraping* ke dalam *database* sistem. |
| *Alternatif flow* | 1. Sistem melakukan *load* dengan *link* berita. 2. Sistem mengecek *link* berita. 3. Validasi gagal dan sistem tidak dapat melakukan *scraping website.* |

|  |  |  |
| --- | --- | --- |
| **Nomor** | 001 | |
| **Nama** | *Scraping* | |
| **Tujuan** | Proses ini digunakan untuk mengambil data dari halaman *website.* | |
| **Deskripsi** | *Usecase* ini menggambarkan proses pengambilan data dari halaman website seperti *link, image,* dan isi artikel. | |
| **Aktor** | *Website* | |
| **Prekondisi** | Aplikasi telah berjalan pada *web server* dan tersambung dengan jaringan internet | |
| **Skenario Utama** | | |
| **Aksi Aktor** | | **Reaksi Sistem** |
| 1. Membuka halaman *web* melalui URL yang sudah ada. | |  |
|  | | 1. Sistem mengambil elemen *website* kemudian menyimpannya ke dalam *database* sistem. |

1. *Usecase Preprocessing*

|  |  |
| --- | --- |
| *Usecase id* | UC-2 |
| *Usecase name* | Melakukan *Preprocessing* |
| *Actors* | *-* |
| *Description* | *Usecase* ini menggambarkan proses *preprocessing* isi website berupa teks yang terdiri dari *case folding, tokenizing, filtering,* dan *stemming.* |
| *Precondition* | Isi *web* hasil *scraping* yang berupa teks sudah tersimpan dalam *database* sistem. |
| *Normal flow* | 1. Sistem memanggil data yang tersimpan dalam *database* sistem. 2. Sistem melakukan proses *preprocessing.* |
| *Alternatif flow* | 1. Sistem memeriksa jumlah data dalam *database.* 2. Data dalam *database* tidak ada dan sistem tidak dapat melakukan proses *preprocessing.* |

|  |  |  |
| --- | --- | --- |
| **Nomor** | 002 | |
| **Nama** | *Preprocessing* | |
| **Tujuan** | Proses ini digunakan untuk *preprocessing* artikel *website.* | |
| **Deskripsi** | *Usecase* ini menggambarkan proses *preprocessing* isi website berupa teks yang terdiri dari *case folding, tokenizing, filtering,* dan *stemming.* | |
| **Aktor** | - | |
| **Prekondisi** | Isi *web* yang berupa teks sudah tersimpan dalam *database* sistem. | |
| **Skenario Utama** | | |
| **Aksi Aktor** | | **Reaksi Sistem** |
|  | | 1. Sistem mengambil data dari *website.* |
|  | | 1. Sistem melakukan proses *preprocessing.* |

1. *Usecase* Kategorisasi

|  |  |
| --- | --- |
| *Usecase id* | UC-3 |
| *Usecase name* | Melakukan Proses Kategori |
| *Actors* | *-* |
| *Description* | *Usecase* ini menggambarkan proses perhitungan teks untuk mengetahui kategori yang sesuai dengan teks tersebut. |
| *Precondition* | Data teks sudah dilakukan proses *preprocessing.* |
| *Normal flow* | 1. Sistem melakukan perhitungan untuk mengetahui bobot setiap kata. 2. Sistem melakukan perhitungan untuk mengetahui kategori dokumen menggunakan algoritma *k-nearest neighbor.* 3. Sistem melakukan perhitungan untuk mengetahui kategori dokumen menggunakan algoritma *naïve bayes classifier.* |
| *Alternatif flow* | - |

|  |  |  |
| --- | --- | --- |
| **Nomor** | 003 | |
| **Nama** | Kategorisasi | |
| **Tujuan** | Proses ini digunakan untuk menghitung nilai teks dimasukan ke dalam kategori yang sesuai. | |
| **Deskripsi** | *Usecase* ini menggambarkan proses perhitungan teks untuk mengetahui kategori yang sesuai dengan teks tersebut. | |
| **Aktor** | Aplikasi | |
| **Prekondisi** | Data teks sudah dilakukan proses *preprocessing* | |
| **Skenario Utama** | | |
| **Aksi Aktor** | | **Reaksi Sistem** |
|  | | 1. Sistem melakukan perhitungan untuk mengetahui bobot setiap kata. |
|  | | 1. Sistem melakukan perhitungan untuk mengetahui kategori dokumen menggunakan algoritma *k-nearest neighbor.* |
|  | | 1. Sistem melakukan perhitungan untuk mengetahui kategori dokumen menggunakan algoritma *naïve bayes classifier.* |

1. *Usecase Display Website*

|  |  |
| --- | --- |
| *Usecase id* | UC-4 |
| *Usecase name* | Menampilkan *Website* |
| *Actors* | *Website Visitor* |
| *Description* | *Usecase* ini menggambarkan proses menampilkan halaman *user interface website* yang berisi artikel-artikel berita yang sudah dikategorikan dan halaman proses kategorisasi. |
| *Precondition* | Aplikasi sudah berjalan pada *web server* dan semua proses *usecase* berjalan sesuai fungsinya. |
| *Normal flow* | 1. Actor membuka halaman *website.* 2. Sistem menampilkan halaman *website* dengan artikel berita yang sudah dikategorikan. 3. Aktor membuka halaman *classification* untuk melihat proses pengkategorian dokumen. 4. Aktor menginputkan *link*  *website* untuk dilakukan perhitungan. 5. Sistem mengecek *link website* yang diinputkan. 6. Sistem melakukan proses *scraping, preprocessing,* menghitung bobot tiap kata, melakukan perhitungan untuk mendapatkan nilai kategori dengan algoritma knn dan *naïve bayes klassifier.* 7. Sistem menampilkan hasil dari proses-proses tersebut. |
| *Alternatif flow* | 1. Actor membuka halaman *website.* 2. Sistem menampilkan halaman *website* dengan artikel berita yang sudah dikategorikan. 3. Aktor membuka halaman *classification* untuk melihat proses pengkategorian dokumen. 4. Aktor menginputkan *link*  *website* untuk dilakukan perhitungan. 5. Sistem mengecek *link website* yang diinputkan. 6. Sistem gagal memvalidasi *link* yang diinputkan dan gagal melakukan proses pengkategorian. |

|  |  |  |
| --- | --- | --- |
| **Nomor** | 004 | |
| **Nama** | *Display Website* | |
| **Tujuan** | Proses ini digunakan untuk menampilkan halaman *user interface* aplikasi. | |
| **Deskripsi** | *Usecase* ini menggambarkan proses menampilkan halaman *user interface website* yang berisi artikel-artikel berita yang sudah dikategorikan dan halaman proses kategorisasi. | |
| **Aktor** | *Website visitor* | |
| **Prekondisi** | Aplikasi sudah berjalan pada *web server* dan semua proses *usecase* berjalan sesuai fungsinya. | |
| **Skenario Utama** | | |
| **Aksi Aktor** | | **Reaksi Sistem** |
| 1. Actor membuka halaman *website.* | |  |
|  | | 1. Sistem menampilkan halaman *website* dengan artikel berita yang sudah dikategorikan. |
| 1. Aktor membuka halaman *classification* untuk melihat proses pengkategorian dokumen. | |  |
| 1. Aktor menginputkan *link*  *website* untuk dilakukan perhitungan. | |  |
|  | | 1. Sistem melakukan proses *scraping, preprocessing,* menghitung bobot tiap kata, melakukan perhitungan untuk mendapatkan nilai kategori dengan algoritma knn dan *naïve bayes klassifier.* |
|  | | 1. Sistem menampilkan hasil dari proses-proses tersebut. |

1. Pemecahan *string* menjadi kalimat
2. *Case folding* (*lowe case* dan penghilangan simbol)
3. *Tokenizing* (pemotongan kata)
4. *Filtering* (menghilangkan *stopword*)
5. *Stemming* (mengembalikan ke kata dasar)

*Scraping website*

*Text preprocessing*

Pembobotan TF-IDF

Pengkategorian dengan *k-nearest neighbour*

Pengkategorian dengan *naïve bayes*

|  |  |
| --- | --- |
| **No** | **Kalimat** |
| 1. | ketua komisi iii dpr bambang soesatyo membenarkan adanya operasi tangkap tangan oleh kpk terhadap salah satu anggota komisi iii |
| 2. | anggota tersebut diduga dari fraksi partai demokrat namun bambang memastikan tidak terkait kegiatan di komisi |
| 3. | saya sudah cek tidak terkait dengan tupoksi komisi iii lebih kepada kegiatan banggar kata bambang saat dikonfirmasi rabu juni |
| 4. | bambang mengaku sedih dengan penangkapan rekan sejawat di komisinya tersebut bambang yang juga ketua dpp golkar menyebut putu sebagai sosok yang humoris |
| 5. | kami semua di komisi iii sedih dan prihatin putu i putu sudiartana adalah sahabat yang baik dan humoris nggak ada putu nggak rame tutur bambang |
| 6. | saat ini kpk juga sudah menyegel ruang anggota tersebut di dpr |
| 7. | tim satuan tugas kpk dikabarkan telah menangkap seorang anggota dpr karena diduga terkait dengan tindak pidana korupsi anggota dewan itu disebut sebut berasal dari komisi iii dpr fraksi partai demokrat berinisial ips |
| 8. | saat dikonfirmasi wakil ketua kpk laode muhammad syarif tidak menampiknya namun menurut syarif penjelasan lebih detail akan diberikan pada konferensi pers nanti |
| 9. | tunggu sore saja detailnya kata syarif |