

KLASIFIKASI DOKUMEN BERITA OLAHRAGA MENGUNAKAN *NAIVE BAYES CLASSIFIER*

Diajukan Sebagai Syarat untuk Menyelesaikan Pendidikan
Program Studi Strata-1 pada Jurusan Teknik Informatika
Fakultas Ilmu Komputer Universitas Sriwijaya



Oleh :

OJHI GUSTI MAROJAHAN SILABAN
09021181320042

Jurusan Teknik Informatika
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2021

LEMBAR PENGESAHAN SKRIPSI

**KLASIFIKASI DOKUMEN BERITA OLAH RAGA
MENGUNAKAN *NAÏVE BAYES CLASSIFIER***

OLEH:

OJHI GUSTI MAROJAHAN SILABAN

09021181320042

Pembimbing I,



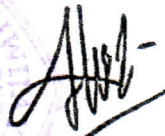
Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

Indralaya, 09 Januari 2021
Pembimbing II,



Dr. Abdiansah, S.Kom., M.CS.
NIP. 198410012009121005

Mengetahui,
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari rabu, tanggal 30 Desember 2020 telah dilaksanakan Ujian Sidang Tugas Akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : Ojhi Gusti Marojahan Silaban

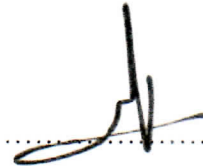
NIM : 09021181320042

Judul : Klasifikasi Dokumen Berita Olahraga Menggunakan *Naïve Bayes Classifier*

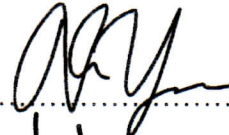
1. Pembimbing I
Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003



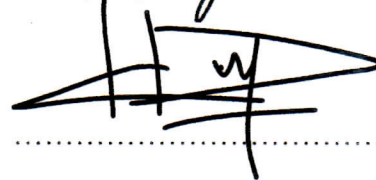
2. Pembimbing II
Dr. Abdiansah, S.Kom., M.CS.
NIP. 198410012009121005



3. Penguji I
Novi Yusliani, M.T.
NIP. 198211082012122001



4. Penguji II
Ali Ibrahim, M.T.
NIP. 198407212019031004



Mengetahui,
Ketua Jurusan Teknik Informatika



Alvi Syahrini Utami, M.Kom.
NIP. 197812222006042003

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Ojhi Gusti Marojahan Silaban

NIM : 09021181320042

Program Studi : Teknik Informatika

Judul Skripsi : Klasifikasi Dokumen Berita Olahraga Menggunakan *Naïve*

Bayes Classifier

Hasil Pengecekan *Software iThenticate/ Turnitine* : 5%

Menyatakan bahwa laporan proyek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/ plagiat. Apabila ditemukan unsur penjiplakan/ plagiat dalam laporan proyek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.

Indralaya, Januari 2021



Ojhi Gusti Marojahan Silaban
NIM. 09021181320042

MOTTO DAN PERSEMBAHAN

- Niat baik akan selalu menghasilkan sesuatu yang baik juga
- Doa dan restu orang tua adalah kunci dalam memperlancar segala urusan
- Intelligence plus character, that is the goal of true education
- Kamu tidak bisa kembali dan mengubah masa lalu, maka dari itu tataplah masa depan dan jangan membuat kesalahan yang sama dua kali

Kupersembahkan karya tulis ini kepada:

- Tuhan Yesus Kristus
- Ibu dan Ayah tercinta
- Seluruh Keluarga Besarku
- Kedua Pembimbing
- Seluruh sahabat dan teman-temanku

Sports News Document Classification Using *Naïve Bayes Classification*

Ojhi Gusti Marojahan Silaban (09021181320042)

Jurusan Teknik Informatika, Fakultas Ilmu Komputer

Universitas Sriwijaya

ABSTRACT

Document classification aims to group unstructured documents into groups that describe the contents of the document. Documents can be text documents such as news articles. One type of news that is in great demand is sports news, especially by men. The number of sports makes readers confused looking for sports news that they want to read. Therefore, document classification is needed in order to classify sports news according to the sports news group. To solve this problem, software was developed using the Naïve Bayes Classifier with simple computation and a fairly high degree of accuracy. The Naïve Bayes Classifier predicts future probabilities based on past experiences. There are three stages, namely, preprocessing, training, and classification. This study uses secondary data taken from online news portals with various categories, namely football, basketball, badminton, MotoGP, formula1, and MMA. The resulting accuracy rate from the software is 73.33% by using 30 training documents and 30 test documents.

Keywords: document classification, sports news, Naïve Bayes Classifier

Klasifikasi Dokumen Berita Olahraga Menggunakan *Naïve Bayes Classifier*

Ojhi Gusti Marojahan Silaban (09021181320042)

Jurusan Teknik Informatika, Fakultas Ilmu Komputer

Universitas Sriwijaya

ABSTRAK

Klasifikasi dokumen bertujuan untuk mengelompokkan dokumen yang tidak terstruktur ke dalam kelompok-kelompok yang menggambarkan isi dari dokumen. Dokumen dapat berupa dokumen teks seperti artikel berita. Salah satu jenis berita yang banyak diminati adalah berita olahraga, khususnya oleh kaum pria. Banyaknya cabang olahraga membuat pembaca kebingungan mencari berita cabang olahraga yang ingin dibaca. Oleh karena itu, klasifikasi dokumen diperlukan untuk dapat mengelompokkan berita olahraga sesuai dengan kelompok cabang olahraga berita tersebut. Untuk mengatasi masalah tersebut, dikembangkan perangkat lunak menggunakan *Naïve Bayes Classifier* dengan komputasi yang sederhana dan tingkat akurasi yang cukup tinggi. *Naïve Bayes Classifier* memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. Terdapat tiga tahapan yaitu, praproses, pelatihan, dan klasifikasi. Penelitian ini menggunakan data sekunder yang diambil dari portal berita *online* dengan berbagai kategori yaitu sepakbola, bola basket, bulutangkis, motogp, formula1, dan MMA. Tingkat akurasi yang dihasilkan dari perangkat lunak sebesar 73,33% dengan menggunakan 30 dokumen latih dan 30 dokumen uji.

Kata Kunci : Klasifikasi dokumen, berita olahraga, *Naïve Bayes Classifier*

KATA PENGANTAR

Segala puji dan syukur saya persembahkan kepada Tuhan Yesus Kristus atas berkah dan karunia-Nya yang telah diberikan kepada Penulis sehingga dapat menyelesaikan Tugas Akhir ini dengan baik. Tugas Akhir yang berjudul “Klasifikasi Dokumen Berita Olahraga Menggunakan *Naïve Bayes Classifier*” disusun untuk memenuhi salah satu syarat dalam menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Jurusan Teknik Informatika di Universitas Sriwijaya.

Pada kesempatan ini Penulis ingin menyampaikan ucapan terima kasih kepada pihak-pihak yang telah memberikan dukungan, bimbingan dan motivasi selama proses penelitian ini dilaksanakan. Secara khusus Penulis ingin mengucapkan terima kasih kepada :

1. Tuhan Yesus Kristus atas segala berkah dan karunia-Nya.
2. Keluarga tercinta, Ibu dan Ayah yang selalu memberi semangat dan dukungan moral dan materil, serta selalu menyebut namaku didalam setiap doanya. Saudaraku, adik-adikku tercinta yang selalu mendukung dalam pengerjaan tugas akhir ini. Keluarga besar beserta kerabat yang juga selalu memberikan semangat dan dukungan.
3. Bapak Jaidan Jauhari, M.T., selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Ibu Alvi Syahrini Utami, M.Kom., selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.
5. Ibu Alvi Syahrini Utami, M.Kom., selaku Dosen Pembimbing I yang telah membimbing, mengarahkan dan memotivasi saya untuk menyelesaikan Tugas Akhir ini.
6. Bapak Dr. Abdiansah, S.Kom., M.CS., selaku Dosen Pembimbing II yang telah membimbing, mengarahkan dan memotivasi saya untuk menyelesaikan Tugas Akhir ini.
7. Ibu Novi Yuslian, M.T., selaku penguji I Tugas Akhir yang telah memberikan nasihat dan saran yang membangun.

8. Bapak Ali Ibrahim, M.T., selaku penguji II Tugas Akhir yang telah memberikan nasihat dan saran yang membangun.
9. Bapak Ali Ibrahim, M.T., selaku dosen pembimbing akademik yang selalu memberikan saran dan arahan selama masa perkuliahan.
10. Seluruh Dosen Jurusan Teknik Informatika yang telah memberikan bekal ilmu selama masa perkuliahan.
11. Teman seperjuangan Teknik Informatika Angkatan 2013.
12. Seluruh Staf Administrasi dan Pegawai yang telah membantu dalam urusan administrasi.

Penulis menyadari masih banyak terdapat kekurangan dalam penyusunan Tugas Akhir yang disebabkan oleh keterbatasan pengetahuan dan pengalaman. Oleh karena itu, kritik dan saran membangun sangat diharapkan untuk menyempurnakan Tugas Akhir ini. Semoga Tugas Akhir ini dapat bermanfaat bagi kita semua.

Indralaya, 11 Januari 2021

Ojhi Gusti Marojahan Silaban

DAFTAR ISI

| | |
|--|------|
| HALAMAN JUDUL..... | i |
| LEMBAR PENGESAHAN SKRIPSI | ii |
| TANDA LULUS UJIAN SIDANG TUGAS AKHIR | iii |
| HALAMAN PERNYATAAN | iv |
| MOTTO DAN PERSEMBAHAN | v |
| ABSTRACT | vi |
| ABSTRAK | vii |
| KATA PENGANTAR | viii |
| DAFTAR ISI..... | x |
| DAFTAR TABEL..... | xivv |
| DAFTAR GAMBAR | xv |
| DAFTAR LAMPIRAN..... | xvi |
| BAB 1 PENDAHULUAN | I-1 |
| 1.1 Pendahuluan | I-1 |
| 1.2 Latar Belakang Masalah..... | I-1 |
| 1.3 Rumusan Masalah | I-3 |
| 1.4 Tujuan Penelitian | I-4 |
| 1.5 Manfaat Penelitian | I-4 |
| 1.6 Batasan Masalah..... | I-4 |
| 1.7 Sistematika Penulisan..... | I-5 |
| 1.8 Kesimpulan | I-6 |
| BAB 2 KAJIAN LITERATUR..... | II-1 |
| 2.1 Pendahuluan | II-1 |
| 2.2 Landasan Teori..... | II-1 |
| 2.2.1 Berita | II-1 |
| 2.2.2 Klasifikasi Berita..... | II-2 |
| 2.3 Text Mining..... | II-3 |
| 2.3.1 Text Preprocessing | II-3 |
| 2.3.1.1 Tokenizing..... | II-4 |

| | |
|---|--------------|
| 2.3.1.2 Filtering | II-5 |
| 2.3.1.3 Stemming | II-5 |
| 2.3.2 Penemuan Pola (Pattern Discovery)..... | II-7 |
| 2.3.3 Kategorisasi dokumen..... | II-8 |
| 2.4 Klasifikasi | II-8 |
| 2.5 Naïve Bayes Classifier pada Klasifikasi Dokumen | II-9 |
| 2.6 Penelitian terkait..... | II-13 |
| BAB 3 METODELOGI PENELITIAN..... | III-1 |
| 3.1 Pendahuluan | III-1 |
| 3.2 Unit Penelitian..... | III-1 |
| 3.3 Metode Pengumpulan Data | III-1 |
| 3.3.1 Jenis Data | III-1 |
| 3.3.2 Sumber Data..... | III-2 |
| 3.3.3 Teknik Pengumpulan Data..... | III-2 |
| 3.4 Tahapan Penelitian | III-2 |
| 3.4.1 Menentukan Ruang Lingkup dan Unit Penelitian..... | III-3 |
| 3.4.2 Menentukan Dasar Teori yang berkaitan dengan Permasalahan | III-4 |
| 3.4.3 Menetapkan Kriteria Pengujian | III-4 |
| 3.4.4 Menentukan Alat yang Digunakan Untuk Penelitian..... | III-6 |
| 3.4.5 Melakukan Pengujian Penelitian..... | III-7 |
| 3.4.6 Melakukan Analisis Hasil Pengujian dan Membuat Kesimpulan..... | III-7 |
| 3.5 Metode Pengembangan Perangkat Lunak..... | III-8 |
| 3.5.1 Fase Inception | III-9 |
| 3.5.2 Fase Elaboration..... | III-9 |
| 3.5.3 Fase Contruction | III-10 |
| 3.5.4 Fase Trantition | III-10 |
| 3.6 Manajemen Proyek Penelitian..... | III-11 |
| BAB 4 PENGEMBANGAN PERANGKAT LUNAK..... | IV-1 |
| 4.1 Pendahuluan | IV-1 |
| 4.2 Fase Insepsi | IV-1 |
| 4.2.1 Pemodelan Bisnis | IV-1 |
| 4.2.2 Kebutuhan Sistem..... | IV-2 |

| | |
|--|------------|
| 4.2.3 Analisis dan Desain | IV-4 |
| 4.2.3.1 Analisis Perangkat Lunak | IV-4 |
| 4.2.3.2 Desain Perangkat Lunak | IV-6 |
| 4.3 Fase Elaborasi | IV-14 |
| 4.3.1 Pemodelan Bisnis | IV-14 |
| 4.3.1.1 Perancangan Data | IV-14 |
| 4.3.1.2 Perancangan Antar Muka..... | IV-14 |
| 4.3.2 Kebutuhan Sistem..... | IV-15 |
| 4.3.3 Diagram Sequence..... | IV-16 |
| 4.4 Fase Konstruksi..... | IV-20 |
| 4.4.1 Kebutuhan Sistem..... | IV-20 |
| 4.4.2 Diagram Kelas | IV-20 |
| 4.4.3 Implementasi | IV-22 |
| 4.4.3.1 Implementasi Kelas..... | IV-22 |
| 4.4.3.2 Implementasi Antarmuka | IV-24 |
| 4.5 Fase Transisi | IV-25 |
| 4.5.1 Pemodelan Bisnis | IV-25 |
| 4.5.2 Kebutuhan Sistem..... | IV-25 |
| 4.5.3 Rencana Pengujian | IV-26 |
| 4.5.3.1 Rencana Pengujian Use Case Melakukan Klasifikasi | IV-26 |
| 4.5.3.3 Rencana Pengujian Use Case Melakukan Pelatihan..... | IV-27 |
| 4.5.3.4 Rencana Pengujian Use Case Melakukan Praproses | IV-27 |
| 4.5.4 Implementasi | IV-28 |
| 4.5.4.1 Pengujian Use Case Melakukan Klasifikasi Berita | IV-29 |
| 4.5.4.2 Pengujian Use Case Melakukan Pelatihan..... | IV-30 |
| 4.5.4.3 Pengujian Use Case Melakukan Praproses | IV-31 |
| 4.6 Kesimpulan | IV-34 |
| BAB 5 HASIL DAN ANALISIS PENELITIAN..... | V-1 |
| 5.1 Pendahuluan | V-1 |
| 5.2 Hasil Penelitian | V-1 |
| 5.2.1 Konfigurasi Penelitian | V-2 |
| 5.3 Hasil Pengujian | V-2 |

| | |
|--|-------|
| 5.3.1 Analisi Tingkat Persentase Data yang Dikenali atau tidak dikenali..... | V-2 |
| 5.4 Analisis Hasil Pengujian | V-7 |
| 5.5 Kesimpulan | V-8 |
| KESIMPULAN DAN SARAN..... | VI-1 |
| 6.1 Kesimpulan | VI-1 |
| 6.2 Saran..... | VI-1 |
| DAFTAR PUSTAKA | xxvii |
| LAMPIRAN..... | L-1 |

DAFTAR TABEL

| | |
|--|--------|
| Tabel III-1 Spesifikasi Perangkat Keras | III-6 |
| Tabel III-2 Spesifikasi Perangkat Lunak..... | III-7 |
| Tabel III-3 Tabel Rancangan Hasil Klasifikasi Dokumen Berita Olahraga | III-8 |
| Tabel III-4 Tabel Penjadwalan Penelitian..... | III-12 |
| Tabel IV-1 Kebutuhan Fungsional..... | IV-3 |
| Tabel IV-2 Kebutuhan Non-Fungsional | IV-4 |
| Tabel IV-3 Definisi Aktor Use Case..... | IV-7 |
| Tabel IV-4 Definisi Use Case | IV-7 |
| Tabel IV-5 Skenario Use Case Melakukan Pelatihan..... | IV-8 |
| Tabel IV-6 Skenario Use Case Melakukan Klasifikasi..... | IV-9 |
| Tabel IV-7 Skenario Use Case Melakukan Praproses | IV-10 |
| Tabel IV-8 Implementasi Kelas | IV-22 |
| Tabel IV-9 Rencana Pengujian Use Case Melakukan Klasifikasi | IV-26 |
| Tabel IV-10 Rencana Pengujian Use Case Melakukan Pelatihan | IV-27 |
| Tabel IV-11 Rencana Pengujian Use Case Melakukan Praproses..... | IV-27 |
| Tabel IV-12 Pengujian Use Case Melakukan Klasifikasi Berita..... | IV-29 |
| Tabel IV-13 Pengujian Use Case Melakukan Pelatihan | IV-30 |
| Tabel IV-14 Pengujian Use Case Melakukan Praproses..... | IV-31 |
| Tabel V-1 Hasil Pengujian 30 Data Teks Berita Olahraga | V-3 |

DAFTAR GAMBAR

| | |
|---|-------|
| Gambar III-1 Diagram Blok Tahapan Penelitian | III-3 |
| Gambar III-2 Urutan Proses Pembangunan Perangkat Lunak | III-6 |
| Gambar IV-1 Representasi Dokumen Masukan | IV-2 |
| Gambar IV-2 Data Teks Berita Olahraga..... | IV-5 |
| Gambar IV-3 Diagram Use Case | IV-6 |
| Gambar IV-4 Diagram Aktivitas Melakukan Pelatihan..... | IV-12 |
| Gambar IV-5 Diagram Aktivitas Melakukan Klasifikasi Berita..... | IV-13 |
| Gambar IV-6 Diagram Aktivitas Melakukan Praproses | IV-14 |
| Gambar IV-7 Rancangan Antarmuka Form Klasifikasi..... | IV-15 |
| Gambar IV-8 Sequence Diagram Melakukan Klasifikasi Berita..... | IV-17 |
| Gambar IV-9 Sequence Diagram Melakukan Pelatihan | IV-18 |
| Gambar IV-10 Sequence Diagram Melakukan Praproses | IV-19 |
| Gambar IV-11 Diagram Kelas Perangkat Lunak | IV-21 |
| Gambar IV-12 Antarmuka Halaman Menu Utama Klasifikasi..... | IV-24 |
| Gambar IV-13 Output Kelas Ketika di Klasifikasi | IV-25 |
| Gambar V-1 Output Kelas Ketika di Klasifikasi | V-1 |
| Gambar V-2 Grafis Hasil Percobaan Data Uji Dikenali dan Tidak Dikenali .. | V-7 |

DAFTAR LAMPIRAN

Hasil Coding L-1

BAB I

PENDAHULUAN

1.1 Pendahuluan

Bab ini menjelaskan secara detail tentang latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan dokumen penelitian.

1.2 Latar Belakang Masalah

Penyebaran informasi berkembang dengan sangat pesat dalam dokumen berita online, dan terus mengalami perkembangan. Salah satu jenis berita yang banyak diminati adalah berita olahraga, khususnya oleh kaum pria. Banyaknya cabang olahraga sering membuat pembaca kebingungan mencari berita cabang olahraga yang ingin dibaca. Dokumen berita olahraga yang beredar di dunia maya terus bertambah dan cenderung kurang efektif dalam pencarian maupun penyajian informasinya (Hamzah dkk, 2008). Oleh karena itu, klasifikasi dokumen ke dalam kategori yang sesuai sangat diperlukan untuk meningkatkan efektifitas dan efisiensi dari dokumen teks. Hal ini mendukung diperlukannya proses klasifikasi terhadap informasi tersebut (Natalius, 2010)

Klasifikasi dokumen adalah proses pengelompokan dokumen sesuai dengan kategori yang dimilikinya. Klasifikasi dokumen merupakan masalah yang mendasar namun sangat penting karena manfaatnya cukup besar mengingat jumlah dokumen terus bertambah setiap hari. Klasifikasi dokumen bertujuan untuk

mengelompokkan dokumen yang tidak terstruktur ke dalam kelompok-kelompok yang menggambarkan isi dari dokumen. Dokumen dapat berupa dokumen teks seperti artikel berita.

Sejumlah penelitian dalam pengklasifikasian dokumen teks telah dilakukan sebelumnya dengan beberapa metode diantaranya; *K-Nearest Neighbour* (Pratama et al, 2018), dan *Decision Tree* (Parmawati, 2005).

Menurut (Nugraha dkk, 2018) algoritma *K-Nearest Neighbour* (KNN) merupakan sebuah metode untuk klasifikasi terhadap objek yang berdasarkan dari data yang jaraknya paling dekat dengan objek tersebut. Algoritma KNN merupakan sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Algoritma KNN adalah algoritma yang kompleks, tapi membutuhkan proses yang lama untuk pengklasifikasiannya.

Parmawati (2005) melakukan pengklasifikasian dokumen menggunakan *Decision Tree* dimana Abstrak dan *keyword* sebagai data masukan untuk proses pengklasifikasiannya menggunakan 30 atribut sebagai acuan pembuatan rulanya. Dengan menggunakan algoritma id3, rata-rata hasil akurasi pengklasifikasian jurnal kurang akurat. Selain akurasi pengklasifikasian yang masih dibawah angka 70%, algoritma id3 ini juga memakan waktu cukup lama untuk mengklasifikasikan dokumen teks karena algoritma ini dirancang untuk membuat tree untuk setiap rulanya.

Teknik yang juga banyak digunakan dalam klasifikasi dokumen adalah *Naïve Bayes Classifier* (NBC) yang memiliki beberapa kelebihan antara lain,

seederhana, cepat, dan berakurasi tinggi. Metode NBC pada klasifikasi dokumen teks menggunakan atribut kata yang muncul dalam satu dokumen sebagai dasar klasifikasinya. Algoritma klasifikasi *Naïve Bayes* memanfaatkan teori probabilitas yang ditemukan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. Metode ini memiliki kelebihan hanya memerlukan komputasi matematika yang tidak terlalu kompleks sehingga sangat efisien. Algoritma NBC yang sederhana dan kecepatannya yang tinggi dalam proses pelatihan dan klasifikasi membuat algoritma ini menarik untuk digunakan sebagai salah satu metode klasifikasi (Chen & Yap, 2006)

Berdasarkan hasil penelitian sebelumnya yang menyatakan bahwa *Naïve Bayes Classifier* memiliki tingkat akurasi yang tinggi, maka penelitian ini akan menerapkan metode *Naïve Bayes Classifier* dalam pengklasifikasian dokumen berita olahraga.

1.3 Rumusan Masalah

Banyaknya artikel berita olahraga yang ada dan miripnya kata per kata di berita olahraga yang berbeda membuat kesulitan membedakan berita cabang olahraga yang ingin dibaca, sehingga pengelompokkan beritan olahraga tersebut menjadi lambat dan kurang efisien. Sementara itu, *Naïve Bayes Classifier* berhasil diterapkan pada pengklasifikasian dokumen. Namun permasalahannya adalah bagaimana membangun perangkat lunak pengklasifikasian dokumen berita

olahraga dengan menerapkan *Naïve Bayes Classifier*, sehingga dapat mengelompokkan berita olahraga sesuai cabang olahraganya.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah:

1. Mengklasifikasikan dokumen berita olahraga yang belum diketahui kelompoknya dengan menggunakan metode *Naïve Bayes Classifier*.
2. Menganalisis tingkat akurasi hasil pengklasifikasian dokumen berita olahraga.

1.5 Manfaat Penelitian

Manfaat penelitian ini adalah menghasilkan perangkat lunak yang dapat mempermudah pengklasifikasian dokumen berita olahraga dan memperoleh informasi dengan cepat tanpa perlu membaca satu persatu dan mengetahui berita tiap cabang olahraga secara otomatis

1.6 Batasan Masalah

Penelitian ini menggunakan beberapa batasan masalah di antaranya :

1. Klasifikasi hanya dilakukan dalam berita berbahasa Indonesia.
2. Data dokumen yang digunakan dalam enam kategori yaitu sepakbola, bola basket, bulutangkis, MMA, motogp, formula1
3. Dokumen berita memiliki isi atau konten kategori yang tunggal.
4. File yang akan digunakan dalam klasifikasi beresktensi *.txt.

5. Jumlah artikel yang digunakan saat training sebanyak 30 artikel pada tiap kategori

1.7 Sistematika Penulisan

Sistematika penulisan tugas akhir ini mengikuti standar penulisan tugas akhir Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya, antara lain :

BAB I. PENDAHULUAN

Pada bab ini dibahas mengenai latar belakang, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan.

BAB II. KAJIAN LITERATUR

Pada bab ini akan dibahas mengenai dasar teori yang digunakan dalam analisis, perancangan, dan implementasi tugas akhir yang dilakukan pada bab-bab selanjutnya.

BAB III. METODOLOGI PENELITIAN

Pada bab ini akan dibahas mengenai tahapan yang akan dilaksanakan pada penelitian ini. Masing-masing rencana tahapan penelitian dideskripsikan dengan rinci dengan mengacu pada suatu kerangka kerja. Di akhir bab ini berisi perancangan manajemen proyek pada pelaksanaan penelitian.

1.8 Kesimpulan

Berdasarkan uraian di atas, maka dalam penelitian ini akan dikembangkan sebuah perangkat lunak untuk mengklasifikasikan dokumen berita olahraga menggunakan metode *Naïve Bayes Classifier*.

DAFTAR PUSTAKA

- Aliwy, A. H., & Ameer, E. H. A. (2017). Comparative study of five text classification algorithms with their improvements. *International Journal of Applied Engineering Research*, 12(14), 4309–4319. <https://doi.org/10.113/J.0973-4562>
- Asian, J., Williams, H. E., & Tahaghoghi, S. M. M. (2005). Stemming Indonesian. *Conferences in Research and Practice in Information Technology Series*, 38(January), 307–314. <https://doi.org/10.1145/1316457.1316459>
- Chen, L., & Yap, K. (2006). Arg Max. *Signal Processing*, 54(4), 1557–1562. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1608568
- Hamzah, A., Soesianto, F., Susanto, A., & Istiyanto, J. E. (2008). Studi Kinerja Fungsi-fungsi Jarak dan Similaritas dalam Clustering Dokumen Teks Berbahasa Indonesia. *Seminar Nasional Informatika 2008 (SemnasIF 2008) UPN "Veteran" Yogyakarta, 2008(semnasIF)*, 21–30.
- Indah Pratiwi, N., & Widodo. (2017). Klasifikasi Dokumen Karya Akhir Mahasiswa Menggunakan Naïve Bayes Classifier (NBC) Berdasarkan Abstrak Karya Akhir Di Jurusan Teknik Elektro Universitas Negeri Jakarta. *PINTER : Jurnal Pendidikan Teknik Informatika Dan Komputer*, 1(1), 31–38. <https://doi.org/10.21009/pinter.1.1.5>
- Kang, D. K., Silvescu, A., & Honavar, V. (2006). RNBL-MN: A recursive Naive Bayes learner for sequence classification. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes*

- in *Bioinformatics*), 3918 *LNAI*(Iis 0219699), 45–54.
https://doi.org/10.1007/11731139_8
- M., Asikin, M. F., Kurniawaty, D., Sari, S. K., & Cholissodin, I. (2016). Implementasi Metode Naïve Bayes Classifier Untuk Seleksi Asisten Praktikum Pada Simulasi Hadoop Multinode Cluster. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 3(4), 273.
<https://doi.org/10.25126/jtiik.201634227>
- Mahmudy, W. F., & Widodo, A. W. (2014). Klasifikasi Artikel Berita Menggunakan Naive Bayes Classifier yang Dimodifikasi. *Tekno*, 21, 1–10.
- Mózo, B. S. (2017). Pengaruh Seleksi Fitur Pada Skema Klasifikasi Naive Bayes Berbasis Gaussian dan Kernel Density. *Journal of Chemical Information and Modeling*, 53(9), 1689–1699.
<https://doi.org/10.1017/CBO9781107415324.004>
- Natalius, S. (2010). *Makalah II2092 Probabilitas dan Statistik-Sem. I Tahun. 3*.
- Nugraha, pratama dwi,, Said al faraby, A. (2018). Klasifikasi Dokumen Menggunakan Metode Knn Dengan Information Gain. *EProceedings of Engineering*, 5(1), 1541–1550.
- Pandhu, A., & Agus, H. (2016). Naive Bayes Classification pada Klasifikasi Dokumen Untuk Identifikasi Konten E-Government. *Journal of Applied Intelligent System*, 1(1), 48-55–55.
- Pramudita, Y. D., Putro, S. S., & Makhmud, N. (2018). Klasifikasi Berita Olahraga Menggunakan Metode Naïve Bayes dengan Enhanced Confix Stripping Stemmer. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 5(3), 269.

<https://doi.org/10.25126/jtiik.201853810>

Rivki, M., & Bachtiar, A. M. (2017). Implementasi Algoritma K-Nearest Neighbor Dalam Pengklasifikasian Follower Twitter Yang Menggunakan Bahasa Indonesia. *Jurnal Sistem Informasi*, 13(1), 31.

<https://doi.org/10.21609/jsi.v13i1.500>

Setiawan, A., Kurniawan, E., & Handiwidjojo, W. (2013). Implementasi Stop Word Removal Untuk Pembangunan Aplikasi Alkitab Berbasis Windows 8. *Jurnal EKSIS*, 6(2), 1–11.

Wong, A. K. C., & Wang, Y. (2003). Pattern discovery: A data driven approach to decision support. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, 33(1), 114–124.

<https://doi.org/10.1109/TSMCC.2003.809869>