

***PERBANDINGAN METODE K-NEAREST NEIGHBOR (K-NN) DAN
SUPPORT VECTOR MACHINE (SVM) DALAM KLASIFIKASI DOKUMEN
MULTI-LABEL (STUDI KASUS TUGAS AKHIR MAHASISWA S1
INFORMATIKA UNS)***

SKRIPSI

**Diajukan untuk memenuhi sebagian persyaratan mendapatkan gelar Strata Satu
Program Studi Informatika**



Disusun Oleh:

AHMAD SYARIFUZ ZAKI

M0515002

PROGRAM STUDI INFORMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS SEBELAS MARET

2020

***PERBANDINGAN METODE K-NEAREST NEIGHBOR (K-NN) DAN
SUPPORT VECTOR MACHINE (SVM) DALAM KLASIFIKASI DOKUMEN
MULTI-LABEL (STUDI KASUS TUGAS AKHIR MAHASISWA S1
INFORMATIKA UNS)***

SKRIPSI

**Diajukan untuk memenuhi sebagian persyaratan mendapatkan gelar Strata Satu
Program Studi Informatika**



Disusun Oleh:

AHMAD SYARIFUZ ZAKI

M0515002

PROGRAM STUDI INFORMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS SEBELAS MARET

2020

HALAMAN PERSETUJUAN

SKRIPSI

**PERBANDINGAN METODE *K-NEAREST NEIGHBOR* (K-NN) DAN *SUPPORT VECTOR MACHINE* (SVM) DALAM KLASIFIKASI DOKUMEN *MULTI-LABEL*
(STUDI KASUS TUGAS AKHIR MAHASISWA S1 INFORMATIKA UNS)**

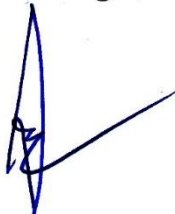
Disusun oleh:

AHMAD SYARIFUZ ZAKI

NIM. M0515002

**Skripsi ini telah disetujui untuk dipertahankan di hadapan dewan penguji
pada tanggal, 20 Januari 2020**

Pembimbing 1,



Dr. Wisnu Widiarto, S.Si, M.T.

NIP. 197006012008011009




Pembimbing 2,

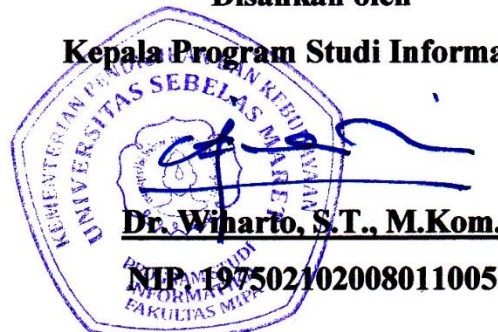


Esti Suryani, S.Si, M.Kom

NIP. 197611292008122001

HALAMAN PENGESAHAN**SKRIPSI****PERBANDINGAN METODE *K-NEAREST NEIGHBOR* (K-NN) DAN *SUPPORT VECTOR MACHINE* (SVM) DALAM KLASIFIKASI DOKUMEN *MULTI-LABEL*
(STUDI KASUS TUGAS AKHIR MAHASISWA S1 INFORMATIKA UNS)****Disusun oleh:****AHMAD SYARIFUZ ZAKI****NIM. M0515002****Telah dipertahankan dihadapan Dewan Penguji****Pada tanggal : 28 Januari 2020****Susunan Dewan Penguji**

- | | | |
|--|--------------|---|
| 1. <u>Dr. Wisnu Widiarto, S.Si, M.T.</u>
NIP. 197006012008011009 | (Ketua) | () |
| 2. <u>Esti Suryani, S.Si, M.Kom</u>
NIP. 197611292008122001 | (Sekretaris) | () |
| 3. <u>Dr. Umi Salamah, S.Si., M.Kom.</u>
NIP. 197002171997022001 | (Anggota) | () |
| 4. <u>Dr. Wiharto, S.T., M.Kom.</u>
NIP. 197502102008011005 | (Anggota) | () |

Disahkan oleh**Kepala Program Studi Informatika****Dr. Wiharto, S.T., M.Kom.****NIP. 197502102008011005**

PERSEMBAHAN

Skripsi ini saya persembahkan untuk:

Ibu, Bapak, Kakak, Adik, dan semua orang yang telah memberikan dukungan moril dan materiil.



***PERBANDINGAN METODE K-NEAREST NEIGHBOR (K-NN) DAN
SUPPORT VECTOR MACHINE (SVM) DALAM KLASIFIKASI DOKUMEN
MULTI-LABEL (STUDI KASUS TUGAS AKHIR MAHASISWA S1
INFORMATIKA UNS)***

AHMAD SYARIFUZ ZAKI

Program Studi Informatika

Fakultas Matematika dan Ilmu Pengetahuan Alam

Universitas Sebelas Maret

ABSTRAK

Klasifikasi adalah proses pengelompokan objek yang memiliki karakteristik atau ciri yang sama ke dalam beberapa kelas. Klasifikasi dokumen dapat dilakukan dengan menggunakan ciri atau fitur kata yang muncul pada dokumen latih. Namun apabila suatu ciri yang menjadi ciri dari satu dokumen merupakan ciri dari dokumen lainnya maka dokumen tersebut memiliki lebih dari satu topik atau disebut *multi-label*. Penelitian ini membandingkan metode klasifikasi *K-Nearest Neighbor* dengan *Support Vector Machine* dalam klasifikasi dokumen multi-label. Menggunakan metode pendekatan *Problem Transformation* yaitu *Binary Relevance (BR)* dan *Label Powerset (LP)* untuk mengadaptasi metode *K-Nearest Neighbor* dan *Support Vector Machine* dalam membangun klasifikasi. Dataset yang digunakan adalah skripsi mahasiswa S1 Informatika UNS. Penelitian ini diuji dengan menggunakan validasi *10-fold cross validation*. Sehingga nilai performa terbaik untuk accuracy dan hamming loss pada SVM sebesar 0.66 dan 0.14 sedangkan untuk KNN sebesar 0.64 dan 0.17. Menurut hasil tersebut, metode SVM memiliki performa yang lebih baik dari pada metode KNN.

Kata kunci: klasifikasi, *multi-label*, *k-nearest neighbor*, *support vector machine*, *problem Transformation*, skripsi.

***PERBANDINGAN METODE K-NEAREST NEIGHBOR (K-NN) DAN
SUPPORT VECTOR MACHINE (SVM) DALAM KLASIFIKASI DOKUMEN
MULTI-LABEL (STUDI KASUS TUGAS AKHIR MAHASISWA S1
INFORMATIKA UNS)***

AHMAD SYARIFUZ ZAKI

**Department of Informatics
Faculty of Mathematics and Natural Sciences
Universitas Sebelas Maret**

ABSTRACT

Classification is the process of grouping objects that have the same characteristics or characteristics into several classes. Document classification can be done using the features or words that appear on the training document. However, this document is a multi-label document. This study compares the K-Nearest Neighbor classification method with the Support Vector Machine in multi-label document classification. Use the method used. Problem Transformation namely Binary Relevance (BR) and Label Powerset (LP) to adapt the K-Nearest Neighbor and Support Vector Machine methods in building integration. The dataset used is a thesis UNS Informatics S1 student. This research was tested using 10-fold cross validation. So that the best performance value for accuracy and hamming loss in SVM is 0.66 and 0.14 while for KNN is 0.64 and 0.17. According to these results, the SVM method has a better performance than the KNN method.

Keyword: *classification, multi-label, k-nearest neighbor, support vector machine, problem transformation, skripsi.*

KATA PENGANTAR

Puji syukur kepada Allah SWT atas segala limpahan rahmat dan karunia-Nya, penulis dapat menyelesaikan penulisan skripsi ini. Sholawat dan salam senantiasa penulis haturkan kepada Rasulullah SAW sebagai pembimbing seluruh umat manusia

Skripsi ini tidak akan selesai tanpa adanya bantuan dari banyak pihak, karena itu penulis menyampaikan terima kasih kepada:

1. Bapak Dr. Wiharto, S.T., M.Kom. selaku Kepala Program Studi Informatika yang telah memberikan dukungan selama proses perkuliahan dan penyusunan Tugas Akhir.
2. Bapak Dr. Wisnu Widiarto, S.Si., M.T. selaku dosen pembimbing I atas ilmu, waktu, dan masukan yang diberikan, serta kesediaan untuk membimbing penulis selama pelaksanaan Tugas Akhir.
3. Ibu Esti Suryani, S.Si., M.Kom. selaku dosen pembimbing II atas ilmu, waktu, dan masukan yang diberikan, serta kesediaan membimbing penulis selama penulisan Tugas Akhir.
4. Orang Tua dan keluarga penulis yang senantiasa memberikan dukungan dan doa kepada penulis untuk menyelesaikan Tugas Akhir.
5. Teman-teman Informatika 2015 atas segala kebersamaan dan dukungannya selama perkuliahan di Program Studi Informatika.
6. Semua pihak yang tidak dapat penulis sebutkan satu persatu.

Penyusunan skripsi ini tentunya masih ada beberapa kekurangan. Oleh sebab itu, kritik dan saran pembaca sangat diperlukan. Semoga skripsi ini bisa bermanfaat bagi pembaca maupun penulis sendiri.

Surakarta, 20 Januari 2020

Penulis

DAFTAR ISI

HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
MOTTO	iii
PERSEMBAHAN.....	iv
ABSTRAK.....	v
ABSTRACT.....	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR LAMPIRAN.....	xiii
DAFTAR GAMBAR.....	xiv
BAB I PENDAHULUAN.....	Error! Bookmark not defined.
1.1. Latar Belakang	Error! Bookmark not defined.
1.2. Rumusan Masalah	Error! Bookmark not defined.
1.3. Batasan Masalah.....	Error! Bookmark not defined.
1.4. Tujuan Penelitian.....	Error! Bookmark not defined.
1.5. Manfaat Penelitian.....	Error! Bookmark not defined.
1.6. Sistematika Penulisan.....	Error! Bookmark not defined.
BAB II TINJAUAN PUSTAKA	Error! Bookmark not defined.
2.1. Dasar Teori.....	Error! Bookmark not defined.
2.1.1. <i>Text Mining</i>	Error! Bookmark not defined.
2.1.2. <i>Text Preprocessing</i>	Error! Bookmark not defined.

2.1.3.	<i>Term Frequency-Inverse Document Frequency (TF-IDF)</i> ..	Error! Bookmark not defined.
2.1.4.	<i>Multi-label Classification</i>	Error! Bookmark not defined.
2.1.5.	<i>Label Powerset (LP)</i>	Error! Bookmark not defined.
2.1.6.	<i>Binary Relevance (BR)</i>	Error! Bookmark not defined.
2.1.7.	<i>Support Vector Machine (SVM)</i>	Error! Bookmark not defined.
2.1.8.	<i>K-Nearest Neighbor (KNN)</i>	Error! Bookmark not defined.
2.1.9.	<i>K-Fold Cross Validation</i>	Error! Bookmark not defined.
2.1.10.	<i>Confusion Matrix</i>	Error! Bookmark not defined.
2.2.	Penelitian Terkait	Error! Bookmark not defined.
BAB III METODOLOGI PENELITIAN		Error! Bookmark not defined.
3.1.	Pengumpulan Dataset	Error! Bookmark not defined.
3.2.	<i>Text Preprocessing</i>	Error! Bookmark not defined.
3.3.	Pembobotan Kata	Error! Bookmark not defined.
3.4.	<i>Problem Tranformation</i>	Error! Bookmark not defined.
3.5.	Klasifikasi SVM dan KNN	Error! Bookmark not defined.
3.6.	Evaluasi	Error! Bookmark not defined.
BAB IV HASIL DAN PEMBAHASAN		Error! Bookmark not defined.
4.1.	Pengumpulan Dataset	Error! Bookmark not defined.
4.2.	Text Preprocessing	Error! Bookmark not defined.
4.3.	Pembobotan Kata	Error! Bookmark not defined.
4.4.	Problem Tranformation	Error! Bookmark not defined.
4.5.	Klasifikasi menggunakan SVM dan KNN	Error! Bookmark not defined.
4.5.1.	Skenario BR+KNN	Error! Bookmark not defined.
4.5.2.	Skenario LP+KNN.....	Error! Bookmark not defined.

4.5.3. Skenario BR+SVM.....**Error! Bookmark not defined.**

4.5.4. Skenario LP+SVM.....**Error! Bookmark not defined.**

4.6. Evaluasi**Error! Bookmark not defined.**

BAB V PENUTUP**Error! Bookmark not defined.**

5.1. Kesimpulan.....**Error! Bookmark not defined.**

5.2. Saran.....**Error! Bookmark not defined.**

DAFTAR PUSTAKA**Error! Bookmark not defined.**

LAMPIRAN.....**Error! Bookmark not defined.**



DAFTAR TABEL

Tabel 2.1 Satu set contoh multi-label	Error! Bookmark not defined.
Tabel 2.2 Format Label Powerset	Error! Bookmark not defined.
Tabel 2.3 Contoh SVM biner model One-vs-One	Error! Bookmark not defined.
Tabel 2.4 Confusion Matrix	Error! Bookmark not defined.
Tabel 4.1 Contoh dataset D1 single-label	Error! Bookmark not defined.
Tabel 4.2 Contoh dataset D2 multi-label	Error! Bookmark not defined.
Tabel 4.3 Proses case folding dataset D1.....	Error! Bookmark not defined.
Tabel 4.4 Proses case folding dataset D2.....	Error! Bookmark not defined.
Tabel 4.5 Proses filtering dataset D1	Error! Bookmark not defined.
Tabel 4.6 Proses filtering dataset D2	Error! Bookmark not defined.
Tabel 4.7 Proses tokenizing dataset D1	Error! Bookmark not defined.
Tabel 4.8 Proses tokenizing dataset D2	Error! Bookmark not defined.
Tabel 4.9 Hasil stopword removal dataset D1	Error! Bookmark not defined.
Tabel 4.10 Proses stopword removal dataset D2.....	Error! Bookmark not defined.
Tabel 4.11 Proses stemming dataset D1	Error! Bookmark not defined.
Tabel 4.12 Proses stemming dataset D2	Error! Bookmark not defined.
Tabel 4.13 Perhitungan TF-IDF dataset D1 dan D2	Error! Bookmark not defined.
Tabel 4.13 Perhitungan TF-IDF dataset D1 dan D2 (lanjutan)	Error! Bookmark not defined.
Tabel 4.13 Perhitungan TF-IDF dataset D1 dan D2 (lanjutan)	Error! Bookmark not defined.
Tabel 4.13 Perhitungan TF-IDF dataset D1 dan D2 (lanjutan)	Error! Bookmark not defined.
Tabel 4.14 Proses binary relevance dataset D1 dan D2.....	Error! Bookmark not defined.
Tabel 4.15 Proses label powerset dataset D1 dan D2	Error! Bookmark not defined.
Tabel 4.16 Persebaran dataset metode label powerset.....	Error! Bookmark not defined.
Tabel 4.17 Contoh vektor persebaran data	Error! Bookmark not defined.
Tabel 4.18 Transformasi label ke bentuk binary	Error! Bookmark not defined.
Tabel 4.19 Hasil perhitungan Euclidean Distance metode BR+KNN	Error! Bookmark not defined.
Tabel 4.20 Mengurutkan data training metode BR+KNN	Error! Bookmark not defined.
Tabel 4.21 Hasil mayoritas label metode BR+KNN	Error! Bookmark not defined.

Tabel 4.22 Hasil perhitungan Euclidenan Distance metode LP+KNN..... **Error! Bookmark not defined.**

Tabel 4.23 Mengurutkan data training metode LP+KNN**Error! Bookmark not defined.**

Tabel 4.24 Dekomposisi 5 kelas ke 10 kelas biner**Error! Bookmark not defined.**

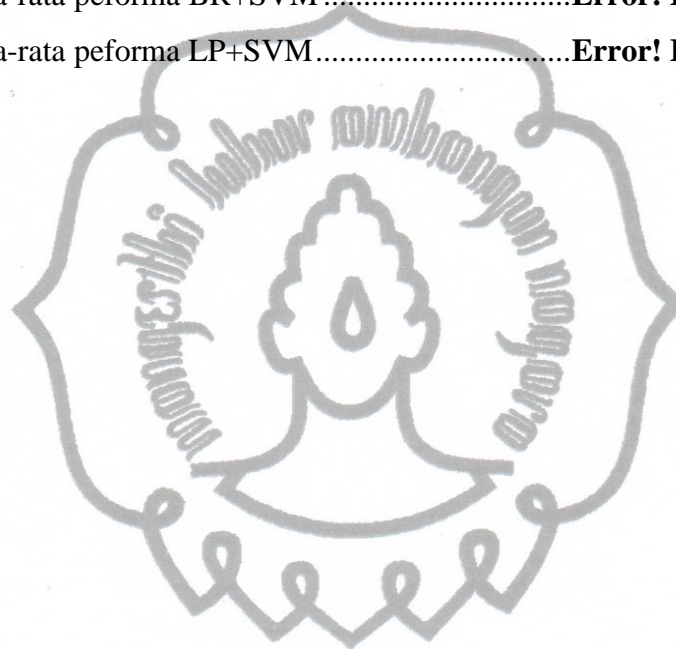
Tabel 4.25 Contoh perbandingan label prediksi dengan label kebenaran ... **Error! Bookmark not defined.**

Tabel 4.26 Hasil rata-rata performa BR+KNN.....**Error! Bookmark not defined.**

Tabel 4.27 Hasil rata-rata performa LP+KNN**Error! Bookmark not defined.**

Tabel 4.28 Hasil rata-rata peforma BR+SVM**Error! Bookmark not defined.**

Tabel 4.29 Hasil rata-rata peforma LP+SVM.....**Error! Bookmark not defined.**



DAFTAR LAMPIRAN

Lampiran 1 Confusion Matrix**Error! Bookmark not defined.**

Lampiran 2 Label Prediksi LP+SVM C=5,7,10 Fold 1**Error! Bookmark not defined.**



DAFTAR GAMBAR

Gambar 2.1 Tranformasi Model Binary Relevance	Error! Bookmark not defined.
Gambar 2.2 SVM berusaha mencari hyperplane terbaik	Error! Bookmark not defined.
Gambar 2.3 Margin pada hyperlane.....	Error! Bookmark not defined.
Gambar 2.4 Contoh klasifikasi dengan metode One-vs-One.....	Error! Bookmark not defined.
Gambar 3.1 Metodologi Penelitian	Error! Bookmark not defined.
Gambar 3.2 Proses Case Folding	Error! Bookmark not defined.
Gambar 3.3 Proses filtering	Error! Bookmark not defined.
Gambar 3.4 Proses Tokenizing	Error! Bookmark not defined.
Gambar 3.5 Proses Stopword Removal	Error! Bookmark not defined.
Gambar 3.6 Proses stemming	Error! Bookmark not defined.
Gambar 4.1 Contoh perbandingan data label CS dan Non-CS metode BR+SVM.....	Error! Bookmark not defined.
Gambar 4.2 Contoh perbandingan data label CS dan label NDS metode LP+SVM.....	Error! Bookmark not defined.
Gambar 4.3 Contoh hasil voting dari metode LP+SVM	Error! Bookmark not defined.
Gambar 4.4 Confusion Matrix BR+SVM C=3 Label Computational Science...	Error! Bookmark not defined.
Gambar 4.5 Confusion Matrix BR+SVM C=3 Label Network Data & Security.....	Error! Bookmark not defined.
Gambar 4.6 Confusion Matrix BR+SVM C=3 Label Sotf Computing	Error! Bookmark not defined.
Gambar 4.7 Confusion Matrix BR+SVM C=3 Label Information System .	Error! Bookmark not defined.
Gambar 4.8 Confusion Matrix LP+SVM C=3.....	Error! Bookmark not defined.
Gambar 4.9 Peforma terbaik tiap skenario	Error! Bookmark not defined.