ARTIKEL

SISTEM BANTU KLASIFIKASI ABSTRAK SKRIPSI MENGGUNAKAN NAIVE BAYES CLASSIFIER

(Studi Kasus: Program Studi Teknik Informatika Universitas Nusantara PGRI Kediri)

Oleh:

MOCHAMMAD RIDHO AL-FATAH

13.1.03.023.0120

Dibimbing oleh :

1. **PATMI KASIH, M.KOM**
2. **MADE AYU DUSEA WIDYA DARA, M.KOM**

TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS NUSANTARA PGRI KEDIRI

2018

SURAT PERNYATAAN
ARTIKEL SKRIPSI TAHUN 2018

Yang bertanda tangan di bawah ini:

Nama Lengkap	: MOCHAMMAD RIDHO AL-FATAH
NPM	: 13.1.03.02.0120
Telepon/HP	: 085784578410
Alamat Surel (Email)	: oneball59@gmail.com
Judul Artikel	: SISTEM BANTU KLASIFIKASI ABSTRAK SKRIPSI MENGGUNAKAN NAIVE BAYES CLASSIFIER (Studi Kasus: Program Studi Teknik Informatika Universitas Nusantara PGRI Kediri)
Fakultas – Program Studi	: TEKNIK – TEKNIK INFORMATIKA
Nama Perguruan Tinggi	: UNIVERSITAS NUSANTARA PGRI KEDIRI
Alamat Perguruan Tinggi	: Jalan K.H. Achmad Dahlan 76 Telepon (0354) 771576 Kediri

Dengan ini menyatakan bahwa :

a. artikel yang saya tulis merupakan karya saya pribadi (bersama tim penulis) dan bebas plagiarisme;
b. artikel telah diteliti dan disetujui untuk diterbitkan oleh Dosen Pembimbing I dan II.

Demikian surat pernyataan ini saya buat dengan sesungguhnya. Apabila di kemudian hari ditemukan ketidaksesuaian data dengan pernyataan ini dan atau ada tuntutan dari pihak lain, saya bersedia bertanggungjawab dan diproses sesuai dengan ketentuan yang berlaku.

Mengetahui		Kediri, 16 Agustus 2018
Pembimbing I	Pembimbing II	Penulis,
Patmi Kasih, M.Kom NIDN. 0701107802	Made Ayu Dusea Widya Dara, M.Kom NIDN. 0729088802	Mochammad Ridho Al-fatah NPM. 13.1.03.02.0120

SISTEM BANTU KLASIFIKASI ABSTRAK SKRIPSI MENGGUNAKAN NAIVE BAYES CLASSIFIER (Studi Kasus: Program Studi Teknik Informatika Universitas Nusantara PGRI Kediri)

Mochammad Ridho Al-Fatah
13.1.03.02.0120
Fakultas Teknik – Prodi Teknik Informatika
Oneball59@gmail.com
Patmi Kasih, M.Kom dan Made Ayu Dusea Widya Dara, M.Kom
UNIVERSITAS NUSANTARA PGRI KEDIRI

ABSTRAK

Mahasiswa yang menyelesaikan skripsi diwajibkan untuk menyerahkan laporan hasil penelitian skripsi kepada petugas program studi untuk di simpan di perpustakaan sehingga dapat di gunakan oleh mahasiswa lain sebagai referensi. Namun dengan banyaknya laporan skripsi dan bidang penelitian yang berbeda, sehingga petugas program studi mengalami kesulitan di dalam pengarsipan. Pengarsipan skripsi dilakukan berdasarkan pengelompokan bidang penelitian, sehingga setiap laporan skripsi yang di serahkan akan di klasifikasikan berdasarkan kelompok bidang penelitian.

Salah satu metode yang dapat digunakan untuk melakukan klasifikasi abstrak yaitu algoritma metode *Naïve Bayes Classifier*. *Naïve Bayes Classifier* adalah merupakan sebuah metode klasifikasi yang berakar pada Teorema Bayes. Ciri utama dari *Naive Bayes Classifier* ini adalah asumsi yang sangat kuat (naif) akan independensi dari masing-masing kondisi/kejadian. Sebelum menjelaskan *Naive Bayes Classifier* ini, akan independensi dari masing-masing kondisi/kejadian.

Setiap abstrak skripsi yang diuji akan melalui beberapa tahapan dalam pengklasifikasian. Tahap pertama yaitu setiap abstrak skripsi akan masuk pada tahap preprocessing dimana di dalamnya meliputi proses tokenizing (pemecahan kalimat menjadi kata per kata dan karakter tidak penting), wordlist (pengambilan kata-kata penting), stemming (pengambilan suku kata). Setelah tahap preprocessing telah selesai maka, tahap selanjutnya ialah pembentukan *matrix term* yang nantinya digunakan dalam proses penghitungan menggunakan *Naive Bayes Classifier*, hasil dari perhitungan inilah yang digunakan untuk mengetahui hasil klasifikasi abstrak skripsi yang diuji.

Kesimpulan hasil penelitian ini adalah yaitu sistem bantu yang dibangun belum bekerja dengan maksimal. Di karenakan ada kesalahan pada pembentukan pada kriteria dan perhitungan yang masih salah. Maka dari itu, berdasarkan pengujian yang telah dilakukan sistem menghasilkan nilai akurasi sebesar 33%.

KATA KUNCI : *Naive Bayes Classifier*, *Text Mining*, klasifikasi, abstrak skripsi.

I. LATAR BELAKANG

Mahasiswa yang menyelesaikan skripsi diwajibkan untuk menyerahkan laporan hasil penelitian skripsi kepada petugas program studi untuk disimpan di perpustakaan sehingga dapat digunakan oleh mahasiswa lain sebagai referensi. Namun dengan banyaknya laporan skripsi dan bidang penelitian yang berbeda, sehingga petugas program studi mengalami kesulitan di dalam pengarsipan. Pengarsipan skripsi dilakukan berdasarkan pengelompokan bidang penelitian, sehingga setiap laporan skripsi yang di serahkan akan di klasifikasikan berdasarkan kelompok bidang penelitian.

Berdasarkan permasalahan di atas, untuk menanggulangi hal tersebut perlu adanya sistem yang dapat digunakan untuk klasifikasi abstrak skripsi secara otomatis. Salah satu metode yang dapat digunakan untuk pengklasifikasian abstrak skripsi yaitu metode *Naive Bayes Clasiffier*.

Dengan metode tersebut, diharapkan pengklasifikasian abstrak skripsi yang diajukan oleh mahasiswa akan lebih akurat dan mempermudah tugas dari petugas akademis Program Studi Akuntansi Universitas Nusantara PGRI Kediri.

II. METODE

A. Metode Penelitian

Tahapan–tahapan kegiatan dalam penelitian ini yang akan dilakukan dalam proses pengerjaan skripsi ini adalah :

1 Pendekatan dan Teknik Penelitian

Pendekatan penelitian yang digunakan pada penelitian ini adalah pendekatan kualitatif, dimana penelitian melakukan analisa terhadap akurasi metode *Naive Bayes Classifier* dalam hal klasifikasi abstrak skripsi. Penelitian ini juga merupakan penelitian eksperimental, karena untuk mendapatkan akurasi terbaik algoritma *Naive Bayes Classifier* dalam mengklasifikasikan, dilakukan beberapa kali percobaan dengan parameter yang berbeda-beda. Teknik penelitian yang sesuai adalah penelitian pengembangan atau rekayasa teknologi informasi.

2 Prosedur Penelitian

Merupakan langkah-langkah yang dikerjakan atau kegiatan yang dilakukan dalam melakukan suatu penelitian. Prosedur yang dilakukan dalam penelitian ini yaitu :

a. Sumber pustaka dari buku dan jurnal yang berkaitan dengan klasifikasi abstrak yang menggunakan metode *Naive Bayes Classifier*.

b. Mendefinisikan dan mengumpulkan semua kebutuhan kemudian kebutuhan tersebut dianalisis dan didefinisikan.

c. Merancang sistem merupakan tahap penyusunan proses, data, aliran proses dan hubungan antar data.

d. Implementasi sistem.

B. Text Mining

Menurut Konchady (2006), mengartikan bahwa text mining ialah sebagai berikut:

> *Text mining* adalah sebuah penerapan yang berasal dari *information retrieval* (IR) dan *natural language processing* (NLP). Definisi *text mining* secara sempit hanya berupa metode yang dapat menemukan informasi baru yang tidak jelas atau mudah diketahui dari sebuah kumpulan dokumen. Sedangkan secara lebih luas, *text mining* mencakup teknik *text-processing* yang lebih umum, seperti pencarian, pengambilan intisari, dan pengkategorian.

Menurut Triawati (2009), permasalahan yang dihadapi pada text mining ialah sebagai berikut:

> Permasalahan yang dihadapi pada *text mining* seperti halnya *data mining* yaitu jumlah data yang besar, dimensi yang tinggi, data dan struktur yang terus berubah, serta data *noise*, sehingga sumber data yang digunakan pada *text mining* adalah kumpulan teks yang memiliki bentuk yang tidak terstruktur atau setidaknya semi terstruktur. Tujuan dari *text mining* adalah untuk mendapatkan informasi yang berguna dari sekumpulan dokumen dalam bentuk teks.

C. Text Preprocessing

Menurut Feldman & Sanger (2007), menjelaskan beberapa tahap ialah sebagai berikut:

> Dalam melakukan *text mining*, teks dokumen yang digunakan harus dipersiapkan terlebih dahulua, setelah itu baru dapat digunakan untuk proses utama. Proses mempersiapkan teks dokumen atau *dataset* mentah disebut juga dengan proses *text preprocessing*. *Text preprocessing* berfungsi untuk mengubah data teks yang tidak terstruktur atau sembarang menjadi data yang terstruktur. Secara umum proses yang dilakukan dalam tahapan *preprocessing* adalah sebagai berikut:
>
> 1. ***Case Folding***
> *Case Folding* adalah proses penyamaan case dalam sebuah dokumen. Hal ini dilakukan untuk mempermudah pencarian. Tidak semua dokumen teks konsisten dalam penggunaan huruf kapital. Oleh karena itu peran *case folding* dibutuhkan dalam menjadi suatu bentuk standar (dalam hal ini huruf kecil dan *lowercase*).
> 2. ***Tokenizing***
> *Tokenizing* adalah proses pemotongan sebuah dokumen menjadi bagian-bagian, yang disebut dengan token. Pada saat bersamaan *tokenizing* juga berfungsi untuk membuang beberapa karakter tertentu yang dianggap sebagai tanda baca.
> 3. ***Stopword Removal***
> *"Stopword Removal* adalah proses menghilangkan kata-kata

yang tidak berkontribusi banyak pada isi dokumen" (Yates & Neto, 1999). Kata-kata yang termasuk ke dalam *stopword* dihilangkan karena memberikan pengaruh yang tidak baik dalam proses *text mining* seperti kata-kata "and", "i", "you", "with", "she", "he", dan lain-lain.

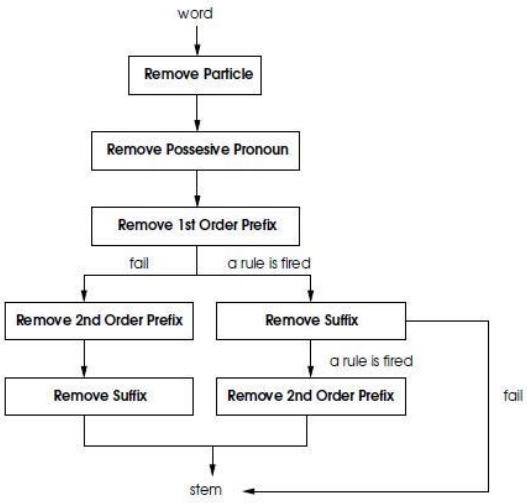
4. ***Stemming***
   Stemming adalah suatu proses pengembalian suatu kata berimbuhan ke dalam bentuk dasarnya (*root*). *Stemming* adalah alat pemprosesan teks dasar yang sering digunakan untuk meningkatkan kinerja pada *text retrieval* dan *text classification*. Namun sama pada halnya *stopword*, kinerja *stemming* juga bervariasi dan sering bergantung pada domain bahasa yang digunakan.

Stemming Algoritma Porter

Menurut Fadillah (2003), mengartikan bahwa *stemming* ialah sebagai berikut:

> *Stemming* merupakan proses yang memetakan bentuk varian kata menjadi kata dasarnya. Dalam pengembangan aplikasi *stemming* dokumen teks berbahasa indonesia menggunakan bahasa pemrograman PHP dan MySql sebagai DBMS (*database management system*).

Alur proses dari algoritma Porter diperlihatkan pada gambar 1 berikut:



Gambar 1 Proses stemming algoritma Porter

Menurut Agusta dalam Lasmedi Afuan (2013), adapun langkah-langkah pada algoritma porter adalah sebagai berikut:

1. Hapus *Particle*,
2. Hapus Possesive Pronoun.
3. Hapus awalan pertama. Jika tidak ada lanjutkan ke langkah 4a, jika ada cari maka lanjutkan ke langkah 4b.
4. a. Hapus awalan kedua, lanjutkan ke langkah 5a.
   b. Hapus akhiran, jika tidak ditemukan maka kata tersebut diasumsikan sebagai *root word.* Jika ditemukan maka lanjutkan ke langkah 5b.
5. a. Hapus akhiran. Kemudian kata akhir diasumsikan sebagai *root word*
   b. Hapus awalan kedua. Kemudian kata akhir diasumsikan sebagai *root word*.

Metode Algoritma Naive Bayes Clasiffier

Dalam penelitian Agus Mulyanto (2009), pengartian *Naive Bayes Classifier* ialah sebagai berikut:

> Merupakan sebuah metode klasifikasi yang berakar pada Teorema Bayes. Ciri utama dari *Naive Bayes Classifier* ini adalah asumsi yang sangat kuat (naif) akan independensi dari masing-masing kondisi/kejadian. Sebelum menjelaskan *Naive Bayes Classifier* ini, akan independensi dari masing-masing kondisi/kejadian. Sebelum menjelaskan *Naive Bayes Classifier* ini, akan dijelaskan terlebih dahulu *Teorema Bayes* yang menjadi dasar dari metode tersebut. Pada *teorema Bayes*, bila terdapat dua kejadian yang terpisah misalkan sebagai berikut:

$$P(A|B) = \frac{P(A)}{P(B)} P(B|A) \ldots\ldots\ldots \text{Persamaan (1)}$$

Simulasi Metode

Membuat simulasi perhitungan *Naive Bayes Classifier* dengan menggunakan tiga abstrak skripsi uji (D0 – D2) dan abstak skripsi penguji (Q). Berdasarkan pada tabel perhitungan selanjutnya sebagai berikut:

Tabel 1 Data abstrak uji dan penguji

id	?	A1	A2	A3	A4
Game	0	0	1	0	0
Media	0	0	0	0	0
Basis	0	1	1	0	0
Desktop	0	0	1	0	0
Teknologi	0	0	0	0	0
Sistem	1	1	0	0	1
Naive	0	1	0	0	0
Bayes	0	1	0	0	0
Web	0	0	0	0	0
Php	0	1	0	0	0
Mysql	0	1	0	0	0
Implementasi	0	0	0	0	0
Jaring	0	0	0	0	0
Firewall	0	0	0	0	0
Filtering	0	0	0	0	0
Komunikasi	0	0	0	0	0
Filter	0	0	0	0	0
Layer	0	0	0	0	0
Protokol	0	0	0	0	0
Mamdani	0	0	0	0	0
Fuzzy	0	0	0	0	0
Rule	0	0	0	0	0
Algoritma	0	0	0	1	0
Centroid	0	0	0	0	0
Neighbor	0	0	0	0	1
Suffix	0	0	0	0	0
Tree	0	0	0	0	0
Agglomerativ	0	0	0	0	0
Hierarchical	0	0	0	0	0
Cluster	0	0	0	1	0
Identifikasi	0	0	0	0	1
Deteksi	1	0	0	0	1
Nearest	0	0	0	0	0
KNN	0	0	0	0	1
Objek	0	0	0	0	0
Citra	0	0	0	0	1
RGB	0	0	0	0	1
Grayscale	1	0	0	0	1
Tepi	1	0	0	0	1
Histogram	0	0	0	0	1
Test	0	0	0	0	1
Trend	0	0	0	0	0
Moment	0	0	0	0	0
Least	0	0	0	0	0
Square	0	0	0	0	0
Single	0	0	0	0	0
Moving	0	0	0	0	0
Average	0	0	0	0	0
Ramal	0	0	0	0	0
Absolute	0	0	0	0	0
Error	0	0	0	0	0
MSE	0	0	0	0	0
Visual	0	0	0	0	0
Forecasting	0	0	0	0	0
Frame	0	0	0	0	0
Atribute	0	0	0	0	0
Decision	0	0	0	0	0
Rute	0	0	0	0	0
Apriori	0	0	0	0	0
Mining	0	0	0	0	0
Frequent	0	0	0	0	0
Perceptron	0	0	0	0	0
Bobot	0	0	0	0	0
Minkowski	1	0	0	0	0
Image	0	0	0	0	0
Topik	?	C	H	F	F

Keterangan:

A1 – A12 : Abstrak 1 – Abstrak 12

Topik (A) : Robotik

Topik (B) : Sistem Informasi

Topik (C) : SPK

Topik (D) : AI

Topik (E) : Citra

Topik (F) : Data Mining

Topik (G) : Jaringan

Topik (H) : Multimedia

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

$$P(Game = 0\ |Topik = SPK) = \frac{1}{1} = 1$$

$$P(Media = 0\ |Topik = SPK) = \frac{1}{1} = 1$$

$$P(Basis = 0\ |Topik = SPK) = \frac{1}{1} = 0.666$$

$$P(Desktop = 0\ |Topik = SPK) = \frac{1}{1} = 1$$

Dan seterusnya.

1) Menghitung hasil probalitas topik :

$$Robotika = \frac{1}{14} = 0.07142$$

$$Sistem\ Informasi = \frac{1}{14} = 0.07142$$

$$SPK = \frac{3}{14} = 0.21428$$

$$AI = \frac{1}{14} = 0.07142$$

$$Citra = \frac{2}{14} = 0.14285$$

$$Data\ Mining = \frac{3}{14} = 0.21428$$

$$Jaringan = \frac{1}{14} = 0.07142$$

$$Multimedia = \frac{2}{14} = 0.14285$$

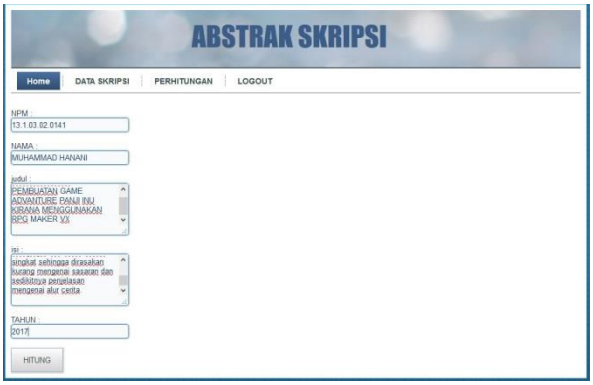
Tabel 2 Probabilitas

Topik	A	B	C	D	Dst
Game	1	1	1	1	Dst
Media	1	1	1	0	Dst
Basis	1	1	0.66	0	Dst
Desktop	1	1	1	1	Dst
Teknologi	1	1	1	0	Dst
Sistem	1	1	0.66	0	Dst
Naive	1	1	0.66	1	Dst
Bayes	1	1	0.66	1	Dst
Web	1	1	1	1	Dst
Php	1	1	0.66	1	Dst
Mysql	1	1	0.66	1	Dst
Implementasi	1	1	0.66	1	Dst
Jaring	1	1	1	1	Dst
Firewall	1	1	1	1	Dst
Filtering	1	1	1	1	Dst
Komunikasi	1	1	1	1	Dst
Filter	1	1	1	1	Dst
Layer	1	1	1	1	Dst
Protokol	1	1	1	1	Dst
Mamdani	1	1	0.66	1	Dst
Fuzzy	0	1	0.66	1	Dst
Rule	1	1	0.66	1	Dst
Algoritma	1	0	0.66	1	Dst
Centroid	1	0	1	1	Dst
Neighbor	1	0	1	1	Dst
Suffix	1	0	1	1	Dst
Tree	1	0	1	1	Dst
Agglomerativ	1	0	1	1	Dst
Hierarchical	1	1	1	1	Dst
Cluster	1	0	1	1	Dst
Identifikasi	1	1	0.66	1	Dst
Deteksi	0	0	0	0	Dst
Nearest	1	1	1	1	Dst
KNN	1	1	1	1	Dst
Objek	1	1	1	1	Dst
Citra	0	1	1	1	Dst
RGB	1	1	1	1	Dst

Grayscale	0	0	0	0	Dst
Tepi	0	0	0	0	Dst
Histogram	1	1	1	1	Dst
Test	1	1	1	1	Dst
Trend	1	1	0.66	1	Dst
Moment	1	1	0.66	1	Dst
Least	1	1	1	1	Dst
Square	1	1	0.66	1	Dst
Single	1	1	0.66	1	Dst
Moving	1	1	1	1	Dst
Average	1	1	0.66	1	Dst
Ramal	1	1	0.66	1	Dst
Absolute	1	1	1	1	Dst
Error	1	1	0.66	1	Dst
MSE	1	1	0.66	1	Dst
Visual	1	1	0.66	1	Dst
Forecasting	1	1	1	1	Dst
Frame	1	1	1	1	Dst
Atribute	1	1	1	1	Dst
Decision	1	1	1	1	Dst
Rute	1	1	1	1	Dst
Apriori	1	1	1	1	Dst
Mining	1	1	1	1	Dst
Frequent	1	1	1	1	Dst
Perceptron	1	1	1	1	Dst
Bobot	1	1	1	1	Dst
Minkowski	0	0	1	1	Dst
Image	1	1	1	1	Dst
Topik	0.07	0.07	0.21	0.07	Dst
Total	59.7	54.7	55.21	58.7	Dst

III. HASIL DAN KESIMPULAN

A. Tampilan Halaman Perhitungan



Gambar 2 Halaman Perhitungan

Setelah proses sudah berjalan, akan muncul hasil klasifikasi abstrak yang sudah diuji.

B. Tampilan Halaman Hasil

Gambar 3 Halaman Hasil

Pada halaman ini sudah proses hasil *Case folding*, *Tokenizing*, *Stopword Removal*, dan *Stemming*. Setelah pra-processing tersebut keluar hasil klasifikasi abstrak.

IV. PENUTUP

Simpulan

Berdasarkan hasil penelitian dan pembahasan tentang pengklasifikasian abstrak skripsi, dapat disimpulkan:

1. sistem bantu yang dibangun belum bekerja dengan maksimal. Di karenakan

ada kesalahan pada pembentukan pada kriteria dan perhitungan yang masih salah. Maka dari itu, berdasarkan pengujian yang telah dilakukan sistem menghasilkan nilai akurasi sebesar 16%.

2. Dengan adanya sistem bantu ini dapat mempermudah petugas akademik program studi Teknik Informatika dalam melakukan pengarsipan abstrak skripsi.
3. Dengan tingkat nilai akurasi yang baik, sistem dapat membantu petugas akademik dalam penanataan abstrak skripsi yang ada di program studi Teknik Informatika dengan mudah.

Saran

Setelah menarik dari beberapa kesimpulan maka saran yang bisa diberikan oleh penulis sebagai berikut:

1. Dari hasil penelitian maka perlu adanya pembenahan dari penerapan metode terhadap klasifikasi abstrak.

V. DAFTAR PUSTAKA

Aida Indriani. 2014. *Program Studi Teknik Informatika, STMIK PPKIA Tarakanita Rahmawati*. Seminar Nasional Aplikasi Teknologi Informasi (SNATI), *ISSN: 1907-5022*

Andini, S. 2013. Klasifikasi Dokumen Teks Menggunakan Algoritma *Naive Bayes* dengan Bahasa Pemrograman Java. Jurnal Teknologi Informasi & Pendidikan, 6(2): 140-147.

Dudung. 2016. Pengertian, Komponen dan Fungsi XAMPP Lengkap dengan Penjelasannya (online). Tersedia: https://www.dosenpendidikan.com/pengertian-komponen-dan-fungsi-xampp-lengkap-dengan-penjelasannya/, diakses 15 juli 2016.

Feldman, R & Sanger, J. 2007. The Text Mining Handbook : Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press : New York.

Fadillah Z. Tala. 2003. *A Study of Stemming Effect on Information Retrieval in Bahasa Indonesia*, Netherland, Universiteit van Amsterdam

Jogiyanto. 2005. Pengertiang Entity Relationship Diagram (ERD). (Online). Tersedia: http://www.teukutaufik.com/2016/02/definisi-dan-simbol-erd-entity.html, diakses 9 Juli 2016.

.

Kusrini, Luthfi, E.T. 2009. Algoritma Data Mining. Yogyakarta: ANDI.

Lembaga Penelitian dan Pengabdian Masyarakat (LPPM). 2015. *Panduan Penulisan Karya Tulis Ilmiah*. Kediri: Universitas Nusantara PGRI Kediri.

Natalius, S. 2010. Metoda *Naive Bayes Classifier* dan Penggunaannya pada Klasifikasi Dokumen. Makalah II2092 Probabilitas dan Statistik – Sem. 1 Tahun 2010/2011.

Nugroho, B. 2005. *Database Relasional dengan MySQL*. Yogyakarta: ANDI.

Ivan Rismi Polontalo. 2013. Penulisan Abstrak dalam sebuah karya tulis ilmiah. (Online) Tersedia : Http://dosen.ung.ac.id/ivanrismipolontalo/home/2013/1/24/penulisan_abstrak_dalam_sebuah_karya_tulis_ilmiah.html

Utomo, M.S. 2013. Implementasi Stemmer Tala Pada Aplikasi Berbasis Web. Jurnal Teknologi Informasi DINAMIK. 18 (1): 41-45.

Zuliarso, H. F. 2012. Rancang Bangun Sistem Perpustakaan untuk Jurnal Elektronik. *Fakultas Teknologi Informasi, Universitas Stikubank*.