

ANALISIS JAWABAN SOAL ESSAY MENGGUNAKAN TEXT MINING DENGAN METODE NAÏVE BAYES

SKRIPSI

Diajukan Untuk Memenuhi Sebagian Syarat Guna
Memperoleh Gelar Sarjana Komputer (S.Kom)
Pada Jurusan Teknik Informatika



Oleh :

FARIS YUDHA DHARMAWAN
NPM : 10.1.03.02.0188

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS NUSANTARA PGRI KEDIRI
2014**



Skripsi oleh:

FARIS YUDHA DHARMAWAN

NPM: 10.1.03.02.0188

Judul:

**ANALISIS JAWABAN SOAL ESSAY MENGGUNAKAN TEXT
MINING DENGAN METODE NAÏVE BAYES**

Telah disetujui untuk diajukan Kepada

Panitia Ujian/Sidang Skripsi Jurusan Teknik Informatika

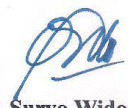
FT UNP Kediri

Tanggal: 09 Februari 2016

Pembimbing I


Drs. Agus Budianto, M.Pd
NIDN. 002 208 650 8

Pembimbing II


Dr. Suryo Widodo, M.Pd
NIDN. 000 202 640 3

Skripsi oleh:

FARIS YUDHA DHARMAWAN

NPM: 10.1.03.02.0188

Judul:

**ANALISIS JAWABAN SOAL ESSAY MENGGUNAKAN TEXT
MINING DENGAN METODE NAÏVE BAYES**

Telah dipertahankan di depan Panitia Ujian/Sidang Skripsi

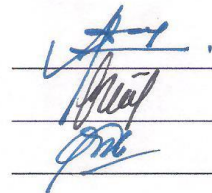
Jurusan Teknik Informatika UNP Kediri

Pada Tanggal: 09 Februari 2016

Dan Dinyatakan telah Memenuhi Persyaratan

Panitia Penguji:

1. Ketua : Drs. Agus Budianto, M.Pd.
2. Penguji I : Fatkur Rohman, M.Pd.
3. Penguji II : Dr. Suryo Widodo, M.Pd.



Mengetahui,

Dekan Fak. Teknik



Dr. Survo Widodo, M.Pd
NIP. 19640202 199103 1

ANALISIS JAWABAN SOAL ESSAY MENGGUNAKAN TEXT MINING DENGAN METODE NAÏVE BAYES

Nama Mahasiswa: Faris Yudha Dharmawan

NPM : 10.1.03.02.0188

Fakultas Teknik – Prodi Teknik Informatika

Email: farisyudharmawan@gmail.com

Dosen Pembimbing 1: Drs. Agus Budianto, M.Pd.

Dosen Pembimbing 2: Dr. Suryo Widodo, M.Pd.

UNIVERSITAS NUSANTARA PGRI KEDIRI

ABSTRAK

Penelitian ini diangkat berdasarkan pada permasalahan umum yang sering terjadi dalam kehidupan sehari-hari, dimana guru atau dosen masih mengoreksi/menganalisis jawaban soal essay dari banyaknya siswa atau mahasiswa secara manual yang menyita waktu dan tenaga. Hal tersebut kadang membuat analisis jawaban secara manual tidak akurat.

Permasalahan dalam penelitian ini adalah (1) Bagaimana merancang sistem pengambil keputusan untuk menganalisis jawaban soal essay secara otomatis dengan text mining? (2) Bagaimana menganalisis jawaban soal essay dengan membandingkan jawaban asli dengan jawaban dari siswa atau mahasiswa menggunakan Algoritma Naïve Bayes?

Dalam membandingkan jawaban essay aplikasi ini menggunakan pendekatan Text Mining yang selanjutnya diproses dengan algoritma Naïve Bayes. Naïve Bayes adalah sebuah kerangka untuk mengambil keputusan berdasarkan pengalaman dimasa lalu. Serta dalam perancangan sistem menggunakan Flowchart dan Use Case. Implementasi dan rancangan program menggunakan bahasa pemrograman PHP dengan database MySql.

Kesimpulan penelitian ini adalah (1) Dengan asumsi bahwa tujuan menganalisis jawaban essay dengan text mining menggunakan Algoritma Naïve Bayes dapat mempermudah memecahkan kasus mengoreksi jawaban soal essay. (2) Model penerapan logika dinamis yang diterapkan dalam Algoritma Naïve Bayes memiliki keuntungan dimana proses membandingkan setiap kata pada jawaban memberikan bentuk perbandingan yang sangat fleksibel.

Dengan adanya Analisis Jawaban Soal Essay dengan Text Mining menggunakan metode Naïve Bayes maka akan memudahkan guru atau dosen mengoreksi jawaban soal essay dari siswa atau mahasiswa untuk menentukan nilai dengan asumsi bahwa soal essay tersebut berjawaban mutlak.

Kata Kunci: Naïve Bayes, text mining, analisis, jawaban, soal, essay, koreksi.

I. LATAR BELAKANG

Saat ini perkembangan teknologi informasi semakin pesat, hal ini dapat dilihat pada tugas – tugas yang biasanya dilakukan secara manual dan menyita waktu sekarang dapat dilakukan secara otomatis dengan waktu yang singkat. Soal essay merupakan soal yang membutuhkan jawaban tertulis, jawaban soal essay dapat dibedakan menjadi dua, soal essay yang pertama yaitu soal essay dengan jawaban terbatas. Soal essay jenis ini sangat menekankan pada batas jumlah jawaban, soal essay dengan jawaban terbatas biasanya mengandung permasalahan yang terbatas dan jawabannya sudah ditentukan spesifikasinya. Soal essay yang kedua yaitu soal essay dengan jawaban bebas. Biasanya jawaban soal essay jenis ini merupakan sebuah pendapat/opini dari penjawab, dan jawaban dari setiap orang mungkin berbeda.

Mengoreksi jawaban dari soal essay secara manual adalah hal yang membutuhkan waktu lama, selain itu banyak kesulitan – kesulitan jika mengoreksi jawaban essay secara manual yaitu ketidak akuratan koreksi, membutuhkan banyak biaya, dan melelahkan. Untuk dapat mengoreksi jawaban dengan cepat dan akurat

dibutuhkan aplikasi analisis jawaban essay oleh karena itu dengan adanya aplikasi analisis jawaban essay dapat membantu guru maupun dosen untuk mengoreksi jawaban dari soal essay dengan sangat mudah dan cepat.

Pada pembuatan aplikasi analisis jawaban soal essay ini metode yang digunakan yaitu Naïve Bayes. Naïve Bayes merupakan metode yang digunakan untuk mengklasifikasikan sekumpulan dokumen. Algoritma ini memanfaatkan metode probabilitas dan statistik, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. Berdasarkan uraian di atas maka peneliti mengambil judul Analisis Jawaban Soal Essay dengan memanfaatkan metode Naïve Bayes, diharapkan dapat digunakan untuk menyelesaikan masalah koreksi jawaban soal essay.

II. METODE NAÏVE BAYES

Naïve Bayes Classifier merupakan sebuah metoda klasifikasi yang berakar pada teorema *Bayes*. Ciri utama dari *Naïve Bayes Classifier* ini adalah asumsi yang sangat kuat (naïf) akan independensi dari masing-masing kondisi/kejadian. Sebelum

menjelaskan *Naïve Bayes Classifier* ini, akan dijelaskan terlebih dahulu Teorema *Bayes* yang menjadi dasar dari metoda tersebut. Pada teorema *Bayes*, bila terdapat dua kejadian yang terpisah (misalkan A dan B), maka teorema *Bayes* dirumuskan sebagai berikut:

$$P(A|B) = \frac{P(A)}{P(B)} P(B|A)$$

Teorema *Bayes* sering pula dikembangkan mengingat berlakunya hukum probabilitas total, menjadi seperti berikut:

$$P(A|B) = \frac{P(A)P(B|A)}{\sum_{i=1}^n P(A_i|B)}$$

dimana $A_1 \cup A_2 \cup \dots \cup A_n = S$

Untuk menjelaskan teorema *Naïve Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, teorema *Bayes* di atas disesuaikan sebagai berikut:

$$P(C|F_1, \dots, F_n)$$

$$\frac{P(C)P(|F_1, \dots, F_n|C)}{P(F_1, \dots, F_n)}$$

Dimana variabel C merepresentasikan kelas, sementara variabel F_1, \dots, F_n merepresentasikan karakteristik-karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel dengan karakteristik tertentu dalam kelas C (*posterior*) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut *prior*), dikali dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas C (disebut juga *likelihood*), dibagi dengan peluang kemunculan karakteristik-karakteristik sampel secara global (disebut juga *evidence*).

Nilai *evidence* selalu tetap untuk setiap kelas pada satu sampel. Nilai dari Posterior tersebut yang nantinya akan dibandingkan dengan nilai-nilai Posterior kelas lainnya untuk menentukan ke kelas apa suatu

sampel akan diklasifikasikan. Penjabaran lebih lanjut rumus *Bayes* tersebut dilakukan dengan menjabarkan $P(F_1, \dots, F_n|C)$ menggunakan aturan perkalian, menjadi sebagai berikut:

$$\begin{aligned} P(F_1, \dots, F_n|C) &= \\ P(F_1|C)P(F_2, \dots, F_n|C, F_1) \\ &= P(F_1|C)P(F_2|C, F_1) \\ &\quad P(F_3, \dots, F_n|C, F_1, F_2) \\ &\quad \vdots \\ &\quad P(F_{n-1}|C, F_1, F_2, \dots, F_{n-2}) \\ &\quad P(F_n|C, F_1, F_2, \dots, F_{n-1}) \end{aligned}$$

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya faktor-faktor syarat yang mempengaruhi nilai probabilitas, yang hampir mustahil untuk dianalisa satu-persatu. Akibatnya, perhitungan tersebut menjadi sulit untuk dilakukan.

Di sinilah digunakan asumsi independensi yang sangat tinggi (naïf), bahwa masing-masing petunjuk (F_1, F_2, \dots, F_n) saling bebas (*independent*) satu sama lain.

Dengan asumsi tersebut, maka berlaku suatu kesamaan sebagai berikut:

$$P(F_i|F_j) = \frac{P(F_i \cap F_j)}{P(F_j)} = P(F_i)$$

untuk $i \neq j$, sehingga

$$P(F_i|C, F_j) = P(F_i|C)$$

Dari persamaan di atas dapat disimpulkan bahwa asumsi independensi naïf tersebut membuat syarat peluang menjadi sederhana, sehingga perhitungan menjadi mungkin untuk dilakukan. Selanjutnya, penjabaran $P(F_1, \dots, F_n|C)$ dapat disederhanakan menjadi seperti berikut:

$$\begin{aligned} P(F_1, \dots, F_n|C) &= P(F_1|C)P(F_2|C) \dots P(F_n|C) \\ P(F_1, \dots, F_n|C) &= \prod_{i=1}^n P(F_i|C) \end{aligned}$$

Dengan kesamaan di atas, persamaan teorema *Bayes* dapat dituliskan sebagai berikut:

$$\begin{aligned} P(F_1, \dots, F_n|C) &= \frac{1}{P(F_1, F_2, \dots, F_n)} P(C) \\ &\quad \prod_{i=1}^n P(F_i|C) \end{aligned}$$

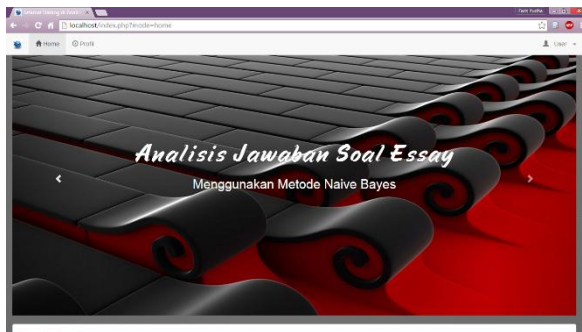
$$P(C|F_1, \dots, F_n) \\ = \frac{P(C)}{Z} \prod_{i=1}^n P(F_i|C)$$

Persamaan di atas merupakan model dari teorema *Naïve Bayes* yang selanjutnya akan digunakan dalam proses klasifikasi dokumen. Adapun Z merepresentasikan *evidence* yang nilainya konstan untuk semua kelas pada satu sampel.

III. HASIL DAN KESIMPULAN

A. Tampilan Aplikasi

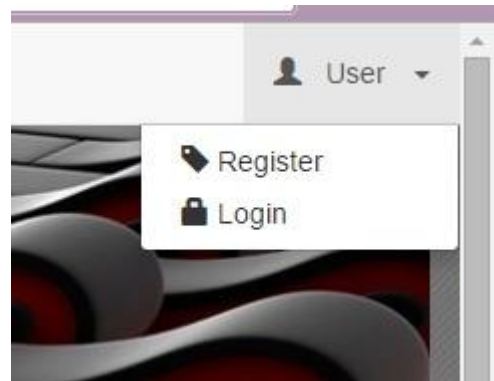
a. Tampilan *Home Page*



Gambar 5.1 Tampilan *Home Page*

Home Page merupakan halaman awal aplikasi analisis jawaban soal essay, berisi menu *Home*, *Profil*, dan *User* untuk *Login* atau Registrasi mahasiswa baru.

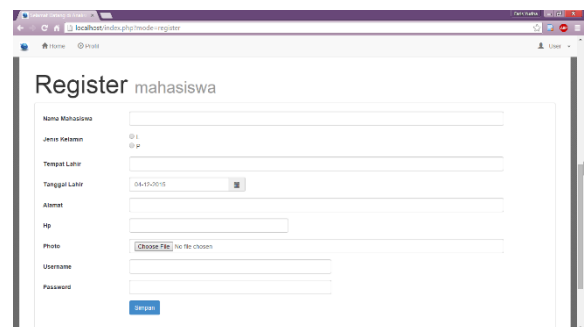
b. Tampilan Menu *User*



Gambar 5.2 Tampilan Menu *User*

Menu *User* terdapat pada pojok kanan atas menu bar, menu ini digunakan untuk *register* mahasiswa baru dan juga *login* sebagai mahasiswa atau *admin*.

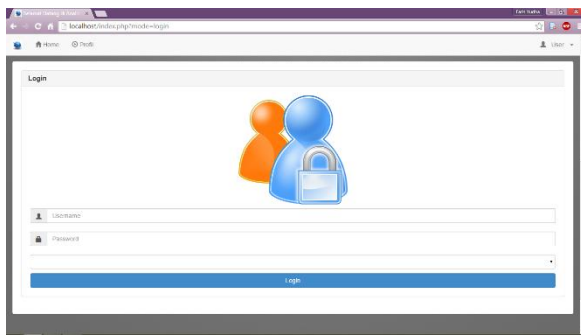
c. Tampilan Register *Page*



Gambar 5.3 Tampilan Register *Page*

Halaman ini digunakan untuk melakukan registrasi mahasiswa baru dengan mengisi identitas mahasiswa, *username* dan *password*.

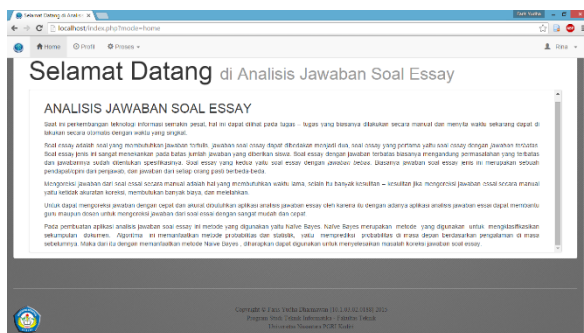
d. Tampilan Login Page



Gambar 5.4 Tampilan Login Page

Tampilan menu *login*, *user* dapat *login* sebagai mahasiswa atau sebagai *admin*, untuk melakukan proses analisis maka *user* harus *login* sebagai mahasiswa terlebih dahulu, untuk melihat, mengganti, dan *mem-backup* data maka *user* harus *login* sebagai *admin*.

e. Tampilan Home Page User Mahasiswa

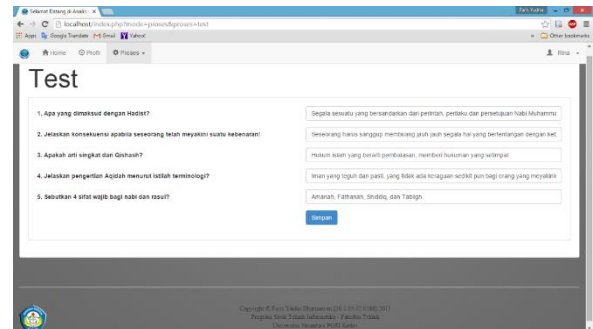


Gambar 5.5 Tampilan Home Page user Mahasiswa

Tampilan awal setelah *login* sebagai mahasiswa, terdapat menu tambahan yaitu menu Proses yang digunakan untuk melakukan *test*

dengan menjawab soal yang telah disediakan.

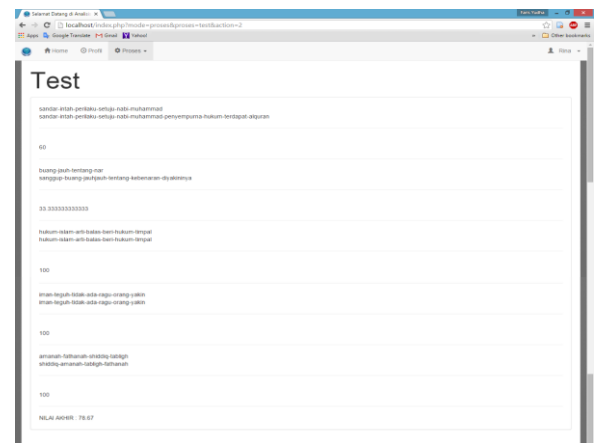
f. Tampilan Test Page



Gambar 5.6 Tampilan Test Page

Tampilan halaman *test*, mahasiswa harus mengisi jawaban pada setiap pertanyaan di *field* yang sudah disediakan, klik *save* untuk diproses analisis oleh program dan melihat nilainya.

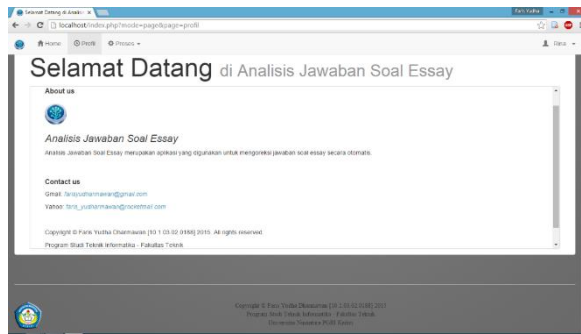
g. Tampilan Test Result Page



Gambar 5.6 Tampilan Test Result Page

Halaman ini menampilkan hasil *text mining* jawaban, nilai pada setiap soal dan nilai total dari proses analisis jawaban menggunakan algoritma *Naïve Bayes*.

h. Tampilan *Profil Page*



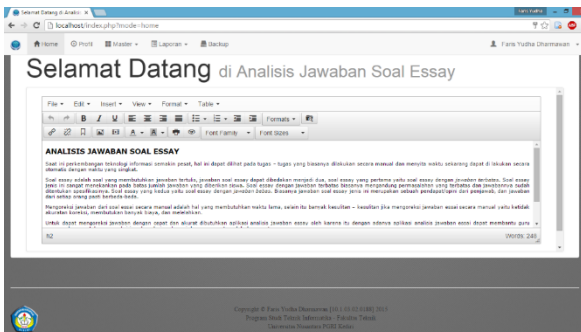
Gambar 5.7 Tampilan *Profil Page*

Merupakan halaman yang berisi informasi tentang aplikasi Analisis Jawaban Soal Essay dan *developer*.

i. *Administrator Page*

Merupakan tampilan *user administrator*, berisi menu *Home*, *Profil*, *Master*, *Laporan*, dan *Backup*.

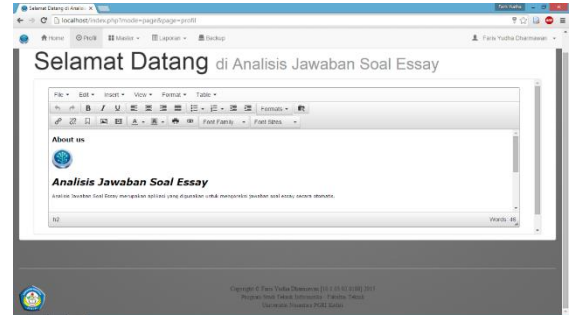
a. Tampilan *Edit Home Page*



Gambar 5.8 Tampilan *Edit Home Page*

Merupakan tampilan sebagai *administrator* yang digunakan untuk mengubah halaman awal atau *Home Page*.

b. Tampilan *Edit Profil Page*

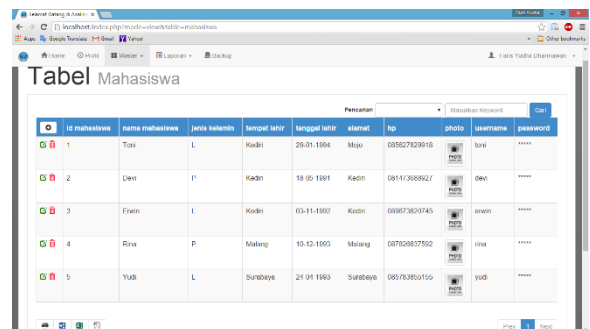


Gambar 5.9 Tampilan *Edit Profil Page*

Merupakan tampilan halaman untuk mengubah *Profil Page*

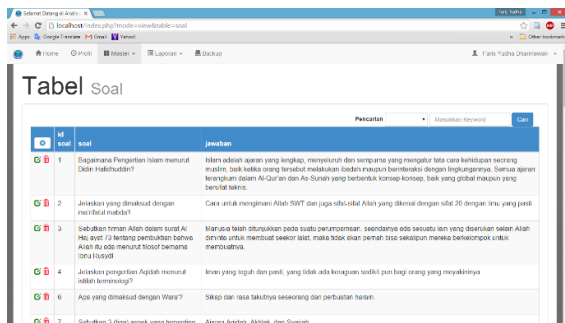
c. Tampilan Halaman pada menu *Master*

Merupakan menu yang digunakan untuk mengubah *database*.



Gambar 5.10 Tampilan *Edit User Page*

Merupakan tampilan halaman untuk mengubah *user mahasiswa*.

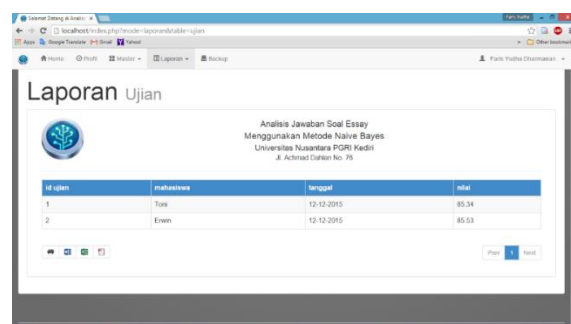


id soal	soal	jawaban
1	Bagaimana Pengertian Islam menurut Quran?	Islam adalah agama yang lengkap, menyeluruh dan sempurna yang mengatur tata cara kehidupan seorang muslim, baik dalam urusan ibadah maupun kehidupan sosial kemasyarakatan. Semua ajaran yang terkandung dalam Al-Quran dan As-Sunnah yang berbentuk konsep-konsep, baik yang global maupun yang bersifat teknis.
2	Jelaskan yang dimaksud dengan meyakini Allah?	Cara untuk mengimani Allah SWT dan juga untuk Allah yang dikawatirkan oleh 20 dengan ilmu yang pasti meyakini Allah?
3	Sebutkan nama Allah dalam surat Al-Fatiha!	Marilah kita tundukkan pada suatu persembahkan, seandainya ada sesuatu yang diserahkan oleh Allah SWT untuk membuat seorang muslim, maka tidak akan pernah bisa sekalipun mereka berakhlak untuk memahaminya.
4	Jelaskan pengertian Allah menurut Islam!	Ilmu yang tahu dan pasti, yang tidak ada keraguan sedikit pun bagi orang yang meyakini.
5	Apakah yang dimaksud dengan Wala?	Sikap dan rasa takutnya seseorang dari perbuatan haram.
6	Apakah yang dimaksud dengan Wala?	Sikap dan rasa takutnya seseorang dari perbuatan haram.
7	Sebutkan 3 macam dosa besar!	Al-Kufr, Al-Nifaq, dan Al-Munafiq.

Gambar 5.11 Tampilan Edit Dataset Page

Merupakan tampilan halaman yang digunakan untuk mengubah *dataset* soal dan jawaban.

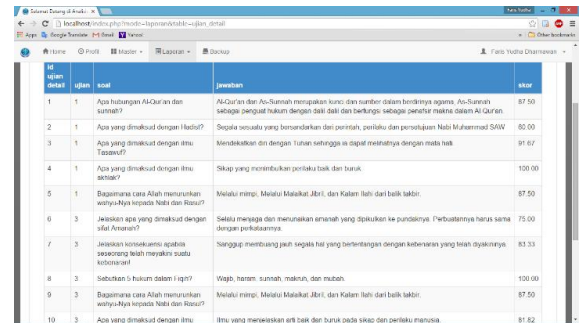
d. Tampilan Halaman pada menu Laporan



id ujian	mahasiswa	tanggal	nilai
1	Toni	12-12-2015	80.34
2	Eren	12-12-2015	80.53

Gambar 5.12 Tampilan Laporan Ujian Page

Merupakan tampilan halaman yang digunakan untuk melihat peserta ujian, tanggal ujian, dan nilai hasil ujian mahasiswa.



id soal	ujian	soal	jawaban	nilai
1	1	Apakah hubungan Al-Quran dan As-Sunnah?	Al-Quran dan As-Sunnah merupakan kunci dan sumber dalam beragama, Al-Quran sebagai petunjuk hukum dengan dalil dari Allah dan As-Sunnah sebagai petunjuk praktik dalam Al-Quran.	87.50
2	1	Apakah yang dimaksud dengan Ibadah?	Segala sesuatu yang bersesuaian dengan perintah, larangan dan petunjuk Nabi Muhammad SAW.	80.00
3	1	Apakah yang dimaksud dengan Ilmu?	Mengetahui dan dengan Tuhan sehingga ia dapat menyalurkan dengan malaikat.	91.67
4	1	Apakah yang dimaksud dengan Ilmu?	Sikap yang menimbulkan perilaku baik dan buruk.	100.00
5	1	Bagaimana cara Allah memurnikan Al-Quran?	Melalui mimpi, Melalui Malaikat Jibril, dan Kalimat lahi dari baik laki-laki.	87.50
6	3	Jelaskan apa yang dimaksud dengan Al-Fatiha?	Selalu mengawali dengan Al-Fatiha yang diucapkan ke pundaknya. (Fatiha) artinya harus sama dengan perkataannya.	75.00
7	3	Jelaskan konsekuensi apabila seseorang telah meyakini Allah?	Sangat membimbing jalan segala hal yang bertentangan dengan kebenaran yang telah dijanjikan.	83.33
8	3	Sebutkan 3 hukum dalam Islam?	Wajib, haram, sunnah, makruh, dan mubah.	100.00
9	3	Bagaimana cara Allah memurnikan Al-Quran?	Melalui mimpi, Melalui Malaikat Jibril, dan Kalimat lahi dari baik laki-laki.	87.50
10	3	Apakah yang dimaksud dengan Ilmu?	Ilmu yang memberikan arah baik dan buruk pada sikap dan perilaku manusia.	81.52

Gambar 5.13 Tampilan Detail Ujian Page

Merupakan tampilan halaman yang digunakan untuk melihat detail ujian mahasiswa dan skor pada tiap soal.

B. Kesimpulan

1. Dengan asumsi bahwa tujuan menganalisis jawaban essay dengan *Text Mining* menggunakan Algoritma *Naïve Bayes* dapat mempermudah memecahkan kasus mengoreksi jawaban soal essay.
2. Model penerapan logika dinamis yang diterapkan dalam *Text Mining* dan Algoritma *Naïve Bayes* memiliki keuntungan dimana proses membandingkan setiap kata pada jawaban memberikan bentuk perbandingan yang sangat fleksibel.
3. Data yang digunakan dapat diganti atau ditambah sesuai kebutuhan.

IV. DAFTAR PUSTAKA

- Agus Mulyanto. 2009. *Sistem Informasi Konsep dan Aplikasi*. Pustaka Pelajar. Yogyakarta.
- Agusta, Ledy. 2009. *Perbandingan Algoritma Stemming Porter Dengan Algoritma Nazief & Adriani Untuk Stemming Dokumen Teks Bahasa Indonesia*. Indonesia: Universitas Kristen Satya Wacana.
- Andini Silfia, *Klasifikasi Dokument Teks Menggunakan Algoritma Naive Bayes Dengan Bahasa Pemrograman Java*. (Online), tersedia: <http://jurnal-tip.net/jurnal-resource/file/13-Vol6-No2Sep2013-Silfia%20Andini.pdf>, diunduh 21 Januari 2014.
- Anugroho Prasetyo, *Klasifikasi Email Spam dengan Metode Naïve Bayes Classifier menggunakan Java Programing*. (Online), tersedia: <http://www2.eepis-its.edu/id/ta/975/Klasifikasi-Spam-Email-Dengan-Metode-Naive-Bayes-Classifier>, diunduh 4 Juni 2014.
- Dzacko Haidar, *Basis Data (Database)*. (Online), tersedia: http://imam_muiz.staff.gunadarma.ac.id/Downloads/files/6535/BASIS+DATA.pdf, diunduh 6 Juni 2014.
- Larose D, T., 2006, *Data Mining Methods and Models*, Jhon Wiley & Sons, Inc. Hoboken New Jersey.
- Natalius Samuel, *Metoda Naïve Bayes Classifier dan Penggunaannya pada Klasifikasi Dokumen*. (Online), tersedia: <http://informatika.stei.itb.ac.id/~rinaldi.munir/Probst/2010-2011/Makalah2010/MakalahProbst/2010-017.pdf>, diunduh 20 Januari 2014.
- Nugroho, Adi. *Rational Rose untuk Pemodelan Berorientasi Objek*. Informatika. Bandung, 2005.
- Pierre Bourque and Robert Dupuis, ed. 2004. *Guide to the Software Engineering Body of Knowledge - 2004 Version*. IEEE Computer Society. pp. 1-1.
- Pramudiono, I. 2006. Apa itu Data Mining?. (Online), tersedia: <http://datamining.japati.net/cgi-bin/indodm.cgi>. Diakses tanggal 21 Januari 2014.
- Rosa A.S. dan M. Shalahuddin. 2013. *Rekayasa Perangkat Lunak*, Bandung: Medula, hlm 5.
- Sukriyawati Gusnia, *Klasifikasi Otomatis Dokumen Berita Kejadian Berbahasa Indonesia Menggunakan Metode Naive Bayes*. (Online), tersedia <http://lib.uin-malang.ac.id/files/thesis/fullchapter/05550057.pdf>, diunduh 4 Juni 2014.
- Turban, Efraim et al. 2005. *Introduction to Information Technology*, 3rd Edition. John Wiley & Sons, Inc. New York. USA.