

Analisis Sentimen Data Kritik Dan Saran Pelatihan Aplikasi Teknologi Informasi Menggunakan Algoritma Support Vector Machine

Alimuddin Hasan Al Kabir^{*1}, Setio Basuki², Galih Wasis Wicaksono³

^{1,2,3}Teknik Informatika/Universitas Muhammadiyah Malang

alimudhasan@gmail.com^{*1}, setiobasuki.umm@gmail.com², galih.w.w@gmail.com³

Abstrak

Opini publik merupakan salah satu instrumen yang bisa digunakan untuk melakukan evaluasi terhadap suatu kegiatan. Penelitian ini dilandasi oleh beberapa masalah diantaranya (1) perlunya peningkatan kualitas pelaksanaan kegiatan Pelatihan Aplikasi Teknologi Informasi, (2) opini peserta yang sudah ditampung belum dimanfaatkan secara maksimal karena terkendala banyaknya data dan tidak mungkin di analisis secara manual. Kritik dan saran diambil dari data periode pelaksanaan 2016/2017 sebanyak 1050 data. Support Vector Machine digunakan sebagai metode dalam analisis sentimen. Proses latih data akan menghasilkan hyperplane terbaik yang dijadikan acuan untuk menentukan kelas sentimen mana sesuai untuk suatu kalimat. Pengujian dilakukan dengan membagi dataset ke dalam data uji sebanyak 20% dan data latih sebanyak 80% sehingga bisa dilakukan proses analisis hingga 5 kali iterasi dengan susunan data yang berbeda. Hasil pengujian menunjukkan perhitungan Akurasi, Precision, Recall, dan F-Measure yang dihasilkan oleh sistem adalah sebesar 82,08%, 83,42%, 81,16%, dan 81,82%.

Kata Kunci: Analisis Sentimen, Support Vector Machine, PATI, Opini

Abstract

Public opinion is one of many instrument that can be used to evaluate of an event. This research based on several problems, they are (1) quality improvement are necessary for Pelatihan Aplikasi Teknologi Informasi, (2) too much of data make the participant's opinion that has been collected has not been maximally utilized. Criticism and suggestions datas are taken from 2016/2017th school year in amount of 1050. Support Vector Machine is used as a method in sentiment analysis. The data training process will produce the best hyperplane used as a reference to determine which sentiment class is much appropriate for a sentence. The test is done by dividing the dataset into the test data as much as 20% and the training data as much as 80% so it can be done the analysis process up to 5 times iteration with different data arrangement. The test results show the calculation of Accuracy, Precision, Recall, and F-Measure generated by system is equal to 82,08%, 83,42%, 81,16%, and 81,82%.

Keywords: Analisis Sentimen, Support Vector Machine, PATI, Opinion

1. Pendahuluan

Pelatihan Aplikasi Teknologi Informasi atau biasa disingkat PATI dan lebih dikenal dengan istilah Pelatihan APLINET (Pelatihan Aplikasi Internet) adalah suatu kegiatan pelatihan yang diwajibkan bagi mahasiswa baru di Universitas Muhammadiyah Malang (UMM) [1]. Kegiatan ini bertujuan untuk memberikan pengetahuan dan pelatihan tentang segala hal yang berhubungan dengan teknologi informasi, khususnya teknologi informasi yang ada di lingkungan kampus. Didampingi oleh 2 (dua) instruktur yang sudah diseleksi dari kalangan mahasiswa UMM, kegiatan ini dilaksanakan selama seminggu dengan sistem langsung praktik. Proses pembelajaran dilakukan di 8 (delapan) laboratorium komputer di lingkungan kampus.

Berdasarkan data pada laporan akhir periode 2014 sampai 2016, kegiatan ini diikuti oleh peserta dengan kisaran 3000 hingga 4000 mahasiswa setiap periode pelaksanaannya. Dengan banyaknya jumlah peserta, tentunya diperlukan manajemen yang baik dari DPP Infokom selaku pelaksana kegiatan. Oleh sebab itu, pada setiap periode pelaksanaan pihak pelaksana selalu meningkatkan kualitas baik itu dalam segi materi, SDM (Sumber Daya Manusia), maupun infrastruktur berdasarkan evaluasi dari pelaksanaan pada periode sebelumnya.

Melakukan evaluasi program pembelajaran tidak cukup jika hanya didasarkan pada penilaian hasil belajar. Ada faktor-faktor lain yang juga mempengaruhinya seperti: siswa, pengelola, lingkungan, kualitas pembelajaran, kurikulum, dan sebagainya [2]. Peningkatan kualitas pembelajaran dan kualitas sistem penilaian diperlukan agar kualitas pendidikan bisa meningkat [3]. Dengan demikian untuk mencapai target pembelajaran diperlukan proses pembelajaran yang baik, sedangkan untuk efektivitas pembelajaran dipengaruhi oleh evaluasi terhadap proses juga hasil pembelajaran. Evaluasi yang tepat dapat mendorong pihak yang terlibat untuk lebih meningkatkan kualitas proses pembelajaran, layanan, maupun fasilitas.

Opini publik merupakan salah satu alat evaluasi yang sudah umum digunakan. Benjamin I. Page mengemukakan bahwa opini atau pendapat publik memiliki pengaruh terhadap pengambilan keputusan untuk suatu kebijakan [4]. Pelaksana kegiatan PATI pun sudah menempuh langkah untuk mengumpulkan opini dari para peserta untuk mengevaluasi kegiatan tersebut. Opini ini berupa kritik dan saran yang ditulis melalui form online di web *e-learning* PATI. Namun jumlah opini yang berupa teks bebas ini terlalu banyak, sehingga data ini tidak bisa dimanfaatkan dengan maksimal karena bentuknya yang tidak terstruktur dan bahasa yang digunakan terkadang tidak baku.

Sentiment analysis atau *opinion mining* merupakan studi yang bertujuan untuk mengolah data tekstual untuk mendapatkan informasi yang terkandung dalam suatu kalimat opini [5]. Pada dasarnya *sentiment analysis* mengelompokkan teks yang ada dalam sebuah kalimat atau dokumen untuk menentukan pendapat yang dikemukakan apakah cenderung positif, negatif, atau netral [6]. Untuk melakukan analisis seperti ini tentunya diperlukan algoritma yang sesuai agar hasilnya sesuai dengan yang diinginkan. Berdasarkan komparasi algoritma klasifikasi oleh Chandani dan Purwanto (2015), diperoleh kesimpulan bahwa *Support Vector Machine (SVM)* memiliki hasil terbaik dengan nilai *accuracy* = 81.10 % dan nilai *AUC* = 0.904 [7]. Hal yang senada juga dipaparkan oleh Hadna, Santosa, dan Winarno (2016) yang mengungkapkan bahwa *Support Vector Machine (SVM)* memiliki hasil yang lebih akurat jika dibandingkan dengan algoritma *Naïve Bayes Classifier (NBC)*.

Amir Hamzah (2014) telah melakukan penelitian mengenai *sentiment analysis* pada opini evaluasi pembelajaran *Naïve Bayes Classifier (NBC)* [8]. Tujuannya adalah untuk menerapkan teknik *sentiment analysis* untuk menganalisa data-data saran/opini mahasiswa, sehingga bisa dimanfaatkan untuk mendukung hasil evaluasi kuesioner yang sudah dilakukan sebelumnya. Hasil dari penelitian ini adalah aplikasi dengan 3 (tiga) fitur utama yaitu analisis opini, pencarian topik, dan *input data training*. Selain tingkat akurasi algoritma yang digunakan masih ada yang lebih baik, aplikasi ini terbilang pun minim fitur. Fitur *input data test*, analisa secara massal, maupun visualisasi data masih belum terdapat dalam aplikasi ini. Merujuk pada latar belakang yang telah dijabarkan, pada penelitian ini akan dikembangkan “Analisis Sentimen Kritik dan Saran Pelatihan Aplikasi Teknologi Informasi (PATI) Menggunakan Algoritma *Support Vector Machine (SVM)*”.

2. Metode Penelitian

2.1 Data Penelitian

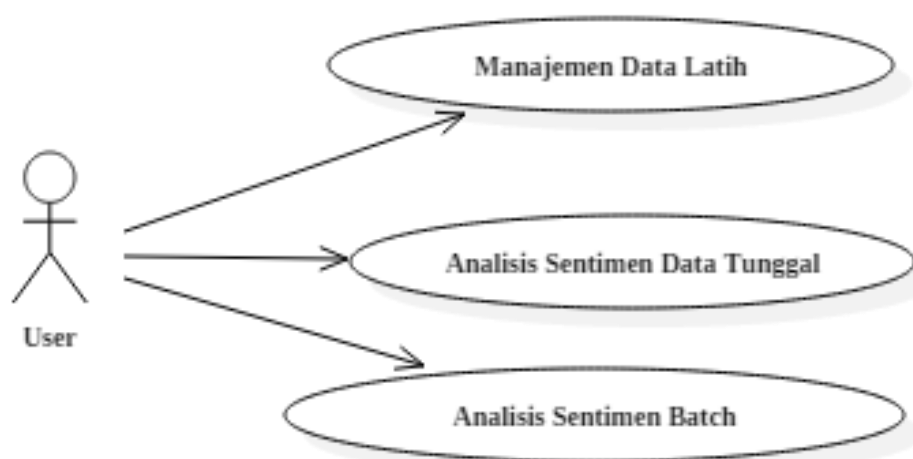
Data yang digunakan dalam penelitian ini adalah data kritik dan saran kegiatan Pelatihan Aplikasi Teknologi dan Informasi (PATI) Universitas Muhammadiyah Malang tahun akademik 2016/2017. Keseluruhan data tersebut berbentuk file *HTML (HyperText Markup Language)* dengan total 4.485 data

2.2 Proses Analisis Data

Data yang sudah diperoleh selanjutnya diolah untuk disesuaikan dengan kebutuhan. Sesuai dengan batasan yang sudah ditentukan sebelumnya, akan diambil sebanyak 1.000 data untuk kemudian dibagi menjadi data latih dan data uji. Komposisi data latih sendiri adalah sebagai berikut: 300 data sentimen positif, 300 data sentimen negatif, dan 300 data sentimen netral. Sementara sisa 100 data akan dijadikan data uji. Data yang sudah dikumpulkan nantinya akan diproses hingga bisa dihitung dan ditemukan persamaan hyperplane terbaiknya.

2.3 Usecase Diagram

Sistem pada Gambar 1 ini memiliki aktor tunggal yang memiliki tiga *usecase*. Pertama manajemen data latih, kedua melakukan analisis sentimen terhadap data tunggal, dan ketiga melakukan analisis sentimen terhadap banyak data sekaligus.



Gambar 1. Usecase Diagram System

2.4 Data Preprocessing

Data preprocessing dilakukan terhadap seluruh data yang ada, baik itu data latih maupun data uji. Ada beberapa tahapan *data preprocessing* yang akan dilakukan terhadap data. Proses ini akan menjadikan file *HTML* berisi kritik dan saran menjadi data siap uji.

Hal pertama yang perlu dilakukan adalah *import data* untuk mengambil konten yang ada dalam file-file tersebut. Kemudian dilanjutkan dengan membersihkan *html tags* yang masih melekat dalam konten file tersebut. Setelah itu tiap dokumen kritik dan saran akan dipecah kembali ke dalam bentuk kalimat karena sistem akan melakukan analisis terhadap dokumen di setiap kalimatnya.

Proses kemudian berlanjut dan dilakukan standarisasi konten. Artinya dokumen dibersihkan dari karakter-karakter yang tidak dibutuhkan dan akan mempengaruhi hasil analisis seperti angka, spasi berlebih, hingga kata tidak baku. Selain itu secara keseluruhan dokumen diubah ke dalam bentuk huruf kecil agar tidak terjadi kesalahan persepsi ketika sistem melakukan analisis.

2.5 Pembobotan TF-IDF

Term Frequency Inverse Document Frequency atau *TF-IDF* merupakan suatu metode yang digunakan dalam melakukan pembobotan terhadap kemunculan kata dalam suatu dokumen. *TF* menyatakan jumlah kata yang muncul dalam suatu dokumen. Sedangkan *IDF* menunjukkan tingkat kepentingan suatu kata yang terdapat dalam kumpulan dokumen. Pada *TF-IDF* terdapat rumus untuk menghitung bobot (W) masing-masing dokumen terhadap kunci yang diformulasikan pada Persamaan 1 dan Persamaan 2.

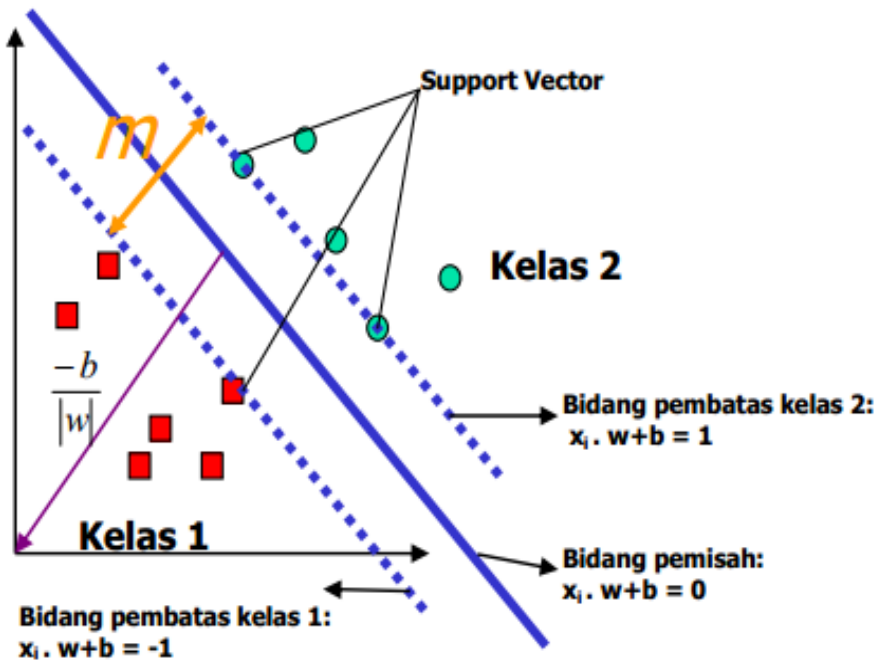
$$w_{ij} = tf \times idf \quad (1)$$

$$w_{ij} = tf_{ij} \times \log \frac{N}{n} \quad (2)$$

2.6 Support Vector Machine

Support Vector Machine (SVM) merupakan salah satu teknik untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi [9]. Teknik ini termasuk dalam metode klasifikasi jenis terpandu (*supervised*) karena memiliki target pembelajaran tertentu. Klasifikasi dilakukan dengan mencari *hyperplane* atau garis pembatas (*decision boundary*) yang memisahkan antara satu kelas dengan kelas lainnya.

Dalam konsep ini, *SVM* berusaha untuk mencari *hyperplane* terbaik diantara fungsi yang tidak terbatas jumlahnya. Fungsi yang tidak terbatas dalam pencarian *hyperplane* di metode *Support Vector Machine* merupakan sebuah keuntungan, dimana pemrosesan pasti akan selalu bisa dilakukan bagaimanapun data yang dimilikinya, seperti pada Gambar 2.



Gambar 2. Ilustrasi Model SVM [10]

Data yang digunakan pada metode Support Vector Machine dinotasikan dengan $x_i \in R^d$ dan label dinotasikan dengan $y_i \in \{1,2\}$ untuk $i = 1,2, \dots, l$ yang mana l adalah banyaknya data. Support Vector Machine dapat diformulasikan ke dalam Persamaan 3 dan Persamaan 4 untuk $y_i = +1$ dan persamaan 2.3 untuk $y_i = +2$.

$$x_i * w + b \geq +1 \quad (3)$$

$$x_i * w + b \leq -1 \quad (4)$$

w adalah normal bidang dan b adalah posisi bidang relatif terhadap pusat koordinat. Secara matematika, formulasi problem optimasi SVM untuk kasus klasifikasi linier di dalam primal space ditunjukkan pada Persamaan 5 dan Persamaan 6.

$$\min \frac{1}{2} \|w\|^2 \quad (5)$$

$$y_i(w x_i + b) \geq 1, i = 1, \dots, n \quad (6)$$

Saat pertama kali diperkenalkan oleh Vapnik, SVM hanya dapat mengklasifikasikan data ke dalam dua kelas (klasifikasi biner). Namun, penelitian lebih lanjut untuk mengembangkan SVM sehingga bisa mengklasifikasi data yang memiliki lebih dari dua kelas, terus dilakukan. Salah satu metode yang mengimplementasikan *multi class* SVM adalah *one-against-all*. Dengan menggunakan metode ini, dibangun k buah model SVM biner (k adalah jumlah kelas). Setiap model klasifikasi ke- i dilatih dengan menggunakan keseluruhan data, untuk mencari solusi permasalahan.

3. Hasil Penelitian dan Pembahasan

Pengujian sistem dilakukan dalam beberapa iterasi untuk memastikan keberhasilan pengujian. Tiap iterasinya dataset akan dibagi secara acak kedalam data uji dan data latih. Namun untuk prosentase pembagiannya sendiri sudah ditetapkan yaitu 20% data uji dan 80% data latih. Berikut ini hasil pengujian setiap iterasinya.

Tabel 1. Hasil Pengujian Iterasi 1 Menggunakan Confusion Matrix

		Prediksi		
		Positif	Negatif	Netral
Aktual	Positif	74	3	9
	Negatif	11	57	0
	Netral	18	0	39

Dari perhitungan yang sudah dilakukan Tabel 1, secara keseluruhan pada **Iterasi ke 1** menghasilkan **Akurasi, Precision Average, Recall Average, dan F-Measure Average** sebesar **80,75%, 82,70%, 79,43%, dan 80,55%**.

Tabel 2. Hasil Pengujian Iterasi 2 Menggunakan Confusion Matrix

		Prediksi		
		Positif	Negatif	Netral
Aktual	Positif	67	1	18
	Negatif	3	58	7
	Netral	11	1	45

Dari perhitungan yang sudah dilakukan Tabel 2, secara keseluruhan pada **Iterasi ke 2** menghasilkan **Akurasi, Precision Average, Recall Average, dan F-Measure Average** sebesar **80,57%, 81,22%, 80,72%, dan 80,58%**.

Tabel 3. Hasil Pengujian Iterasi 3 Menggunakan Confusion Matrix

		Prediksi		
		Positif	Negatif	Netral
Aktual	Positif	79	1	6
	Negatif	5	56	7
	Netral	10	1	46

Dari perhitungan yang sudah dilakukan Tabel 3, secara keseluruhan pada **Iterasi ke 3** menghasilkan **Akurasi, Precision Average, Recall Average, dan F-Measure Average** sebesar **85,78%, 86,19%, 84,97%, dan 85,33%**.

Tabel 4. Hasil Pengujian Iterasi 4 Menggunakan Confusion Matrix

		Prediksi		
		Positif	Negatif	Netral
Aktual	Positif	77	0	9
	Negatif	10	56	2
	Netral	13	1	43

Dari perhitungan yang sudah dilakukan Tabel 4, secara keseluruhan pada **Iterasi ke 4** menghasilkan **Akurasi, Precision Average, Recall Average, dan F-Measure Average sebesar 83,41%, 84,96%, 82,44%, dan 83,29%**.

Tabel 5. Hasil Pengujian Iterasi 5 Menggunakan Confusion Matrix

		Prediksi		
		Positif	Negatif	Netral
Aktual	Positif	79	1	6
	Negatif	5	56	7
	Netral	10	1	46

Dari perhitungan yang sudah dilakukan Tabel 5, secara keseluruhan pada **Iterasi ke 4** menghasilkan **Akurasi, Precision Average, Recall Average, dan F-Measure Average sebesar 80,09%, 82,04%, 78,25%, dan 79,36%**.

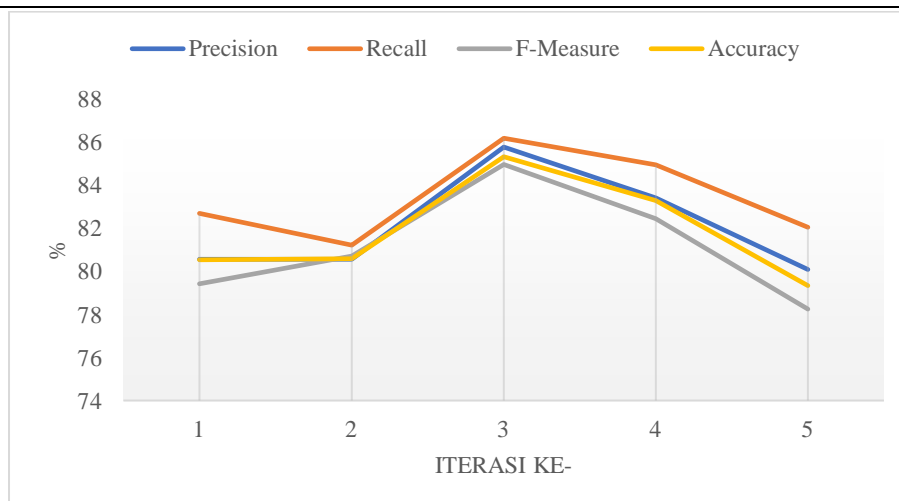
Berdasarkan seluruh pengujian yang sudah dilakukan, maka diperoleh hasil yang ditunjukkan pada Tabel 6.

Tabel 6. Hasil Pengujian Sistem

Iterasi ke-	Accuracy	Precision	Recall	F-Measure
1	80,57	82,70	79,43	80,55
2	80,57	81,22	80,72	80,58
3	85,78	86,19	84,97	85,33
4	83,41	84,96	82,44	83,29
5	80,09	82,06	78,25	79,34
Average	82,08	83,42	81,16	81,82

Dari data tersebut dapat digambarkan sebuah grafik yang menunjukkan hasil penelitian di setiap iterasinya.

Berdasarkan hasil pengujian yang sudah dilakukan pada Gambar 3, dapat didapatkan rata-rata **Akurasi, Precision, Recall, dan F-Measure sebesar 82,08%, 83,42%, 81,16%, dan 81,82%**.



Gambar 3. Grafik Hasil Penelitian

4. Kesimpulan

Mempersiapkan data kritik dan sara yang masih mentah dilakukan dalam beberapa tahap diantaranya import data, konversi data, dan dan menyimpan data ke dalam database. Agar bisa mengimplementasikan algoritma Support Vector Machine perlu dilakukan preprocessing terhadap data yang sudah didapat agar seluruh fitur yang dibutuhkan bisa dihitung bobotnya. Mengukur keberhasilan metode bisa dilakukan dengan menggunakan confusion matrix yang nantinya menghasilkan Akurasi, Precision, Recall, dan F-Measure. Akurasi, Precision, Recall, dan F-Measure yang dihasilkan oleh sistem adalah sebesar 82,08%, 83,42%, 81,16%, dan 81,82%. Ini merupakan hasil yang baik sehingga kinerja sistem untuk analisis sentimen sudah optimal.

Referensi

- [1] Panitia PATI UMM, "Pelatihan Aplikasi Teknologi Informasi (PATI) Universitas Muhammadiyah Malang," 2013. [Online]. Available: <https://ppaiumm.wordpress.com/>. [Accessed: 06-Sep-2017].
- [2] S. E. P. Widoyoko, "Evaluasi Program Pembelajaran," *J. Ilmu Pendidik.*, pp. 1–16, 2000.
- [3] D. Mardapi, "Pengukuran Penilaian dan Evaluasi Pendidikan," *Yogyakarta Nuha Med.*, 2012.
- [4] B. I. Page and R. Y. Shapiro, "Effect of Public Opinion on Policy." pp. 175–190, 1983.
- [5] J. Riany, M. Fajar, and M. P. Lukman, "Penerapan Deep Sentiment Analysis pada Angket Penilaian Terbuka Menggunakan K-Nearest Neighbor," vol. 6, no. 1, pp. 147–156, 2016.
- [6] M. DeHaff, "Sentiment Analysis, Hard But Worth It!," 2010. [Online]. Available: http://customerthink.com/sentiment_analysis_hard_but_worth_it/. [Accessed: 03-Mar-2017].
- [7] V. Chandani, R. S. Wahono, and . Purwanto, "Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen Review Film," *J. Intell. Syst.*, vol. 1, no. 1, pp. 55–59, 2015.
- [8] A. Hamzah, "Sentiment Analysis untuk Memanfaatkan Saran Kuesioner Dalam Evaluasi Pembelajaran dengan Menggunakan Naive Bayes Classifier (NBC)," *Snast*, no. November, pp. 18–24, 2014.
- [9] B. Santosa, "Tutorial Support Vector Machine," pp. 1–23, 1995.
- [10] K. Sembiring, "Tutorial SVM Bahasa Indonesia," *Training*, no. September, pp. 1–28, 2007.

