

Penerapan Klasifikasi Bayes Untuk Memprediksi Jenis Latihan Siswa Pencak Silat (Studi Kasus Pencak Silat PSHT)

Tedy Setiadi

Teknik Informatika, Fakultas Teknologi Industri
Universitas Ahmad Dahlan
Yogyakarta, Indonesia
tedy.setiadi@tif.uad.ac.id

Jamaludin

Teknik Informatika, Fakultas Teknologi Industri
Universitas Ahmad Dahlan
Yogyakarta, Indonesia
ju67020@gmail.com

Abstrak – Perguruan Pencak Silat PSHT saat ini memiliki ratusan siswa dengan berbagai latar belakang dan kemampuan yang berbeda. Selama ini pelatih dalam penentuan jenis latihan pencak silat di PSHT hanya menggunakan perkiraan intuisi semata, akibatnya sering terjadi ketidaksesuaian antara calon siswa dengan jenis latihan yang diikuti. Kondisi ini menyebabkan siswa sering tidak bisa maksimal dalam berlatih, disamping berdampak prestasi siswa tidak optimal dalam kompetisi. Dalam penelitian ini telah dikembangkan aplikasi untuk memprediksi jenis latihan siswa berdasarkan karakteristik yang sesuai menggunakan klasifikasi Naive Bayes. Aplikasi ini diharapkan dapat membantu pelatih dalam meningkatkan kemampuan siswa sesuai karakteristiknya sehingga mampu meningkatkan prestasinya. Hasil pengujian menggunakan *confusion matrix*, dengan 400 data *training* dan 20 data *testing* menghasilkan nilai presisi sebesar 95 %, nilai akurasi sebesar 95 % dan nilai *recall* sebesar 98%.

Kata Kunci: *Confusion Matrix*, Klasifikasi, Naive Bayes, Pencak Silat.

I. PENDAHULUAN

Perguruan pencak silat organisasi PSHT merupakan perguruan pencak silat yang sudah mempunyai nama besar dan memiliki ratusan siswa. Nama besar perguruan disebabkan dengan banyaknya prestasi yang diraih siswanya dalam berbagai event pertandingan yang diikuti. Faktor yang mempengaruhi prestasi para siswanya selain bakat dan kemampuan yang dimiliki adalah kesesuaian antara jenis latihan yang sesuai dengan karakter siswanya. Penentuan jenis pelatihan siswa ditentukan oleh pelatih atau pengurus pencak silat. Selama ini, dalam proses menentukan jenis latihan Pelatih masih memiliki beberapa masalah. Salah satu masalahnya adalah menentukan kesesuaian calon siswa dengan jenis latihan yang akan diikuti.

Berdasarkan fakta di lapangan, proses memprediksi jenis latihan siswa oleh pelatih saat ini masih menggunakan

perkiraan/intuitif semata. Akibatnya, sering terjadinya ketidak sesuaian dalam memprediksi jenis latihan siswa sehingga memberikan dampak tidak baik terhadap siswa pencak silat PSHT, seperti jenis latihan siswa yang tidak sesuai dengan batas kemampuan yang dimiliki oleh siswa. Akibatnya sering siswa merasa kelelahan yang berlebihan dari jenis latihan yang diikuti serta prestasi siswa yang kurang optimal.

Berdasarkan kriteria yang terdapat pada masing-masing siswa pencak silat PSHT, pelatih diharapkan dapat melakukan klasifikasi jenis latihan siswa yang sesuai. Pelatih dan pengurus organisasi nantinya akan tidak kesulitan dalam menentukan kelas jenis latihan siswa berdasarkan kriteria yang ada pada setiap siswa. Ada dua kelas jenis latihan siswa PSHT, yaitu kelas jenis latihan umum dan privat. Untuk kelas jenis latihan umum yaitu kelas jenis latihan yang akan diikuti siswa pencak silat PSHT dengan waktu latihan kurang lebih 8 jam latihan, sedangkan untuk kelas jenis latihan privat yaitu kelas jenis latihan yang akan diikuti oleh siswa di pencak silat PSHT dengan waktu latihan kurang lebih 5 jam latihan.

Disisi lain, sebagai organisasi pencak silat yang sudah mapan, PSHT telah mengelola dalam sistem komputer data siswa dan jenis latihan, data prestasi siswa dari tahun ke tahun. Namun saat ini data historis siswa tersebut belum digunakan secara maksimal. Untuk itu dalam penelitian ini dikembangkan sistem klasifikasi untuk memprediksi kecocokan calon siswa dengan jenis latihan yang diikuti. Harapannya memudahkan pelatih dalam meningkatkan kemampuan siswa sehingga prestasi yang diraih lebih maksimal.

II. KAJIAN PUSTAKA

A. Data Mining

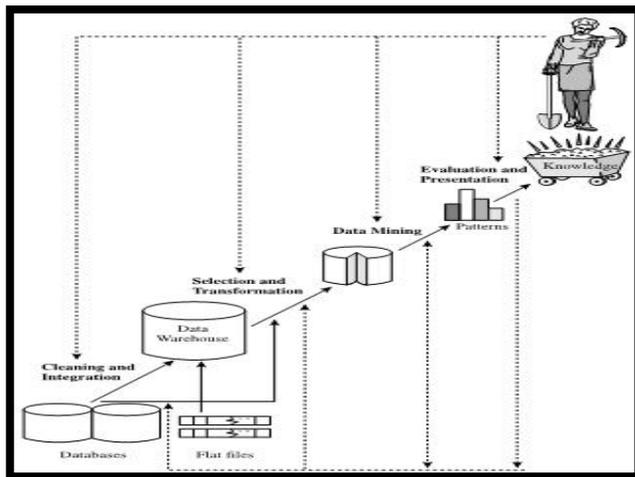
Banyak pendapat bagi peneliti tentang definisi *data mining*, diantaranya adalah sebagai berikut:

- a. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar [1].

- b. *Data mining* di sisi lain adalah kegiatan meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar, keluaran dari *data mining* ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan [2].
- c. *Data mining* atau sering disebut sebagai *knowledge discovery* (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam data berukuran besar [3].

B. Tahap-Tahap *Data Mining*

Data mining dapat dibagi menjadi beberapa tahap, dapat dilihat pada gambar 1.



Gambar 1. Tahap Proses *Mining* [3].

Berikut adalah tahap-tahap *data mining*, antara lain sebagai berikut :

- 1) **Pembersihan Data**
Pembersihan data merupakan proses menghilangkan data yang tidak konsisten atau tidak relevan (*noise*). Pada umumnya data yang diperoleh baik dari *database* maupun hasil eksperimen memiliki isian-isian yang tidak sempurna seperti data yang hilang, Data yang tidak valid atau hanya sekedar salah ketik. Selain itu ada juga atribut-atribut data yang tidak relevan dengan hipotesis *data mining* yang dimiliki. Pembersihan data juga akan mempengaruhi performansi dari teknik *data mining* karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.
- 2) **Seleksi Data**
Data yang ada pada *database* sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai kebutuhan saja untuk dianalisa yang akan diambil dari *database*.
- 3) **Transformasi Data**
Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*.
- 4) **Proses *Mining***

Merupakan proses utama saat metode diterapkan untuk menemukan informasi atau pengetahuan berharga dari data.

- 5) **Evaluasi Pola**
Untuk mengidentifikasi pola-pola menarik ke dalam basis pengetahuan yang ditemukan.
- 6) **Presentasi Pengetahuan**
Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

C. Penelitian Terkait

Penelitian untuk menentukan seleksi atlet pencak silat dengan membuat sistem klasifikasi yang mampu mengklasifikasi penerimaan seleksi atlet pencak silat yang layak lolos memenangkan pertandingan dengan menggunakan metode *Support Vector Machine* (SVM) [4]. Data yang digunakan sebanyak 110 data yang dibagi menjadi data latihan dan data uji dengan dua kelas hasil penerimaan seleksi yaitu lolos dan tidak lolos. Hasil rata-rata akurasi menggunakan metode SVM pada klasifikasi penerimaan seleksi atlet pencak silat sebesar 69,09%.

Penelitian membandingkan metode *Simple Logistic Classifier* dengan *Support Vector Machine* dalam memprediksi kemenangan atlet berdasarkan data kesehatan dan kondisi fisik atlet [5]. Hasil menunjukkan bahwa penggunaan metode SVM lebih unggul daripada SLC, baik dari segi kecepatan maupun dari nilai akurasi yang dihasilkan. Hasil dari pengujian juga telah ditunjukkan bahwa penggunaan histori data kondisi fisik dan kondisi kesehatan atlet dapat digunakan untuk memprediksi kemenangan atlet. Prediksi yang paling baik dilakukan dengan menggunakan semua fitur kesehatan dan fisik atlet

D. Algoritma Naive Bayes

Algoritma Naive Bayes merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema Bayes (atau aturan Bayes) dengan asumsi *independent* (ketidaktergantungan) yang kuat (naif). Dengan kata lain, dalam Naive Bayes, model yang digunakan adalah “model fitur *independent*”. Dalam Bayes (terutama Naive Bayes), maksud *independent* yang kuat pada fitur adalah bahwa sebuah fitur pada sebuah data tidak berkaitan dengan ada atau tidaknya fitur lain dalam data yang sama. Dalam Bayes, hal tersebut tidak dipandang sehingga masing-masing fitur seolah tidak memiliki hubungan apa pun [6].

Teorama Bayes memprediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya. Prediksi Bayes didasarkan pada teorama bayes dengan formula umum sebagai berikut :

$$P(H|X) = \frac{P(X|H) \times P(H)}{P(X)} \tag{1}$$

Klasifikasi Naive Bayes yang mengacu pada Teorama Bayes mempunyai persamaan sebagai berikut :

$$P(C_1|X) = \frac{P(X|H) \times P(C_1)}{P(X)} \tag{2}$$

Berikut adalah tabel penjelasan algoritma Naive Bayes, ditunjukkan pada Tabell.

Tabel 1. Penjelasan Algoritma Naive Bayes.

Parameter	Keterangan
P (H X)	Probabilitas akhir bersyarat (<i>conditionnal probability</i>) suatu hipotesis H terjadi jika diberikan bukti (<i>evidence</i>) X terjadi.
P (X H)	Probabilitas sebuah bukti E terjadi akan memengaruhi hipotesis H.
P (H)	Probabilitas awal (prior) hipotesis H terjadi tanpa memandang bukti apapun.
P (X)	Probabilitas awal (prior) bukti X terjadi tanpa memandang hipotesis atau bukti yang lain.
P	Peluang munculnya kelas (<i>prior</i>).
X	Data dengan kelas yang belum diketahui.
H	Hipotesis data X merupakan suatu <i>class</i> spesifik.
C	Variabel yang mempresentasikan kelas.

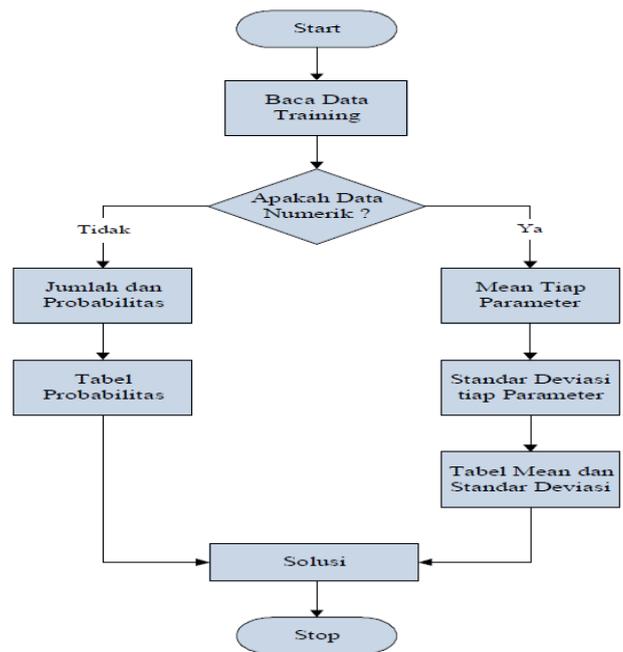
Ide dasar dari aturan Bayes adalah hasil dari hipotesis atau peristiwa (H) diperkirakan berdasarkan pada beberapa bukti (X) yang diamati. Ada beberapa hal penting dari aturan Bayes tersebut, yaitu [7]:

- 1) Sebuah probabilitas awal atau prior H atau P(H) adalah probabilitas dari suatu hipotesis sebelum bukti diamati.
- 2) Sebuah probabilitas akhir H atau P(H|X) adalah probabilitas dari suatu hipotesis setelah bukti diamati.

Contoh dalam suatu peramalan cuaca untuk memperkirakan terjadinya hujan, ada faktor yang memengaruhi terjadinya hujan, yaitu mendung. Jika diterapkan dalam Naive Bayes, probabilitas terjadinya hujan, jika bukti mendung sudah diamati, dinyatakan dengan:

$$\frac{P(\text{mendung}|hujan) \times P(hujan)}{P(\text{Mendung})} \tag{3}$$

Umumnya, Naive Bayes mudah dihitung untuk fitur bertipe kategori seperti pada kasus klasifikasi jenis latihan. Berikut adalah alur dari metode Naive Bayes dapat dilihat pada gambar 2.



Gambar 2. Alur Metode Naive Bayes.

Selain mengulas cara kerja dalam contoh kasus dengan menggunakan metode Naive Bayes, berdasarkan konsep dasar yang telah dibahas sebelumnya Naive Bayes juga mempunyai kelebihan dan kekurangan. Kelebihan dan kekurangannya yaitu sebagai berikut :

Kelebihan Naive Bayes:

- 1) Menangani kuantitatif dan data diskrit.
- 2) Mengeksplor batasan-batasan keoptimalan algoritma klasifikasi Naive Bayes dan mencoba menjelaskan mengenai alasan mengapa algoritma klasifikasi Naive Bayes berfungsi dengan baik pada berbagai jenis data.
- 3) Manfaat lain dari algoritma Naive Bayes adalah metode ini hanya membutuhkan jumlah data pelatihan (training data) yang kecil untuk menentukan estimasi parameter (rata-rata dan variansi dari variabel) yang diperlukan dalam proses pengklasifikasian.

Kekurangan Naive Bayes:

- 1) Metode Naive Bayes hanya bisa digunakan untuk persoalan klasifikasi dengan *supervised learning* dan data-data kategorikal.
- 2) Metode Naive Bayes memerlukan pengetahuan awal untuk dapat mengambil suatu keputusan. Tingkat keberhasilan metode ini sangat tergantung pada pengetahuan awal yang diberikan.
- 3) Tidak berlaku jika probabilitas kondisionalnya adalah nol. Apabila nol maka probabilitas prediksi akan bernilai nol juga.

D. Confussion Matrix

Confussion matrix adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep *data mining*. Rumus ini melakukan perhitungan

dengan 3 keluaran, yaitu : *recall*, *precision*, dan *accuracy*, ditunjukkan pada tabel 2.

Tabel 2. *Confusion Matrix*.

Class	Class Positif	Class Negatif
Prediksi Positif	TP	FN
Prediksi Negatif	FP	TN

Keterangan :

- 1) Jika hasil prediksi negatif dan data sebenarnya negatif (FN).
- 2) Jika hasil prediksi positif sedangkan nilai sebenarnya negatif (TN).
- 3) Jika hasil prediksi negatif sedangkan nilai sebenarnya positif (FP).
- 4) Jika hasil prediksi positif dan nilai sebenarnya positif (TP).

III. HASIL DAN PEMBAHASAN

A. Analisis Kebutuhan Data

Proses analisis ini dimulai dari identifikasi data yang dibutuhkan. Kebutuhan yang dimaksud adalah kebutuhan data/atribut untuk proses klasifikasi jenis latihan siswa. Digunaan 400 data siswa (*data training*) dari tahun 2012-2014 dengan 12 atribut. Atribut tersebut adalah Nama, Jenis Kelamin, Alamat, Umur, Agama, Status keluarga, Tanggungan, Status Siswa, Fisik, Kesehatan, Waktu Tempuh, dan Jenis Latihan. Dari 12 atribut tersebut sebagai atribut kelas adalah Jenis Latihan sedangkan atribut lainnya adalah atribut variabel.

B. Implementasi Sistem

Implementasi sistem sesuai tahap *data mining* yang telah ditentukan. Antara lain sebagai berikut:

1. Impor Data

Pada proses impor data ini akan menjadi proses awal *data mining* klasifikasi untuk memprediksi jenis latihan siswa, dimana *data training* yang akan digunakan diubah menjadi format CSV akan diimpor ke dalam *database* dengan jumlah 400 *record* dan 12 atribut. Dapat dilihat pada gambar 4.

Gambar 4. Proses Import Data.

2. Cleaning

Setelah proses impor data, akan dilakukan proses *cleaning* terhadap *data training* siswa pencak silat PSHT, variabel yang tidak memiliki informasi lengkap (*IS NOT NULL*) akan di-*cleaning*. Hasil *cleaning* dapat dilihat pada gambar 5.

Gambar 5. Proses *Cleaning*.

3. Selection

Setelah proses *cleaning*, dilakukan proses *selection*. Proses *selection* ini digunakan untuk menentukan atribut-atribut yang akan dipakai dalam proses klasifikasi. Hasil atribut setelah *selection* dapat dilihat pada gambar 6.

Gambar 6. Proses *Selection*.

4. Transformasi

Transformasi yang dilakukan mengubah atribut Umur serta Jarak Latihan siswa dari bertipe numerik menjadi tipe kategori. Hasil setelah proses transformasi dapat dilihat pada gambar 7.

Gambar 7. Proses Transformasi.

5. Klasifikasi Naive Bayes

Setelah tahapan *preprocessing* sebelumnya, selanjutnya dilakukan proses klasifikasi menggunakan metode Naive Bayes untuk memprediksi jenis latihan siswa. Proses klasifikasi dapat dilihat pada gambar 8.

Klasifikasi Probabilitas Fitur dan Kelas					
No.	Variabel	Umum		Privat	
		Probabilitas Jenis Kelamin = Pria Kelas Umum: 0.8302	Probabilitas Jenis Kelamin = Wanita Kelas Umum: 0.1698	Probabilitas Jenis Kelamin = Pria Kelas Privat: 0.7763	Probabilitas Jenis Kelamin = Wanita Kelas Privat: 0.2237
1	Jenis Kelamin	0.8302	0.1698	0.7763	0.2237
2	Umur	0.8765	0.1235	0.4342	0.5658
3	Status Perkawinan	0.4228	0.5772	0.9884	0.1316
4	Tanggungans	0.3025	0.6975	0.0132	0.9868
5	Waktu Tempuh	0.4259	0.5741	0.2237	0.7763
6	Status Siswa	0.7377	0.2623	0.7237	0.2763
7	Kondisi Fisik	0.8846	0.0154	0.5759	0.4211
8	Kesehatan	0.9105	0.0895	0.5526	0.4474
Jumlah Peserta Kelas Umum		324		0.8100	
Jumlah Peserta Kelas Privat			76		0.1900

Gambar 8. Proses Klasifikasi.

6. Prediksi jenis latihan siswa

Setelah nilai probabilitas fitur dan kelas telah diperoleh dari klasifikasi maka dapat dilakukan proses prediksi jenis latihan yang sesuai untuk seorang calon anggota. Proses ditunjukkan pada gambar 9.

Tahap Prediksi Calon Anggota

Prediksi Klasifikasi

Nama Lengkap:

Jenis Kelamin:

Agama:

Alamat:

Umur:

Status Perkawinan:

Tanggungans:

Waktu Tempuh:

Status Siswa:

Fisik:

Kesehatan:

Hasil Prediksi Klasifikasi

Jenis Kelamin:

Umur:

Status Perkawinan:

Tanggungans:

Waktu Tempuh:

Status Siswa:

Fisik:

Kesehatan:

Jenis Latihan:

Gambar 9. Proses Prediksi Jenis Latihan.

C. Pengujian

Pengujian menggunakan model *confusion matrix* untuk menguji keakuratan prediksi dari hasil yang telah diklasifikasikan. Diketahui 400 data digunakan sebagai *data training* dan 20 data sebagai *data testing*. Terdapat 308 data diprediksi bernilai *True* dan hasilnya benar Positif (TP), 4 data diprediksi bernilai *False* ternyata hasilnya sesuai

Negatif (FN), 72 data diprediksi *True* dan hasilnya ternyata Negatif (TN), dan 16 data diprediksi bernilai *False* ternyata hasilnya Positif (FP) sesuai dengan prediksi yang dilakukan dengan metode Algoritma Naive Bayes. Hasil ditunjukkan pada Tabel 3.

Tabel 3. *Confussion Matriks Data Mining*

Classification	Predicted Class	
	Kelas Umum	Kelas Privat
Prediksi Positif	308	4
Prediksi Negatif	16	72

$$\text{Precision} = \frac{Tp}{TP+FP} = \frac{308}{324} = 95\% \quad (4)$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{308}{312} = 98\% \quad (5)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{380}{400} = 95\% \quad (6)$$

Dari *data training* berjumlah 400 *record* data dan 20 *record data testing*, dihasilkan nilai *accuracy* sebesar 95%, nilai *precision* sebesar 95%, untuk nilai *recall* sebesar 98%. Dari hasil pengujian sistem dengan *confusion matrix* tersebut maka dapat disimpulkan bahwa hasil dari klasifikasi untuk memprediksi jenis latihan siswa dikatakan baik.

IV. KESIMPULAN

Penelitian ini telah berhasil membangun sistem klasifikasi dan prediksi jenis latihan pencak silat dengan metode Naive Bayes yang dapat membantu pengurus dan pelatih silat untuk memprediksi jenis latihan siswa pencak silat PSHT yang lebih baik dibanding secara intuitif.

Hasil pengujian dengan *confussion matrix* memberikan nilai akurasi 95%, nilai presisi 95%, dan dari *Recall* 98%. Berdasarkan hasil tersebut, maka sistem klasifikasi dan prediksi jenis latihan siswa diorganisasi pencak silat PSHT dapat digunakan dengan baik.

REFERENSI

[1] Ridwan, M. (2013). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Journal EECCIS*, Vol. 7, No.1, Juni 2013.

[2] Meilani, B.D. & Susanti, N. (2015). Aplikasi Data Mining Untuk Menghasilkan Pola Kelulusan Siswa Dengan Metode Naive Bayes. *Journal Ilmiah NERO*, Vol. 1, No. 3.

[3] Fithri, D.L. (2015). Model Data Mining Dalam Penentuan Kelayakan Pemilihan Tempat Tinggal Menggunakan Metode Naive Bayes. *Journal SIMETRIS*, Vol. 7 No. 2, November 2015.

[4] Harahap, E.H. (2017). Implementasi Algoritma Support Vector Machine (SVM) Untuk Penentuan Seleksi Atlet Pencak Silat. *Jurnal Pengembangan Teknologi Informasi*

- dan Ilmu Komputer, Vol. 2, No. 10, Oktober 2017, Hal. 3843-3848.
- [5] Rainarli, E. Perbandingan Simple Logistic Classifier dengan Support Vector Machine dalam Memprediksi Kemenangan Atlet. (2017). *Journal of Information Systems Engineering and Business Intelligence*, Vol. 3, No. 2, Oktober 2017.
- [6] Prasetyo, E. (2013). *Data Mining–Konsep dan Aplikasi*. Yogyakarta: ANDI.
- [7] Saputra, R.A. (2014). Penerapan Algoritma Naive Bayes Untuk Prediksi Penyakit Tuberculosis. *Swabumi*, V, September 2014.