

ALGORITMA NAÏVE BAYES UNTUK MELIHAT FAKTOR-FAKTOR YANG MEMPENGARUHI KULIT TERBAKAR

Rahmawati

Abstract— The skin is the most outer part of the body that has a strategic position for all living creatures, especially humans. How to know the parts of the burned skin can be classified with method Simple Naïve Bayesian algorithm based on attributes Hair, Height, Weight, Lotion. Simple Naïve Bayesian classifier is a potentially simple classification method based on the application of Bayes Theorem assuming an independent explanatory variables (independent). This algorithm utilizes probability and statistical methods proposed by the British scientist Thomas Bayes, which predicts the probability of the future based on the experience of earlier decades. In this case try to discuss the data sunburn where any factors that can affect the skin burned with Naïve Bayes algorithm and K-Nearest Neighbour.

Intisari— Kulit merupakan bagian yang paling luar dari tubuh yang mempunyai posisi yang strategis bagi semua makhluk hidup terutama manusia. Bagaimana cara untuk mengetahui bagian kulit yang terbakar dapat diklasifikasikan dengan metode algoritma Simple Naïve Bayesian berdasarkan atribut Hair, Height, Weight, Lotion. Simple Naïve Bayesian classifier merupakan salah satu metode pengklasifikasian berpeluang sederhana yang berdasarkan pada penerapan Teorema Bayes dengan asumsi antar variable penjelas saling bebas (independent). Algoritma ini memanfaatkan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas dimasa depan berdasarkan pengalaman dimasa sebelumnya. Dalam kasus ini mencoba membahas tentang data sunburn dimana faktor-faktor apa saja yang dapat mempengaruhi kulit terbakar dengan algoritma Naïve Bayes dan K-Nearest Neighbour.

Kata Kunci— Algoritma Naïve Bayes, Faktor-Faktor yang Mempengaruhi Kulit Terbakar , K-Nearest Neighbour.

I. PENDAHULUAN

Data mining atau biasa disebut dengan *Knowledge Discovery in Databases* (KDD) adalah pengambilan informasi yang tersembunyi, dimana informasi tersebut sebelumnya tidak dikenal dan berpotensi bermanfaat. Dalam data mining memiliki beberapa tahap seleksi, Pra-pengolahan, Transformasi, Data mining, sampai dengan tahap Interpretasi/evaluasi untuk pencarian pola. Dalam tahap input diperlukan sebuah data untuk diolah dengan menentukan algoritma tertentu yang kemudian akan diproses untuk menghasilkan suatu output atau model yang akan bermanfaat dimasa mendatang. Dalam kasus ini menganalisa faktor-faktor yang mempengaruhi kulit terbakar menggunakan algoritma naïve bayes.

Data yang digunakan adalah data sunburn yang diperoleh dari repository.

II. KAJIAN LITERATUR

A. DEFINISI DATA MINING

- 1) Melakukan ekstraksi untuk mendapatkan informasi penting yang sifatnya implisit dan sebelumnya tidak diketahui, dari suatu data [1].
- 2) Kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola dan hubungan dalam set data berukuran besar [2].

Data mining atau biasa disebut dengan *Knowledge Discovery in Databases* (KDD) adalah pengambilan informasi yang tersembunyi, dimana informasi tersebut sebelumnya tidak dikenal dan berpotensi bermanfaat [3]

Proses KDD (*Knowledge Discovery in Databases*) mempunyai beberapa tahapan. Ada beberapa Tahapan-tahapan menurut Fayyad [4][5] dalam KDD sebagai berikut:

- a) Seleksi. Dalam tahap ini dilakukan pemahaman terhadap masalah yang akan dicari solusinya melalui penemuan pengetahuan, pengetahuan sebelumnya yang relevan, penentuan tujuan dan pemilihan data yang akan dianalisis.
- b) Pra-pengolahan. Dalam tahap ini dilakukan proses pra-pengolahan data, termasuk pembersihan data.
- c) Transformasi. Dalam tahap ini dilakukan pengurangan dan/atau pengubahan tipe data, sehingga data siap untuk dipresentasikan ke teknik-teknik data mining.
- d) Data mining. Dalam tahap ini dilakukan pemilihan tugas data mining yang sesuai dengan tujuan dan algoritma data mining untuk pencarian pola.
- e) Interpretasi/evaluasi. Dalam tahap ini dilakukan visualisasi dan interpretasi terhadap pola yang ditemukan untuk dijadikan pengetahuan. Pengetahuan didefinisikan sebagai pola yang memenuhi kriteria *interestingness measure* tertentu (kesederhanaan, kepastian, potensi pemanfaatan atau kebaruan).

Dalam tahap ini dilakukan pemilihan tugas data mining yang sesuai dengan tujuan dan algoritma data mining untuk pencarian pola. Untuk contoh kasus yang akan dibahas akan dijelaskan di bagian pembahasan.

Proses data mining meliputi :

Input(data) → Metode(Algoritma Data Mining) → Output(Pola/Model)

Sumber: Kusri (2009)

Gambar 1. Proses Data Mining

Dimana dalam tahapan *input* dibutuhkan sebuah data yang akan dibahas, kemudian dengan menentukan dengan algoritma tertentu akan diproses untuk menghasilkan suatu output atau model yang akan bermanfaat dimasa mendatang

B. PERAN DATA MINING DAN ALGORITMA

Bidang ilmu Data mining mempunyai perannya sendiri untuk memecahkan masalah yang terbagi menjadi beberapa, antara lain:

- 1) *Estimation* (Estimasi):
Linear Regression, Neural Network, Support Vector Machine, etc
- 2) *Prediction/Forecasting* (Prediksi/Peramalan):
Linear Regression, Neural Network, Support Vector Machine, etc
- 3) *Classification* (Klasifikasi):
Naive Bayes, K-Nearest Neighbor, C4.5, ID3, CART, Linear Discriminant Analysis, etc
- 4) *Clustering* (Klastering):
K-Means, K-Medoids, Self-Organizing Map (SOM), Fuzzy C-Means, etc
- 5) *Association* (Asosiasi):
FP-Growth, A Priori, etc

III. METODE PENELITIAN

Dari fungsi-fungsi yang ada dalam data mining, metode yang digunakan pada paper ini untuk mengetahui faktor-faktor yang membuat kulit terbakar adalah algoritma Naïve Bayes dan K-Nearest Neighbour. Berikut penjelasannya.

A. Algoritma Naïve Bayes

Simple Naïve Bayesian classifier merupakan salah satu metode pengklasifikasian berpeluang sederhana yang berdasarkan pada penerapan Teorema Bayes dengan asumsi antar variable penjelas saling bebas (independen). Algoritma ini memanfaatkan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas dimasa depan berdasarkan pengalaman dimasa sebelumnya

Kelebihan :

1. Mudah diimplementasikan
2. Hasilnya cukup baik untuk sebagian besar kasus
3. Bila asumsi saling bebas dipenuhi, maka tingkat akurasi sangat tinggi

Kekurangan:

- a. Adanya asumsi saling bebas antar atributnya terkadang akan menurunkan tingkat akurasi
- b. Biasanya dalam kehidupan nyata selalu ada hubungan antar atribut sehingga asumsi saling bebas menjadi tidak terpenuhi

Kelebihan klasifikasi Bayesian secara umum:

1. Tingkat performansinya dapat dibandingkan metode klasifikasi lainnya seperti dengan

2. *Decision tree* dan *neural network* (terutama untuk Naïve Bayesian).
3. Setiap training data secara incremental dapat menambah/mengurangi peluang hipotesa benar.
4. Metode Bayesian dapat memberikan standar tertentu untuk pembuatan keputusan yang optimal dibandingkan metode lainnya

Dasar dari teorema Naïve digunakan dalam pemrograman adalah :

$$P(A|B) = (P(B|A) * P(A))/P(B)$$

Artinya Peluang kejadian A sebagai B ditentukan dari Peluang B saat A, Peluang A dan Peluang B

$$P(C|X) = P(X|C) \cdot P(C) / P(X)$$

Dimana :

P(X) bernilai konstan utk semua klas

P(C) merupakan frek relatif sample klas C

Dicari P(C|X) bernilai maksimum, sama halnya dengan P(X|C)·P(C) juga bernilai maksimum

IV. HASIL DAN PEMBAHASAN

Dalam kasus ini membahas tentang data sunburn dimana faktor-faktor apa saja yang dapat mempengaruhi kulit terbakar. Faktor-faktor tersebut terdiri dari atribut Hair, Height, Weight, Lotion. Berapa besar probabilitas faktor tersebut mempengaruhi kulit terbakar. Berikut pembahasannya

Tabel 1. Data Sunburn

No	Condition Attributes					Decision Attributes
	Name	Hair	Height	Weight	Lotion	Result
1	Sarah	blonde	average	Light	no	sunburned (positive)
2	Dana	blonde	tall	Average	yes	none (negative)
3	Alex	brown	short	Average	yes	none
4	Annie	blonde	short	Average	no	sunburned
5	Emily	red	average	Heavy	no	sunburned
6	George	red	average	Heavy	no	none
7	Pete	brown	Tall	Heavy	no	none
8	John	brown	average	Heavy	no	none
9	Katie	blonde	Short	Light	yes	none

Sumber: Kjellerstrand (2003)

Dari data tersebut dapat kita bagi menjadi beberapa data. Data konsisten : data yang setiap atributnya memiliki nilai target yang sama. Data tidak konsisten : jika data setiap

atributnya sama tapi nilai atributnya berbeda(memiliki nilai yang berbeda dalam satu keputusan). Data bias : memiliki target/keputusan yang berbeda sedangkan instance pada semua atributnya sama. Data set digunakan untuk memprediksi suatu kejadian dari fakta atau kenyataan yang diketahui sebelumnya. Prediksi dari suatu kejadian disebut Hipotesa. Hipotesa dituliskan dengan $H(\text{attribute}_1, \text{attribute}_2, \dots, \text{attribute}_n) = \text{keputusan}$

Metode Bayes menggunakan probabilitas bersyarat sebagai dasarnya. Dalam ilmu probabilitas bersyarat dinyatakan sebagai:

$$P(c | x) = \frac{P(x | c) P(c)}{P(x)}$$

Likelihood Class Prior Probability

Poster Probability Predictor Prior Probability

$P(c|x)$ posterior probability dari class (target) tiap predictor (attribute)

$P(c)$ prior probability dari class

$P(x|c)$ likelihood : probability dari predictor tiap class

$P(x)$ prior probability dari predictor

Pengujian rasio kemiripan(*likelihood ratio test*) adalah sebuah pengujian statistic dimana ratio dihitung diantara maksimum dari sebuah likelihood function dibawah null hypothesis dan maksimum dari batasannya.

Posterior probability dapat dihitung dengan cara membuat frequency table atribut terhadap target. Table ini kemudian dijadikan likelihood tables. Kemudian dengan persamaan Naïve Bayes dihitung posterior probabilitynya. Class yang memiliki Probabilitas tertinggi adalah outcome dari prediksinya

Dari Tabel 1 data sunburn maka dapat di analisa

- a. Probabilitas kemunculan setiap nilai untuk atribut Hair (A1)

Tabel 2. Probabilitas atribut Hair

Frequency Table		Result	
		None	Sunburned
Hair	Blonde	2	2
	Brown	3	0
	Red	1	1
Jumlah		6	3

Sumber: Hasil Penelitian (2014)

Perhitungan Probabilitas Hair

Tabel 3. Perhitungan atribut Hair

Likelihood Table		Probabilitas		
		None	Sunburned	
Hair	Blonde	2/6	2/3	5/9
	Brown	3/6	0/3	3/9
	Red	1/6	1/3	2/9
Jumlah		6/9	3/9	

Sumber: Hasil Penelitian (2014)

Keterangan:

The zero-frequency problem adalah kejadian dimana tidak frequency ada kemunculan sama sekali pada kejadian sebelumnya. Hal ini bisa diatasi dengan penambahan 1 untuk setiap atribut (*laplace estimator*). Contoh Hair=Brown tidak muncul pada kelasnya(0).

frequency table / likelihood table untuk Hair

Perhitungan → $P(x|c)$

$$P(x|c) = P(\text{Blonde} | \text{None}) = 2/6 = 0,33$$

$$P(x|c) = P(\text{Blonde} | \text{Sunburned}) = 2/3 = 0,67$$

$$P(x|c) = P(\text{Brown} | \text{None}) = 3/6 = 0,5$$

$$P(x|c) = P(\text{Brown} | \text{Sunburned}) = 0/3 = 0$$

$$P(x|c) = P(\text{Red} | \text{None}) = 1/6 = 0,17$$

$$P(x|c) = P(\text{Red} | \text{Sunburned}) = 1/3 = 0,33$$

Perhitungan → $P(c)$

$$P(c) = P(\text{None}) = 6/9 = 0,67$$

$$P(c) = P(\text{Sunburned}) = 3/9 = 0,33$$

Perhitungan → $P(x)$

$$P(x) = P(\text{Blonde}) = 5/9 = 0,56$$

$$P(x) = P(\text{Brown}) = 3/9 = 0,33$$

$$P(x) = P(\text{Red}) = 2/9 = 0,22$$

Posterior Probability

$$P(\text{None} | \text{Blonde}) = \frac{P(x | c) \cdot P(c)}{P(x)} = \frac{0,33 \times 0,67}{0,56} = 0,3948$$

$$P(\text{Sunburned} | \text{Blonde}) = \frac{P(x | c) \cdot P(c)}{P(x)} = \frac{0,67 \times 0,33}{0,56} = 0,3948$$

$$P(\text{None} | \text{Brown}) = \frac{P(x | c) \cdot P(c)}{P(x)} = \frac{0,5 \times 0,67}{0,33} = 0,3948$$

$$P(\text{Sunburned} | \text{Brown}) = \frac{P(x | c) \cdot P(c)}{P(x)} = \frac{0 \times 0,33}{0,33} = 0$$

$$P(\text{None} | \text{Red}) = \frac{P(x | c) \cdot P(c)}{P(x)} = \frac{0,17 \times 0,67}{0,22} = 0,517$$

$$P(\text{Sunburned} | \text{Red}) = \frac{P(x | c) \cdot P(c)}{P(x)}$$

$$= \frac{0,33 \times 0,33}{0,22} = 0,495$$

b. Probabilitas kemunculan setiap nilai untuk atribut Height (A2) = 0,33

Tabel 4. Probabilitas atribut Height

Frequency Table		Result	
		None	Sunburned
Height	Average	2	2
	Tall	2	0
	Short	2	1
Jumlah		6	3

Sumber: Hasil Penelitian (2014)

Tabel 5. Perhitungan atribut Height

Likelihood Table		Probabilitas		
		None	Sunburned	
Height	Average	2/6	2/3	4/9
	Tall	2/6	0/3	2/9
	Short	2/6	1/3	3/9
Jumlah		6/9	3/9	

Sumber: Hasil Penelitian (2014)

Perhitungan Probabilitas untuk Height

frequency table / likelihood table untuk Height

Perhitungan → $P(x|c)$

$$P(x|c) = P(\text{Average} | \text{None}) = 2/6 = 0,33$$

$$P(x|c) = P(\text{Average} | \text{Sunburned}) = 2/3 = 0,67$$

$$P(x|c) = P(\text{Tall} | \text{None}) = 2/6 = 0,33$$

$$P(x|c) = P(\text{Tall} | \text{Sunburned}) = 0/3 = 0$$

$$P(x|c) = P(\text{Short} | \text{None}) = 2/6 = 0,33$$

$$P(x|c) = P(\text{Short} | \text{Sunburned}) = 1/3 = 0,33$$

Perhitungan → $P(c)$

$$P(c) = P(\text{None}) = 6/9 = 0,67$$

$$P(c) = P(\text{Sunburned}) = 3/9 = 0,33$$

Perhitungan → $P(x)$

$$P(x) = P(\text{Average}) = 4/9 = 0,44$$

$$P(x) = P(\text{Tall}) = 2/9 = 0,22$$

$$P(x) = P(\text{Short}) = 3/9 = 0,33$$

Posterior Probability

$$P(\text{None} | \text{Average}) = \frac{P(x|c) \cdot P(c)}{P(x)} = \frac{0,33 \times 0,67}{0,44} = 1,020$$

$$P(\text{Sunburned} | \text{Average}) = \frac{P(x|c) \cdot P(c)}{P(x)} = \frac{0,67 \times 0,67}{0,44} = 1,020$$

$$P(\text{None} | \text{Tall}) = \frac{P(x|c) \cdot P(c)}{P(x)} = \frac{0,33 \times 0,67}{0,22} = 1,005$$

$$P(\text{Sunburned} | \text{Tall}) = \frac{P(x|c) \cdot P(c)}{P(x)}$$

$$= \frac{0 \times 0,33}{0,22} = 0$$

$$P(\text{None} | \text{Short}) = \frac{P(x|c) \cdot P(c)}{P(x)}$$

$$= \frac{0,33 \times 0,67}{0,33} = 0,67$$

$$P(\text{Sunburned} | \text{Short}) = \frac{P(x|c) \cdot P(c)}{P(x)}$$

$$= \frac{0,33 \times 0,33}{0,33} = 0,33$$

c. Probabilitas kemunculan setiap nilai untuk atribut Weight (A3) = 0,2475

Tabel 6. Probabilitas atribut Wight

Frequency Table		Result	
		None	Sunburned
Weight	Average	2	1
	Light	1	1
	Heavy	3	1
Jumlah		6	3

Sumber: Hasil Penelitian (2014)

Perhitungan Probabilitas untuk Wight

Tabel 7. Probabilitas atribut Wight

Likelihood Table		Probabilitas		
		None	Sunburned	
Weight	Average	2/6	1/3	3/9
	Light	1/6	1/3	2/9
	Heavy	3/6	1/3	4/9
Jumlah		6/9	3/9	

Sumber: Hasil Penelitian (2014)

frequency table / likelihood table untuk Weight

Perhitungan → $P(x|c)$

$$P(x|c) = P(\text{Average} | \text{None}) = 2/6 = 0,33$$

$$P(x|c) = P(\text{Average} | \text{Sunburned}) = 1/3 = 0,33$$

$$P(x|c) = P(\text{Tall} | \text{None}) = 1/6 = 0,17$$

$$P(x|c) = P(\text{Tall} | \text{Sunburned}) = 1/3 = 0,33$$

$$P(x|c) = P(\text{Short} | \text{None}) = 3/6 = 0,5$$

$$P(x|c) = P(\text{Short} | \text{Sunburned}) = 1/3 = 0,33$$

Perhitungan → $P(c)$

$$P(c) = P(\text{None}) = 6/9 = 0,67$$

$$P(c) = P(\text{Sunburned}) = 3/9 = 0,33$$

Perhitungan → $P(x)$

$$P(x) = P(\text{Average}) = 3/9 = 0,33$$

$$P(x) = P(\text{Tall}) = 2/9 = 0,22$$

$$P(x) = P(\text{Short}) = 4/9 = 0,44$$

Posterior Probability

$$P(\text{None} | \text{Average}) = \frac{P(x|c) \cdot P(c)}{P(x)} = \frac{0,33 \times 0,67}{0,33} = 1,67$$

$$P(\text{Sunburned} | \text{Average}) = \frac{P(x|c) \cdot P(c)}{P(x)} = \frac{0,33 \times 0,67}{0,33} = 1,67$$

$$P(\text{None} | \text{Tall}) = \frac{P(x|c) \cdot P(c)}{P(x)}$$

$$= \frac{0,17 \times 0,67}{0,22} = 0,517$$

$$P(\text{Sunburned} | \text{Tall}) = \frac{P(x|c) \cdot P(c)}{P(x)}$$

$$= \frac{0,33 \times 0,33}{0,22} = 0,495$$

$$P(\text{None} | \text{Short}) = \frac{P(x|c) \cdot P(c)}{P(x)}$$

$$= \frac{0,5 \times 0,67}{0,44} = 0,76$$

$$P(\text{Sunburned} | \text{Short}) = \frac{P(x|c) \cdot P(c)}{P(x)}$$

$$= \frac{0,33 \times 0,33}{0,44} = 0,2475$$

d. Probabilitas kemunculan setiap nilai untuk atribut Lotion (A4) = 1,5

Tabel 8. Probabilitas atribut Lotion

Frequency Table		Result	
		None	Sunburned
Lotion	Yes	3	0
	No	3	3
Jumlah		6	3

Sumber: Hasil Penelitian (2014)

Perhitungan Probabilitas Lotion

Tabel 9. Perhitungan atribut Lotion

Likelihood Table		Probabilitas	
		Non	Sunburn
		e	ed
Lotion	Yes	3/6	0/3
	No	3/6	3/3
Jumlah		6/9	3/9

Sumber: Hasil Penelitian (2014)

frequency table / likelihood table untuk Lotion

Perhitungan → $P(x|c)$

$$P(x|c) = P(\text{Yes} | \text{None}) = 3/6 = 0,5$$

$$P(x|c) = P(\text{Yes} | \text{Sunburned}) = 0/3 = 0$$

$$P(x|c) = P(\text{No} | \text{None}) = 3/6 = 0,5$$

$$P(x|c) = P(\text{No} | \text{Sunburned}) = 3/3 = 1$$

Perhitungan → $P(c)$

$$P(c) = P(\text{None}) = 6/9 = 0,67$$

$$P(c) = P(\text{Sunburned}) = 3/9 = 0,33$$

Perhitungan → $P(x)$

$$P(x) = P(\text{Yes}) = 3/9 = 0,33$$

$$P(x) = P(\text{No}) = 6/9 = 0,22$$

Posterior Probability

$$P(\text{None} | \text{Yes}) = \frac{P(x|c) \cdot P(c)}{P(x)}$$

$$= \frac{0,5 \times 0,67}{0,33} = 1,015$$

$$P(\text{Sunburned} | \text{Yes}) = \frac{P(x|c) \cdot P(c)}{P(x)}$$

$$= \frac{0 \times 0,33}{0,33} = 0$$

$$P(\text{None} | \text{No}) = \frac{P(x|c) \cdot P(c)}{P(x)}$$

$$= \frac{0,5 \times 0,67}{0,22} = 1,5227$$

$$P(\text{Sunburned} | \text{No}) = \frac{P(x|c) \cdot P(c)}{P(x)}$$

$$= \frac{1 \times 0,33}{0,22} = 1,5$$

Probabilitas kemunculan setiap nilai untuk atribut Result (A5)

Tabel 8. Probabilitas kemunculan tiap atribut

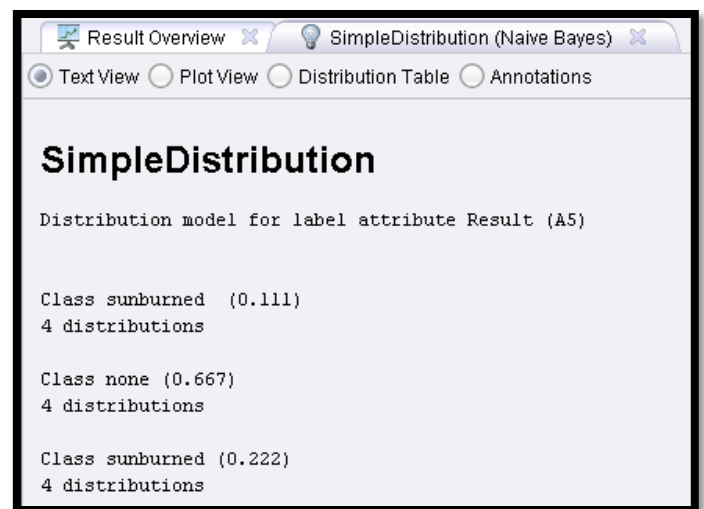
Result	Jumlah kejadian yg dipilih		Probabilitas	
	None	Sunburned	None	Sunburned
Jumlah	6	3	1	1

Sumber: Hasil Penelitian (2014)

Menghitung probabilitas setiap kejadian

None → $P(\text{None} | \text{Height=Average})$. $P(\text{None} | \text{Height=Average})$ Sunburned

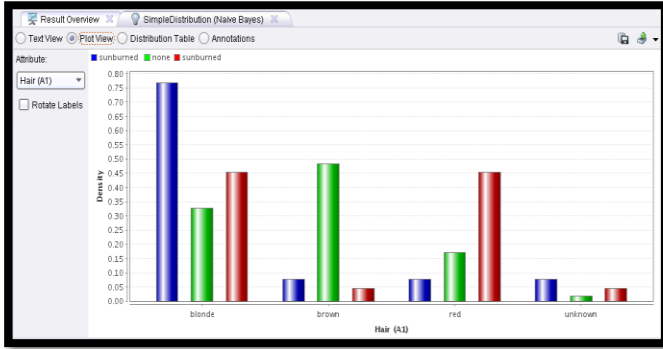
Hasil yang ditampilkan Sebagai text view pada RapidMiner, dimana data kelas sunburned bernilai 0.111



Sumber: Hasil Penelitian (2014)

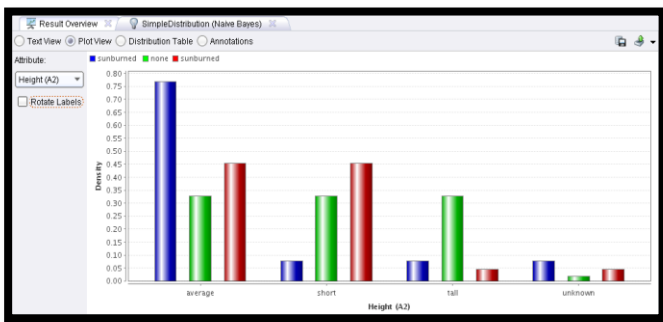
Gambar 2. Tampilan Sample Distribution

Pada atribut Hair, blonde merupakan sunburned yang paling tinggi



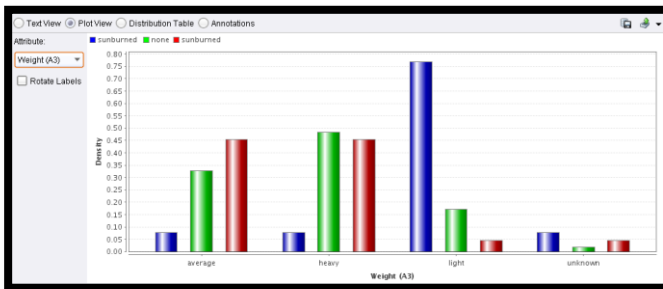
Sumber: Hasil Penelitian (2014)
Gambar 3. Tampilan hasil untuk atribut Hair

Pada atribut Height, average merupakan sunburned yang paling tinggi.



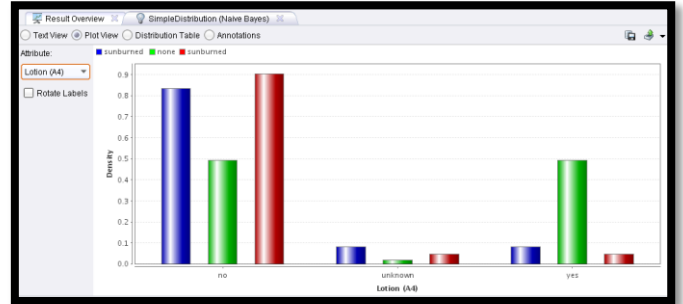
Sumber: Hasil Penelitian (2014)
Gambar 4. Tampilan hasil untuk atribut Height

Pada atribut Weight, light merupakan sunburned yang paling tinggi.



Sumber: Hasil Penelitian (2014)
Gambar 5. Tampilan hasil untuk atribut Weight

Pada atribut Lotion, data sunburned yang paling tinggi adalah No



Sumber: Hasil Penelitian (2014)
Gambar 6. Tampilan hasil untuk atribut Lotion

V. KESIMPULAN

Berdasarkan data sunburned tersebut, dimana ada beberapa faktor yang menyebabkan yaitu

1. Data mining tidak hanya digunakan sebagai tempat penyimpanan data/informasi tetapi juga dapat menyelesaikan permasalahan
2. Algoritma Naïve Bayes digunakan untuk menghitung probabilitas kulit yang terbakar
3. Terdapat kesimpulan probabilitas lass sunburned bernilai 0,111 dan 0, 222 dan juga class none bernilai 0,667
4. Semakin banyak jumlah data yang digunakan untuk training maka semakin tinggi keakuratannya
5. Semakin sering dilakukan learning atau penambahan knowledge maka akan semakin cepat mendapatkan klasifikasi dengan benar.

REFERENSI

- [1] Witten, Ian H., Frank Eibe, and Mark A. Hall., *Data Mining Practical Machine Learning Tools and Technique 3rd edition*. Elsevier, 2011.
- [2] Santosa, Budi, *DataMiningTeknik Pemanfaatan Data untuk Keperluan Bisnis*, Graha Ilmu: Yogyakarta, 2007.
- [3] Susanto, Sani dan Dedi Suryadi., *Pengantar Data Mining – Menggali Pengetahuan Dari Bongkahan Data*. Andi Publisher : Jakarta, 2010.
- [4] Kusrini, *Algoritma Data Mining*, Andi Offcet : Yogyakarta, 2009.
- [5] Larose, Daniel T., *Discovering Knowledge in Data: An introduction to Data Mining*, Wiley: New Jersey, 2005.
- [6] Hakan Kjellerstrand. 2003. <http://www.hakank.org/weka/> [Online]. Available: <http://www.hakank.org/weka/sunburn.arff>



Rahmawati, M.Kom. Tahun 2014 lulus dari Program Strata Satu (S1) Program Studi Sistem Informasi STMIK Nusa Mandiri Jakarta. Tahun 2014 lulus dari Program Strata Dua (S2) Program Studi Ilmu Komputer STMIK Nusa Mandiri. Saat ini menjadi Staf Pengajar di Bina Sarana Informatika.