
**IMPLEMENTASI DATA MINING UNTUK PREDIKSI
PENYEBARAN VIRUS HIV/AIDS DI BANDAR LAMPUNG
DENGAN TEKNIK *DECISION TREE*****Agustian Prakarsya¹⁾, Arif Prambayun²⁾**^{1,2)}*STMIK Lembah Dempo*Pos-el : ¹⁾*agustian.prakarsya@gmail.com*, ²⁾*prambayun@lembahdempo.ac.id***ABSTRACT**

Such free sex behavior if not balanced with the understanding of the dangers of disease as a result of this risky behavior will lead to the ease of contracting dangerous diseases. One of them is HIV / AIDS infection which until now more complex and is in an alarming situation as the number continues to increase especially in urban areas. Based on this problem, this research aims to analyze People who infected hiv diseases so as to facilitate the health of the world to know there are brapa people who terjangkit virus hiv with The development of information technology today is very useful for all circles. as well as in the world of health, So now information technology has become one of the most important needs, information technology can facilitate human in knowing all information. In research conducted by Health Research Association of hiv virus in Lampung, HIV case cumulatively in Province noted that the discovery of as many as 1,578 people while for AIDS as many as 629 cases. This suggests that hiv / aids disease is one of the most common health problems One method that can be used to predict hiv / aids disease is by using data mining. Data mining is a series of processes to explore the added value of a data set of knowledge that has not been known manually. The method for analyzing data in the application of data mining is Knowledge Discovery in Database (KDD) which consists of several steps, namely Cleaning and Integration, Selection and Transformation, Data Mining, and Evaluation and Interpretation

Keywords : HIV / Aids, Datamining, Virus, Health**ABSTRAK**

Perilaku seks bebas seperti itu jika tidak seimbang dengan pemahaman akan bahaya penyakit akibat perilaku berisiko ini akan mengakibatkan kemudahan tertular penyakit berbahaya. Salah satunya adalah infeksi HIV / AIDS yang sampai sekarang lebih kompleks dan dalam situasi yang mengkhawatirkan karena jumlahnya terus meningkat terutama di daerah perkotaan. Berdasarkan permasalahan ini, penelitian ini bertujuan untuk menganalisis orang yang terinfeksi penyakit hiv sehingga dapat memudahkan kesehatan dunia untuk mengetahui ada brapa orang yang terjangkit virus hiv dengan perkembangan teknologi informasi saat ini sangat bermanfaat bagi semua kalangan. Begitu juga dengan dunia kesehatan, maka kini teknologi informasi telah menjadi salah satu kebutuhan yang paling penting, teknologi informasi dapat memudahkan manusia

dalam mengetahui segala informasi. Dalam penelitian yang dilakukan oleh Asosiasi Riset Kesehatan hiv virus di Lampung, kasus HIV secara kumulatif di Provinsi mencatat bahwa ditemukan 1.578 orang sedangkan untuk AIDS sebanyak 629 kasus. Hal ini menunjukkan bahwa penyakit hiv / aids merupakan salah satu masalah kesehatan yang paling umum. Salah satu metode yang dapat digunakan untuk memprediksi penyakit hiv / aids adalah dengan menggunakan data mining. Data mining adalah serangkaian proses untuk mengeksplorasi nilai tambah kumpulan data pengetahuan yang belum diketahui secara manual. Metode untuk menganalisa data dalam aplikasi data mining adalah Knowledge Discovery in Database (KDD) yang terdiri dari beberapa tahap, yaitu Pembersihan dan Integrasi, Seleksi dan Transformasi, Data Mining, dan Evaluasi dan Interpretasi.

Kata kunci: *Hiv / Aids, Datamining, Virus, Kesehatan*

I. PENDAHULUAN

Indonesia merupakan negara berkembang dengan laju pertumbuhan penduduk sebesar 1,5% per tahun. Berdasarkan sensus penduduk yang terakhir pada tahun 2005 diperkirakan jumlah penduduk mencapai lebih dari 206 juta dan sebagian besar penduduk Indonesia hidup di pulau sumatra (39%) khususnya daerah bandar lampung. Kasus HIV/AIDS dilaporkan pertama kali di Provinsi Lampung pada tahun 2002 dari salah satu Kabupaten di Provinsi Lampung. Jumlah kasus HIV/AIDS yang dilaporkan dari seluruh Kabupaten/Kota tahun 2002 sampai dengan 2015 terus meningkat. Fase epidemiologis, perkembangan dan tingkat kematangan epidemi HIV di seluruh dunia dapat dibagi dalam 3 fase yaitu: fase awal: epidemi derajat

rendah, derajat menengah, epidemi terkonsentrasi dan fase lanjut/yang meluas (general epidemic). Indikator prevalensi HIV diharapkan <0,5%, di Provinsi Lampung tahun 2010-2015 berkisar 0,03%-0,04%. Gambaran jumlah kasus HIV/AIDS. (sumber : http://www.depkes.go.id/resources/download/profil/PROFIL_KES_PROVINSI_NSI_2015/08_Lampung_2015.pdf di akses 3 desember, 12:45 WIB)

Data BPS menunjukkan jumlah penduduk bandar lampung pada bulan Juli 2005 sebanyak 7,47 juta orang. namun sebagian besar rakyat Indonesia tergolong masyarakat miskin. Menurut Oey pada tahun 1998-1999 penduduk miskin Indonesia mencapai sekitar 24% dari jumlah penduduk atau hampir 40 juta orang. Tahun 2002-2003 terjadi penurunan menjadi 16,0% dan tahun

2004-2005 terjadi kenaikan lagi menjadi 16,7%. Sejak tahun 2006, tingkat kemiskinan nasional meningkat menjadi 17,8% dibandingkan tahun 2002-2005. Padatnya penduduk dan kemiskinan di daerah perkotaan serta kebutuhan ekonomi yang makin meningkat, menyebabkan banyak perempuan turut mencari nafkah terutama menjadi pekerja seks komersial karena tidak membutuhkan keterampilan dan uangnya mudah diperoleh. Perilaku seks bebas seperti ini jika tidak diimbangi dengan pemahaman tentang bahaya penyakit sebagai akibat dari perilaku berisiko ini akan menimbulkan mudahnya tertular penyakit berbahaya. Salah satunya adalah infeksi HIV/AIDS yang sampai saat ini makin kompleks dan berada pada situasi yang mengkhawatirkan karena jumlahnya meningkat terus khususnya di daerah perkotaan. Berdasarkan permasalahan hal tersebut maka penelitian ini bertujuan untuk menganalisis

Masyarakat yang terinfeksi penyakit hiv sehingga mempermudah dunia kesehatan untuk mengetahui ada brapa masyarakat yang terjangkit virus hiv dengan

Perkembangan teknologi informasi saat ini sangat berguna bagi semua kalangan. termaksud juga di dunia kesehatan , Sehingga Sekarang ini teknologi informasi telah menjadi salah satu kebutuhan yang sangat penting, Teknologi informasi dapat mempermudah manusia dalam mengetahui semua informasi. Sehingga, teknologi informasi dapat diimplementasikan dalam berbagai bidang, termasuk dunia kedokteran dan kesehatan.

Kesehatan merupakan aspek terpenting didalam setiap kehidupan, khususnya bagi anak mudah, masalah kesehatan yang sering muncul pada generasi mudah dengan pergaulan bebas . Pada penelitian yang dilakukan oleh Riset Kesehatan Dasar tentang virus hiv di bandar lampung, Kasus HIV secara kumulatif di Provinsi mencatat bahwa ditemukannya sebanyak 1.578 orang sedangkan untuk AIDS sebanyak 629 kasus. Hal ini menunjukkan bahwa penyakit hiv/aids merupakan salah satu masalah kesehatan yang sering muncul. (sumber

:<https://lampungai.com/berita-lampung-terkini/kasus-hiv-di->

lampung-1-578-orang/ di akses 3 desember 2017, 13:32 WIB)

Salah satu metode yang dapat digunakan untuk memprediksi penyakit hiv/aids adalah dengan menggunakan data mining. Data mining adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Metode untuk menganalisis data dalam penerapan data mining ini adalah Knowledge Discovery in Database (KDD) yang terdiri dari beberapa tahapan, yaitu Cleaning and Integration, Selection and Transformation, Data mining, dan Evaluation and Interpretation .

Adapun teknik yang digunakan dalam data mining adalah decision tree (pohon keputusan) dengan menggunakan algoritma C4.5. Dengan menggunakan teknik decision tree menggunakan algoritma C4.5, akan menghasilkan pola atau pattern, yaitu berupa rules yang dapat memprediksi penyakit.

II. METODE PENELITIAN

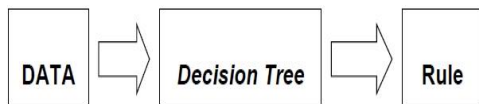
2.1 Data Mining

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode atau algoritma data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses Knowledge Discovery in Database (KDD). Adapun teknik yang digunakan pada proses data mining adalah dengan teknik decision tree dengan menggunakan algoritma C4.5.

2.1.1 Decision Tree

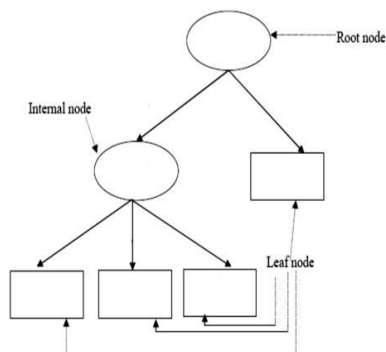
Decision tree merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Metode decision tree mengubah fakta yang sangat besar menjadi pohon keputusan yang mempresentasikan aturan. Aturan dapat dengan mudah dipahami dengan bahasa alami. Dan mereka juga dapat diekspresikan dalam bentuk basis data seperti Structure Query Language (SQL) untuk mencari record pada data tertentu. Sebuah decision tree adalah sebuah struktur

yang dapat digunakan untuk membagi kumpulan data yang besar menjadi himpunan-himpunan record yang lebih kecil dengan menerapkan serangkaian aturan keputusan. Pada decision tree setiap simpul daun menandai label kelas. Simpul yang bukan simpul akhir terdiri dari akar dan simpul internal yang terdiri dari kondisi tes atribut pada sebagian record yang mempunyai karakteristik yang berbeda. Simpul akar dan simpul internal ditandai dengan bentuk oval dan simpul daun ditandai dengan bentuk segi empat. Berikut adalah struktur decision tree seperti yang ditunjukkan pada Gambar 4.



Gambar 4. Struktur decision tree

Berikut adalah konsep decision tree seperti yang ditunjukkan pada Gambar 5.



Gambar 5. Konsep decision tree

Ada beberapa konsep dalam decision tree, antara lain:

- 1) Data dinyatakan dalam bentuk Tabel dengan atribut dan record.
- 2) Atribut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan tree. Misalkan untuk menentukan main
- 3) Tennis, kriteria yang diperhatikan adalah cuaca, angin dan temperatur. Salah satu atribut merupakan atribut yang menyatakan data solusi per-item data yang disebut dengan target atribut.
- 4) Atribut memiliki nilai-nilai yang dinamakan dengan instance

2.1.2 Algoritma C4.5

Algoritma C4.5 adalah salah satu algoritma yang digunakan dalam decision tree. Algoritma C4.5 merupakan pengembangan dari algoritma ID3 (Iterative Dichotomiser 3) yang dikembangkan oleh J.Ross Quinlan. Secara umum, langkah-langkah algoritma C4.5 dalam membangun decision tree adalah sebagai berikut.

- 1) Pilih atribut sebagai akar.
- 2) Buat cabang untuk tiap-tiap nilai.

- 3) Bagi kasus dalam cabang.
- 4) Ulangi proses untuk setiap proses cabang sampai semua kasus pada cabang memiliki kelas yang sama.

C4.5 merupakan salah satu algoritma yang telah secara luas digunakan, khususnya di area machine learning yang memiliki beberapa perbaikan dari algoritma sebelumnya, ID3, yaitu dalam hal metode pemangkasanya (prunning). Adapun perbaikannya adalah sebagai berikut :

- 1) Algoritma C4.5 menghitung gain ratio untuk masing-masing atribut, dan atribut yang memiliki nilai yang tertinggi akan dipilih sebagai simpul. Penggunaan gain ratio ini memperbaiki kelemahan dari ID3 yang menggunakan information gain.
- 2) Pemangkasan dapat dilakukan pada saat pembangunan pohon (tree) ataupun pada saat proses pembangunan pohon selesai.
- 3) Mampu menangani continues attribute.
- 4) Mampu menangani missing data.
- 5) Mampu membangkitkan rule dari sebuah pohon.

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus seperti tertera dalam persamaan 1 berikut.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Keterangan:

S = Himpunan

Kasus A = Atribut

n = Jumlah partisi atribut A

|Si| = Jumlah kasus pada partisi kei

|S| = Jumlah kasus dalam S

JUMLAH KASUS KEHILANGAN DAN SIFILIS MENURUT ENERGI KEHAYATAN DAN HABITAT/NOTA
 PROVINSI JAWA BARU
 TAHUN 2010

NO	KELOMPOK USIA	HIV			AIDS			SIFILIS			JUMLAH KEHAYATAN HABITAT AIDS					
		L	P	UAF	L	P	UAF	L	P	UAF	L	P	UAF			
1	1-4 TAHUN	5	11	36	4,30	3	4	7	5,47			0	0,00			0
2	5-14 TAHUN			0	0,00			0	0,00			0	0,00			0
3	15-24 TAHUN	11	1	12	1,39	1	1	1	0,78			0	0,00			0
4	25-34 TAHUN	27	15	42	15,51	11	6	17	15,28	1	1	1	14,29			0
5	35-44 TAHUN	124	100	224	77,26	69	17	86	75,00	3	3	6	66,78	14	3	18
6	45-54 TAHUN	3	0	13	3,36	3	4	7	5,47			0	0,00			0
JUMLAH HABITAT		222	141	363		87	41	128		4	3	7		14	3	18
PROPORSI ENERGI KEHAYATAN		46,02	38,23			17,97	12,03			17,14	42,86			70,88	26,52	

NO: jumlah kasus adalah seluruh kasus yang ada di wilayah kerja puskesmas tersebut termasuk kasus yang ditemukan di RS
 Sumber: Data RS Dinas Kesehatan Provinsi Lampung

Sementara itu, perhitungan nilai entropi dapat dilihat pada persamaan 2 berikut ini.

$$Entropy(S) = - \sum_{i=1}^n p_i * \log_2 p_i$$

Keterangan:

s= himpunan

Kasus A = Fitur

n = Jumlah partisi S

Pi = Proporsi dari Si terhadap S

III. HASIL DAN PEMBAHASAN

3.1 Proses pengolahan data

Data yang peneliti gunakan dalam penelitian ini adalah mengambil sample darah. Data diperoleh dari persentase donor darah di skring terhadap hiv - aids. Data yang dikumpulkan yaitu data jumlah pendonor yang di lakukan tes kepada sukarelawan di kota lampung

JUMLAH KASUS HIV, AIDS, DAN SIFILIS MENURUT JENIS KELAMIN DAN KABUPATEN/KOTA PROVINSI LAMPUNG TAHUN 2015

NO	KABUPATEN/KOTA	HIV			AIDS			SIFILIS			JUMLAH KEDATANGAN ANEBAT AIDS		
		L	P	L+P	L	P	L+P	L	P	L+P	L	P	L+P
1	LAMPUNG BARAT			0			0			0			0
2	TANGGAMUS			0	2	1	3			0			0
3	LAMPUNG SELATAN			0	14	3	17			0			0
4	LAMPUNG TIMUR	5	4	9	7	5	12			0			0
5	LAMPUNG TENGAH	5	12	17	6	5	11			0			0
6	LAMPUNG UTARA			0	3	1	4			0			0
7	WAY KANAN			0	1	1	2			0			0
8	TULANG BAWANG	9	3	12	6	3	9			0			0
9	PESAWARAN		1	1	4	2	6			0			0
10	PENDEBEHU	2	2	4	3	3	6			0			0
11	MESUJI			0	2		2			0			0
12	TULANG BAWANG BARAT	3	5	8			0			0			0
13	PESISIR BARAT			0			0			0			0
14	BANDAR LAMPUNG	198	116	314	40	17	57	4	3	7	14	5	19
15	METRO			0	3	3	6			0			0
JUMLAH (KAB/KOTA)		222	143	365	87	41	128	4	3	7	14	5	19
PROPOSIS JENIS KELAMIN		60,81	38,18		67,97	32,03		57,34	42,66		70,68	29,32	

Ket: Jumlah kasus adalah seluruh kasus baru yang ada di wilayah kerja puskesmas termasuk termasuk kasus yang ditemukan di RS
Sumber: Sekeloa PZ Dinas Kesehatan Provinsi Lampung

JUMLAH KASUS HIV, AIDS, DAN SIFILIS MENURUT JENIS KELAMIN DAN KABUPATEN/KOTA PROVINSI LAMPUNG TAHUN 2015

NO	KELOMPOK UMUR	HIV			AIDS			SIFILIS			JUMLAH KEDATANGAN ANEBAT AIDS		
		L	P	L+P	L	P	L+P	L	P	L+P	L	P	L+P
1	1-4 TAHUN	5	11	16	4,30	5	4	7	5,47			0	0,00
2	5-14 TAHUN			0	0,00			0	0,00			0	0,00
3	15-24 TAHUN	11	1	12	3,20	1	1	2	1,76			0	0,00
4	25-34 TAHUN	27	15	42	11,51	11	6	17	13,28	1	1	2	1,56
5	35-44 TAHUN	174	108	282	77,20	89	27	116	90,00	3	3	6	4,71
6	45-54 TAHUN	5	8	13	3,56	3	4	7	5,47			0	0,00
JUMLAH (KAB/KOTA)		222	143	365		87	41	128		4	3	7	
PROPOSIS JENIS KELAMIN		60,81	38,18		67,97	32,03		57,34	42,66		70,68	29,32	

Ket: Jumlah kasus adalah seluruh kasus baru yang ada di wilayah kerja puskesmas termasuk termasuk kasus yang ditemukan di RS
Sumber: Sekeloa PZ Dinas Kesehatan Provinsi Lampung

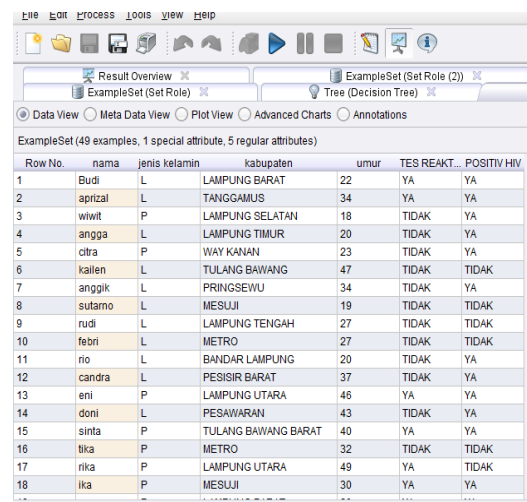
PERSENTASE DONOR DARAH DITINGKAT TERHADAP HIV - AIDS MENURUT UTA-PRA-PROVINSI LAMPUNG TAHUN 2015

NO	UNIT TRANSFUSI DARAH/UNIT DONOR DARAH (UDD)	DONOR DARAH																				
		JUMLAH PERDONOR						REAKTIF HIV			POSITIF HIV											
		L	P	L+P	JUMLAH	%	JUMLAH	%	JUMLAH	%	JUMLAH	%										
1	1	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	UTD-Praktik PM Provinsi Lampung	20.000	22.523	42.523	28.882	100	10.330	100	42.158	100	56	0,13	22	0,07	78	0,18	58	0	22	0	78	0
2	UTD-PM Kota Metro	5.660	1.802	11.358	9.833	100	1.555	100	11.345	100	15	0,20	33	0,29	22	0,20	18	0	3	0	22	0
3	UTD-PM Lampung Utara	3.660	684	4.344	3.668	100	680	100	4.336	100	4	0,11	1	0	0,11	0	0,11	0	0	1	0	0
4	UTD-PM Lampung Tengah	3.385	1.077	4.462	3.385	100	1.077	100	4.422	100	6	0,18	11	0,25	8	0,18	0	0	0	0	0	0
5	UTD-PM Lampung Selatan	2.785	454	3.239	2.785	100	458	100	3.223	100	7	0	0	0	7	0	0	0	0	0	0	0
6	UTD-PM Lampung Timur	880	231	1.111	705	97	230	97	1.015	97	11	0	0	0	0	0	0	0	0	0	0	0
7	UTD-PM Tang. Bawang	1.121	393	1.514	1.121	100	398	100	1.501	100	11	1	0	0	11	1	11	0,89	0	0	11	0,89
JUMLAH		30.440	27.087	57.527	38.407	100	17.970	100	60.019	100	105	0,20	201	0,35	231	0,39	168	0,17	215	0	111	0,20

Sumber: PM Provinsi Lampung

3.2 Proses Data Mining

Pada tahap selanjutnya data akan di proses menggunakan bantuan Tools Rapid Miner dan untuk pengujianya menggunakan cross validation yang adadi dalam Rapid Miner, maka data akan terlihat nilai akurasi, pohon keputusan, dan rulenya. Berikut adalah Data yang akan di import ke dalam Rapid Miner maka akan terlihat nilai akurasi,pohon keputusan dan rulenya. Berikut adalah data yang akan di inport ke dalam ripped miner :



memprediksi virus hiv/aids, dilakukan evaluasi dengan supplied test set menggunakan tes dna dan tes darah maka Berdasarkan permasalahan hal tersebut penelitian ini bertujuan untuk menganalisis Masyarakat yang terinveki penyakit hiv sehingga mempermudah dunia kesehatan untuk mengetahui ada brapa masyarakat yang teerjangkit virus hiv dengan Perkembangan teknologi informasi saat ini sangat berguna bagi semua kalangan. jadi hal ini mempermudah menunjukkan bahwa terdapat 265 instances yang akurat dan 21 instances yang error atau prediksinya salah.

V. SARAN

Untuk meningkatkan kinerja dan menyempurnakan penelitian yang telah dibuat, peneliti memberikan saran agar penelitian ini dapat dikembangkan lagi dengan menggabungkan atau membandingkan dengan algortima klasifikasi lain

untuk mendapatkan hasil prediksi yang lebih baik.

DAFTAR PUSTAKA

- Han J dan Kamber M. 2006. Data mining Concepts and Techniques 2nd Edition. The Morgan Kaufmann Publisher, San Fransisco.
- http://www.depkes.go.id/resources/download/profil/PROFIL_KES_PROVINSI_2015/08_Lampung_2015.pdf di akses 3 desember, 12:45 WIB
- <https://lampungsai.com/berita-lampung-terkini/kasus-hiv-di-lampung-1-578-orang/> di akses 3 desember 2017, 13:32 WIB
- Muzakir, Ari. 2016, <http://journal.unnes.ac.id/nju/> (di akses tanggal 1 november 2017)
- Pramudiono, I .2006. Apa itu Data Mining? Dalam <http://datamining.japati.net/cgi-bin/indodm.cgi>. Diakses tanggal 26 November 2015