

SISTEM PENDUKUNG KEPUTUSAN DETEKSI PENYAKIT KANKER PAYUDARA MENGGUNAKAN ALGORITMA NAIVE BAYES

Ivandari¹⁾, Erni Rahmawatie²⁾, M. Adib Al Karomi³⁾

¹ Program Studi Teknik Informatika, STMIK Widya Pratama
email: ivandarialkaromi@gmail.com

² Program Studi Sistem Informasi, STMIK Widya Pratama
email: erni.stmik.wp@gmail.com

³ Program Studi Teknik Informatika, STMIK Widya Pratama
email: adib.comp@gmail.com

Abstract

Cancer is one of the biggest causes of death in the world. Data from the International Agency for Research of Cancer (IARC) states that in 2012 more than 8.2 million people died from cancer. From these data it is known that breast cancer is the most common type of cancer suffered by 19.2% of all cancer cases. The amount of data and records related to cancer patient cases can be useful if from this data an information or new knowledge can be retrieved. Data mining is a field of knowledge that processes past data to be used as new information and knowledge. From the comparative research of data mining algorithms for detection of breast cancer in 2017 naive bayes is the best algorithm. Naive Bayes is proven to have a higher level of accuracy than other algorithms. In this study a decision support system for the detection of breast cancer was made. The system created using this Excel application can be one of the recommendations. The method used for calculation is the probability of naive bayes..

Keywords: *Naive bayes, Microsoft excell, decission support system*

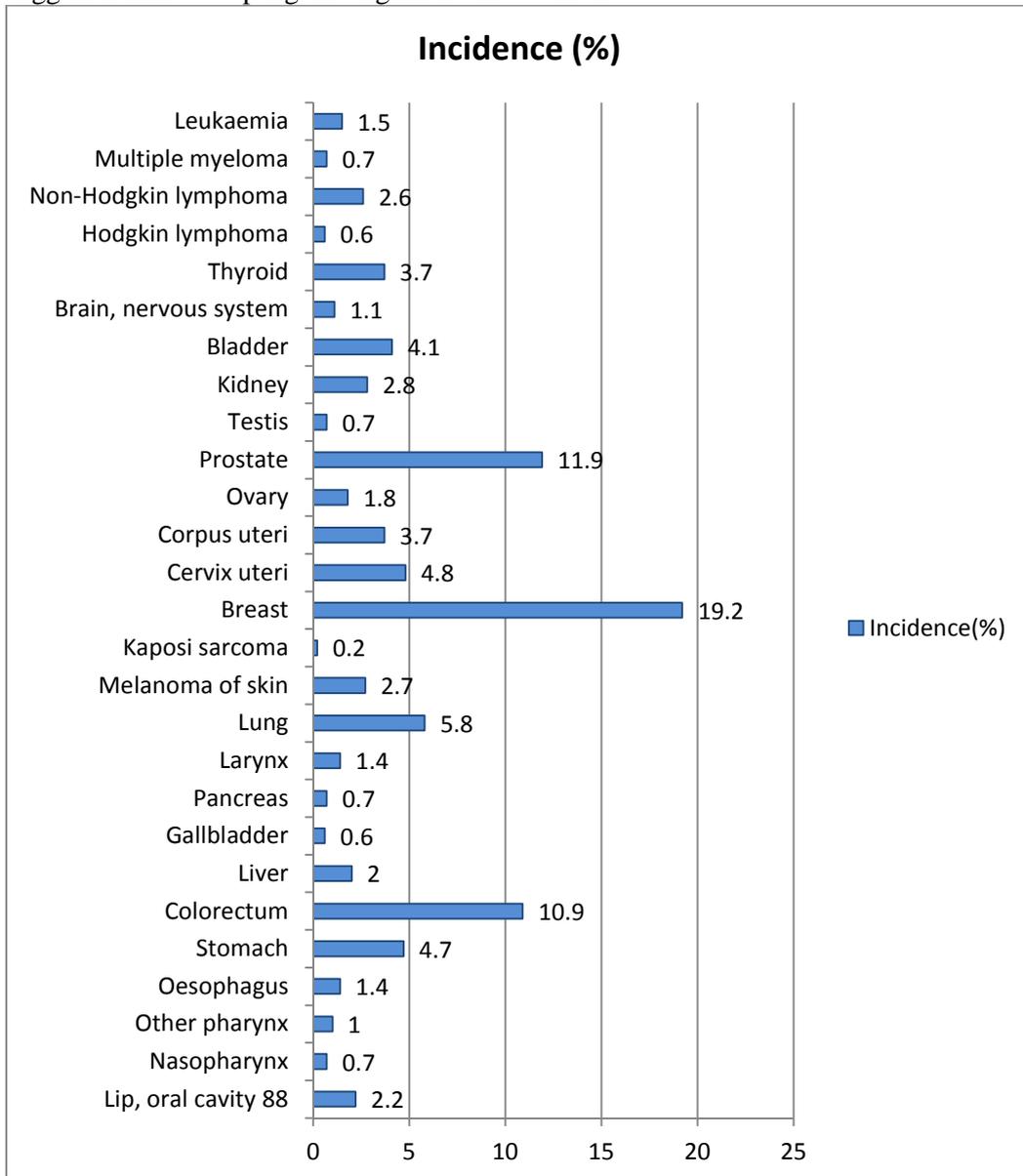
1. PENDAHULUAN

Kanker merupakan salah satu penyakit mematikan. Pada tahun 2012 *International Agency for Research of Cancer* (IARC) mencatat kasus penyakit kanker sebanyak 14.067.894 jiwa dan lebih dari 8,2 juta jiwa meninggal dunia akibat penyakit kanker. Sedangkan dalam 5 tahun terakhir tercatat penderita kanker payudara merupakan yang terbanyak yaitu 19,2% dari keseluruhan kasus. Gambar 1 merupakan grafik presentase penderita jenis kanker dalam 5 tahun terakhir berdasarkan data yang dihimpun oleh *GLOBOCAN International Agency for Research of Cancer* (Iarc. 2012).

Pencatatan terhadap penyakit kanker banyak dilakukan guna mengantisipasi dan menganalisa pasien sejak dini agar dapat dilakukan pencegahan. Dengan mengetahui lebih dini penyakit kanker maka penanganannya pun akan semakin mudah karena sel kanker belum berkembang lebih jauh. Beberapa pencatatan menghasilkan data yang valid dan telah teruji secara klinis. Banyaknya *record* yang ada membuat data menjadi susah dibaca secara manual. Data yang besar bisa saja tidak berarti dan hanya akan menjadi sampah. Data mining dapat menjadikan data yang sebelumnya tidak berarti menjadi sebuah informasi atau pola dengan proses tertentu (Witten, Frank, and Hall 2011). Data mining juga memungkinkan terbentuknya sebuah model ataupun aturan dalam data yang sebelumnya tidak berharga (Prasetyo 2012). Klasifikasi merupakan peranan data minig yang memungkinkan perolehan informasi dan pengetahuan baru dari data lampau. Salah satu algoritma klasifikasi terbaik adalah naive bayes (Wu 2009).

Komparas algoritma klasifikasi terbaik terhadap dataset kanker payudara pernah dilakukan dengan hasil naive bayes menjadi algoritma dengan tingkat akurasi terbaik (Kurniawan and Ivandari 2017). Melihat beberapa kelebihan dari algoritma naive bayes

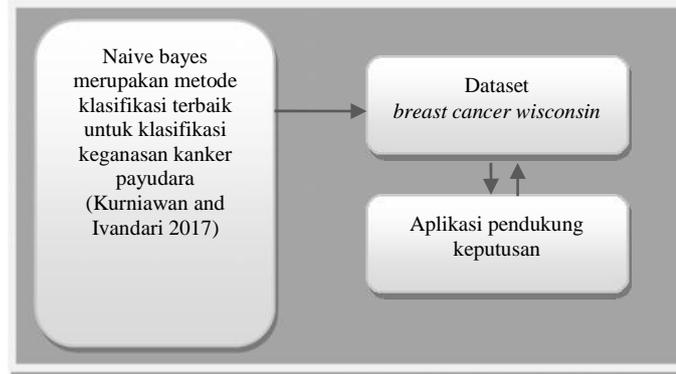
tersebut, maka akan dibuat sebuah sistem pendukung keputusan deteksi penyakit kanker payudara menggunakan algoritma naive bayes. Aplikasi ini akan dibuat dengan menggunakan metode pengembangan sistem waterfall



Gambar 1. Persentase Penderita Jenis Kanker 5 tahun terakhir (Iarc. 2012)

2. KAJIAN LITERATUR DAN PENGEMBANGAN HIPOTESIS

Klasifikasi tingkat keganasan kanker payudara pernah diimplementasikan menjadi sebuah aplikasi dengan menggunakan algoritma C4.5. Dalam pengembangan aplikasi tersebut digunakan metode pengembangan sistem *System Development Live Cycle Model* (SDLC) atau biasa juga disebut dengan istilah *Waterfall*. Bahasa pemrograman yang digunakan adalah bahasa java dengan menggunakan database SQL. Dataset yang digunakan diambil dari data UCI repository dengan hasil tingkat akurasi sebesar 95,57% (Nursela 2014). Gambar 2 merupakan kerangka pemikiran dalam penelitian ini.



Gambar 2 Kerangka Pemikiran

3. METODE PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini adalah eksperimental dengan metode pengembangan sistem waterfall. Beberapa tahapan dalam metode waterfall dalam penelitian ini adalah sebagai berikut:

3.1 Analisa kebutuhan sistem

Tahapan analisa dilakukan dengan mengumpulkan data terkait untuk melakukan perhitungan. Pengumpulan data ini dilakukan dengan mengambil dari dataset publik yaitu data uci repository. Dataset breast cancer wisconsin yang digunakan memiliki 699 record dengan 11 atribut. 1 atribut menjadi label yaitu atribut kelas dengan isian 2 untuk kanker jinak dan 4 untuk kanker ganas. 10 atribut lainnya antara lain atribut id, clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nucleoli, serta mitoses. Setelah melakukan pengumpulan data dilakukan analisa kebutuhan sistem yaitu mengacu pada sistem pendukung keputusan. Software yang dibutuhkan antara lain untuk database menggunakan sql dan untuk yang lain hanya menggunakan aplikasi bawaan windows seperti halnya notepad dan browser

3.2 Desain sistem

Dalam tahapan desain akan dilakukan dalam dua tahapan, yaitu desain sistem akan menggunakan diagram use case serta diagram activity. Diagram activity dan use case digunakan untuk menggambarkan alur program serta alur data yang sesuai dengan tahapan algoritma. Desain ini digunakan sebagai acuan dan sebagai alat untuk penanggulangan kesalahan. Sedangkan desain tampilan akan menggunakan Lembar Kerja Tampilan (LKT). LKT digunakan sebagai gambaran tampilan program aplikasi yang akan dibuat. Dalam LKT akan terdapat beberapa form yang dapat disesuaikan dengan kebutuhan sistem yang dilakukan sebelumnya

3.3 Koding

Penulisan kode program atau koding merupakan tahapan utama dalam metode waterfall. Karena dengan melakukan koding aplikasi yang direncanakan akan dibuat sesuai dengan rencana. Baik buruknya aplikasi sebagian besar ditentukan dalam tahapan ini. Pada awalnya tahapan ini akan dilakukan dengan menggunakan bahasa pemrograman php. Database akan digunakan dengan bantuan sql. Aplikasi yang akan dibuat nantinya dalam bentuk web dapat digunakan untuk deteksi dini penyakit kanker payudara. Akan tetapi seiring perkembangan dan singkatnya waktu penelitian dan penggunaan anggaran maka aplikasi ini dibuat dengan memanfaatkan microsoft excell. Microsoft excell dianggap memadai dan dapat melakukan perhitungan naive bayes dengan menyesuaikan dataset yang ada.

3.4 Pengujian Program

Tahapan berikutnya setelah dilakukan pengkodean adalah tahap pengujian. Dalam tahapan ini akan digunakan metode whitebox dan blackbox. Metode whitebox akan digunakan untuk pengujian kode program serta alur program. Pengujian ini akan menyesuaikan kebutuhan awal dengan program yang tercipta setelah tahap pengkodean. Tahap pengujian dengan whitebox ini merupakan pengujian program secara terperinci. Sedangkan pengujian blackbox digunakan untuk menguji tampilan dari aplikasi yang tercipta.

2.5 Penerapan dan Pemeliharaan

Tahapan penerapan merupakan tahap akhir dari pengembangan sistem. Tahap penerapan pada umumnya terkait implementasi sistem terhadap user yang ada. Dalam penelitian ini tahapan penerapan hanya akan diuji dengan data sampling yang ada. Sedangkan untuk pemeliharaan akan dilakukan setelah sistem benar benar teruji dan digunakan dalam dunia kesehatan.

4. HASIL PENELITIAN

4.1 Hasil Penelitian

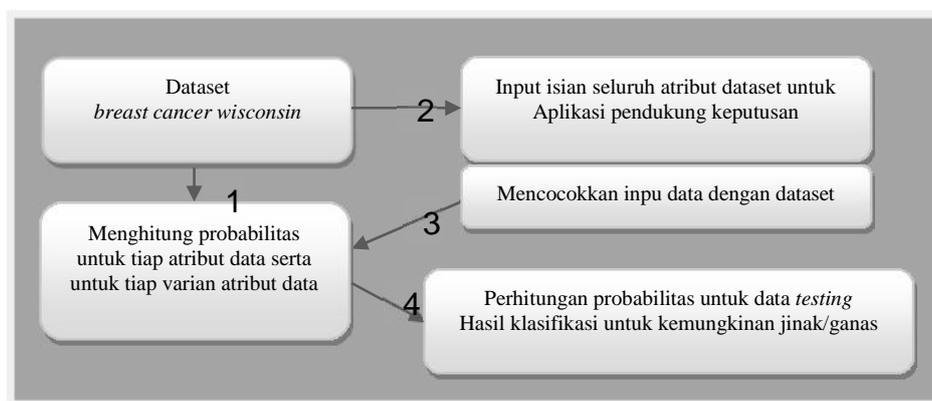
Dari hasil perhitungan menggunakan rapid miner, naive bayes memiliki tingkat akurasi sebesar 97,43%. Tingkat akurasi ini dihitung dari rata-rata percobaan yang dilakukan menggunakan 10 folds cross validation. Tingkat akurasi ini didapat dari perhitungan menggunakan confusion matrix. Gambar 3 merupakan hasil confusion matrix dari aplikasi rapid miner.

accuracy: 97.43% +/- 2.37% (mikro: 97.42%)			
	true 2	true 4	class precision
pred. 2	442	2	99.55%
pred. 4	16	239	93.73%
class recall	96.51%	99.17%	

Gambar 3 Tingkat akurasi naive bayes

4.2 Alur Aplikasi

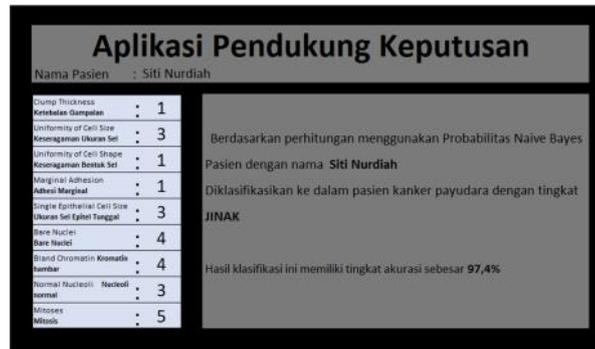
Microsoft excell dalam hal ini yang digunakan untuk membuat aplikasi merupakan program yang kompleks dan mudah digunakan (*user friendly*). Perancangan aplikasi menggunakan microsoft excell dapat dilakukan dengan memanfaatkan fitur yang ada. Beberapa link yang digunakan juga sangat mudah dengan adanya inisialisasi cell pada microsoft excell. Gambar 4 merupakan alur aplikasi yang memanfaatkan microsoft excell dalam pembuatannya. (Kurniawan and Ivandari 2017).



Gambar 4 Alur aplikasi

4.3 Hasil Sistem

Aplikasi yang tercipta dapat diunduh dari url: <http://repository.stmik-wp.ac.id/index.php/download/file/UkVQTzE1MzE4MDY3MTMjMTEtMDgxMS0wMDAxIzN3eXNwbnhndXk5>. Tampilan menu utama aplikasi sebagaimana dalam gambar 5 berikut. Terdapat beberapa isian yang harus diisi antara lain: ketebalan gumpalan, keseragaman ukuran sel, keseragaman bentuk sel, adhesi marginal, ukuran sel epitel tunggal, bare nuclei, kromotin hambar, nuclei normal, serta mitosis. Dari kesemua isian tersebut harus diisi dengan angka antara 1 sampai dengan 10.



Gambar 5 Tampilan aplikasi yang tercipta

Di bagian atas harus diisi nama pasien (boleh dengan nama alias). Sedangkan bagian kanan terdapat bagian hasil perhitungan dari aplikasi. Dalam gambar 5 tersebut Siti Nurdiah diklasifikasikan mengidap penyakit kanker jinak. Tingkat akurasi dari perhitungan ini adalah 97,4%. Tingkat akurasi naive bayes ini dihitung sebelumnya dengan menggunakan aplikasi rapid miner menggunakan *X-Validation* dan *confusion matrix*.

5. SIMPULAN

Dari hasil penelitian yang telah dilakukan didapatkan beberapa kesimpulan antara lain:

1. Terciptanya sebuah aplikasi klasifikasi penyakit kanker payudara. Aplikasi yang tercipta berbasis atau menggunakan microsoft excell.
2. Tingkat akurasi algoritma naive bayes untuk klasifikasi keganasan kanker payudara adalah 97,43%. Akurasi ini didapat dengan menggunakan dataset breast cancer wisconsin dengan memanfaatkan aplikasi rapid miner. Perhitungan menggunakan x-validation serta confusion matrix.

6. REFERENSI

- Iarc., International Agency for Research on Cancer. World Health Organization. 2012. "GLOBOCAN 2012: Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012." *Globocan*. doi:10.1002/ijc.27711.
- Kurniawan, M. Faisal, and Ivandari. 2017. "Komparasi Algoritma Data Mining Untuk Klasifikasi Kanker Payudara." *IC Tech I* April 20: 1–8.
- Nursela, Dwi Ayu. 2014. "Penerapan Algoritma C4 . 5 Untuk Klasifikasi Tingkat Keganasan Kanker Payudara," 1–5.
- Prasetyo, Eko. 2012. *Data Mining Konsep Dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi Offset.
- Witten, I. H., E. Frank, and M. A. Hall. 2011. *Data Mining: Practical Machine Learning Tools and Techniques 3rd Edition*. Vol. 40. Elsevier. doi:10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C.
- Wu, Xindong. 2009. *The Top Ten Algorithms in Data Mining*. Edited by Vipin Kumar. New York: Taylor & Francis Group, LLC.