

## Klasifikasi Jenis Kanker Berdasarkan Struktur Protein Menggunakan Algoritma Naive Bayes

Tawang Wulandari<sup>1</sup>, Marji<sup>2</sup>, Lailil Mufflikah<sup>3</sup>

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya  
Email: <sup>1</sup>tawang.wulandari@yahoo.com, <sup>2</sup>marji@ub.ac.id, <sup>3</sup>lailil@ub.ac.id

### Abstrak

Kanker adalah beberapa sel tubuh yang mulai membelah tanpa berhenti dan menyebar ke jaringan sekitarnya. Setiap tahun terdapat ribuan kasus baru kanker yang menyerang warga Indonesia. Terlambatnya deteksi dini menyebabkan banyak kasus kematian akibat kanker. Faktor penyebab kanker adalah faktor genetik dan lingkungan yang dapat merubah struktur DNA. Perubahan DNA tersebut merugikan proses pembelahan sel dan menguntungkan proses mutasi. Pada proses mutasi dapat menghasilkan gen p53, perubahan genetik tersebut paling umum ditemukan pada kanker manusia. Dari permasalahan tersebut dibutuhkan sistem untuk mengklasifikasikan jenis kanker yang diderita oleh pasien. Salah satu metode yang digunakan adalah *Naive Bayes*. *Naive Bayes* merupakan sebuah pengklasifikasian probabilitas sederhana yang mengaplikasikan Teorema Bayes dengan asumsi ketidaktergantungan yang tinggi. Algoritma tersebut diketahui telah banyak digunakan dalam bidang kedokteran. Algoritma ini diterapkan pada hal-hal yang berhubungan dengan diagnosa medis. Diagnosa dilakukan dengan cara melihat gejala-gejala yang berkaitan kemudian melihat probabilitas kemungkinan dari penyakit. Pengujian dilakukan dengan menggunakan 5 dataset yaitu 320, 400, 480, 588 dan 848 data dari data total sebanyak 848 data. Data dibagi menjadi data latih dan data uji. Data uji diambil 10% hingga 60% dari dataset. Hasil akurasi yang didapatkan pada pengujian 848 data dengan persentase data uji 60% didapatkan akurasi sebesar 79,17%.

**Kata kunci:** kanker, struktur protein, naive bayes

### Abstract

*Cancer is some cells that begins to divide without stopping and spreading to the surrounding tissue. Every year there are thousands of new cases of cancer affecting Indonesians. The late detection caused many deaths from cancer. Cancer-causing factors are genetic and environmental factors that can alter the structure of DNA. DNA changes are detrimental to the process of cell division and beneficial to the mutation process. In the mutation process can produce p53 gene, the genetic changes are most commonly found in human cancers. From these problems required a system to classify the types of cancer suffered by patients. One of the methods used is Naive Bayes. Naive Bayes is a simple probability classification that applies Bayes Theorem with the assumption of high dependence. The algorithm is known to have been widely used in the field of medicine. This algorithm applies to matters relating to medical diagnosis. Diagnosis is done by looking at the related symptoms then looking at the probability of the possibility of the disease. Tests were conducted using 5 datasets ie 320, 400, 480, 588 and 848 data from a total of 848 data. Data is divided into training data and test data. The test data is taken 10% to 60% of the dataset. Accuracy results obtained in the test 848 data with the percentage of 60% test data obtained accuracy of 79.17%.*

**Keywords:** cancer, protein structure, naive bayes

### 1. PENDAHULUAN

Kanker adalah beberapa sel tubuh yang mulai membelah tanpa berhenti dan menyebar ke jaringan sekitarnya. Kanker bisa bermula di

manapun dalam tubuh manusia, karena tubuh manusia terdiri dari triliunan sel aktif. Penyakit kanker telah ditemukan sejak 1,7 juta tahun lalu dan sampai saat ini masih menyerang manusia tanpa diketahui penyebab pastinya. Setiap tahun

terdapat ribuan kasus baru kanker yang menyerang warga Indonesia. Jenis kanker yang sering menyerang penduduk Indonesia antara lain kanker paru, perut, hati, usus besar dan kanker payudara. WHO telah memprediksi bahwa jumlah penderita kanker 20 tahun mendatang akan mencapai 24 juta penderita (WHO, 2014).

Penyebab penyakit kanker tidak dapat diungkapkan secara pasti. Menurut Infodatin Kemenkes RI faktor penyebab kanker antara lain genetik, karsinogen dan gaya hidup. Ketiga faktor tersebut dapat menyebabkan perubahan genetik. Perubahan genetik terjadi saat berubahnya struktur DNA disebabkan oleh perubahan sel. Perubahan sel dapat disebabkan dari paparan sinar UV, sinar X dan bahan-bahan kimia. DNA dapat mengikat faktor-faktor tersebut dan strukturnya pun berubah. Perubahan tersebut merugikan pembelahan sel, sebaliknya menguntungkan proses mutasi (Osterath, 2014). Pada proses mutasi dapat menghasilkan gen p53, perubahan genetik tersebut paling umum ditemukan pada kanker manusia.

Teknologi dapat membantu manusia untuk meringankan beban mereka dengan cara menggunakannya berdampingan dengan ilmu pengetahuan. Dari permasalahan tersebut dibutuhkan alat untuk mengklasifikasikan jenis kanker yang diderita oleh pasien. Salah satu metode yang digunakan adalah *Naive Bayes*. *Naive Bayes* adalah pengelompokan probabilistik sederhana berdasarkan penerapan Teorema *Bayes* dengan asumsi independensi yang kuat (naif) antara fitur. Algoritma tersebut diketahui telah banyak digunakan dalam bidang kedokteran. Algoritma ini diterapkan pada hal-hal yang berhubungan dengan diagnosa medis. Diagnosa dilakukan dengan cara melihat gejala-gejala yang berkaitan kemudian melihat probabilitas kemungkinan dari penyakit.

Penyebab terbesar kematian penduduk dunia salah satunya adalah karena kanker. Menurut WHO, pada tahun 2012 sendiri terdapat 8,2 juta kasus kematian di dunia yang disebabkan oleh kanker. Keterlambatan diagnosa dan mahalnya biaya penyembuhan adalah dua dari banyak faktor kematian akibat kanker. Karena faktor tersebut, seharusnya kanker dideteksi secara dini untuk mempermudah dan menghemat biaya proses penyembuhan. Untuk membantu pendeteksian dini kanker dibutuhkan sistem yang dapat

mendeteksi apakah pasien tersebut terkena kanker atau tidak, dan jika iya apa jenis dari kanker yang diderita oleh pasien. Seperti yang telah diutarakan pada paragraf dua dan tiga bahwa pada proses mutasi dapat menghasilkan gen p53 dan *Naive Bayes* telah banyak digunakan pada bidang kesehatan, maka dibutuhkan sistem yang bisa melakukan klasifikasi penyakit kanker berdasarkan struktur protein.

Algoritma *Naive Bayes* pernah digunakan pada penelitian sebelumnya untuk mendiagnosis penyakit kanker payudara, memiliki tingkat keakuratan sebesar 96,14% (Hermawati, 2015). Penelitian dengan menggunakan data yang sama juga pernah dilakukan dengan menggunakan algoritma yang berbeda yaitu MKNN, memiliki akurasi sebesar 45,53% (Retwitasari, 2016).

Berdasarkan latar belakang dan penelitian sebelumnya, maka judul yang bisa diambil adalah **“Klasifikasi Jenis Kanker Berdasarkan Struktur Protein Menggunakan Algoritma *Naive Bayes*”**.

## 2. KANKER

Kanker bermula ketika terdapat sel yang tidak normal membelah dengan tidak terkontrol. Kanker dapat terbentuk karena berubahnya suatu gen, Karena sifat sel yang terus membelah tanpa henti yang seharusnya membentuk sel baru untuk mengganti sel lama, sebaliknya membuat sel yang tidak normal terus membelah dan terus tumbuh. Sel kanker terkadang dapat menyebar dan tumbuh ke bagian tubuh lain. Penyebaran sel kanker ini biasa disebut *secondary tumor* atau metastasis (Cancer Research UK). Penamaan sel kanker dapat dilihat dari awal mula sel berasal dan dimana letak sel tersebut berkembang (Kelvin dkk., 2011).

Faktor genetik, karsinogen dan gaya hidup adalah beberapa faktor penyebab kanker (PUSDATIN Kementrian Kesehatan RI, 2015). Faktor lainnya adalah terdapatnya gen-gen yang berfungsi pada proses pertumbuhan tumor atau pada saat proses siklus sel. Pada proses siklus sel terdapat 2 macam gen, yaitu gen pemicu terjadinya tumor dan gen penekan terjadinya tumor. Gen pemicu terjadinya tumor atau dengan nama medis *tumor oncogenes*, contohnya adalah gen *c-myc* dan *ras*. Sedangkan gen penekan terjadinya tumor atau dengan nama medis *tumor suppressor*, contohnya adalah p53 dan Rb (Prayitno dkk., 2005). Disamping faktor diatas,

faktor lain adalah keturunan dari keluarga meskipun sangat jarang ditemukan.

Sel kanker dapat hilang 100% dari tubuh seseorang meskipun sangat kecil persentasenya, hal tersebut terbukti karena adanya pasien yang sembuh dari penyakit ini. Tetapi juga terdapat kemungkinan sel kanker akan tumbuh lagi di bagian tubuh berbeda ataupun dibagian tubuh yang sama. Cara untuk melawan kanker adalah dengan mengontrol dan mengendalikan pertumbuhan sel tersebut (Cancer Research UK). Pengobatan yang sering digunakan adalah kemoterapi. Pengobatan-pengobatan tersebut dapat mempengaruhi sirkulasi darah, sistem imun, sistem hormon dan sistem limpa. Pada pasien yang menderita kanker, akan diberikan tingkatan berdasarkan seberapa besar dan cepat sel kanker tumbuh, yang biasa disebut dengan stadium.

### 3. PROTEIN

Protein tersusun atas satu atau lebih rantai polipeptida yang terbentuk sebagai benang panjang asam-asam amino yang dihubungkan satu sama lain dengan ikatan-ikatan peptida dalam susunan linier yang spesifik. Terdapat 20 asam amino yang berbeda-beda. Rantai-rantai peptida tertentu memiliki urutan 100-150 asam-asam amino, dan sebagaimana DNA jumlah kemungkinan strukturnya sangat besar (Harris, 1994). Terdapat 4 struktur protein, antara lain (Marks dkk., 2000):

#### 1. Struktur Primer

Struktur protein ini terdiri dari urutan linear asam amino pada rantai polipeptida. Masih belum terjadi percabangan pada struktur ini.

#### 2. Struktur Sekunder

Struktur ini terdiri dari daerah-daerah lokal rantai polipeptida yang memiliki konformasi reguler yang distabilkan oleh ikatan hidrogen.

#### 3. Struktur Tersier

Konformasi tiga-dimensi total dari keseluruhan suatu rantai polipeptida yang mencakup heliks- $\alpha$ , lembar- $\beta$ , dan daerah berbentuk globular (sferis).

#### 4. Struktur Kuartener

Konformasi tiga-dimensi suatu protein multisubunit yang terdiri dari sejumlah rantai polipeptida (subunit) disatukan oleh interaksi nonkovalen.

### 3.1. Gen P53

P53, disebut juga sebagai TP53 atau protein tumor (EC: 2.7.1.37) adalah gen yang

mengkodekan protein yang mengatur siklus sel dan karenanya berfungsi sebagai penekan tumor. Hal ini sangat penting bagi sel-sel dalam organisme multiseluler untuk menekan kanker. P53 telah digambarkan sebagai “penjaga genom”, mengacu pada perannya dalam melestarikan stabilitas dengan mencegah mutasi genom (Strachan and Read, 1999). Nama ini disebabkan oleh massa molekulernya: dalam fraksi 53 kilodaton (kDa) dari sel protein.

P53 diidentifikasi pada tahun 1979 oleh Arnold Levine, David Lane dan William Old, masing-masing bekerja di Universitas Princeton, Universitas Dundee (Inggris) dan Rumah Sakit Sloan-Kettering Memorial. P53 telah dihipotesiskan sebelumnya sebagai target virus SV40, strain yang menginduksi perkembangan tumor. Meskipun pada awalnya dianggap sebagai onkogen, karakternya sebagai gen supresor tumor terungkap pada tahun 1989.

Protein p53 adalah phosphoprotein yang tersusun dari 393 asam amino. Terdiri dari empat unit atau domain:

1. Domain yang mengaktifkan faktor transkripsi.
2. Domain yang mengenali urutan DNA tertentu (domain inti).
3. Domain yang bertanggung jawab atas tetramerisasi protein.
4. Domain yang mengenali DNA yang rusak, seperti pasangan basa yang tidak sejajar atau DNA beruntai tunggal.

P53 tipe wild adalah protein labil, terdiri dari daerah terlipat dan tidak terstruktur yang berfungsi secara sinergis (Bell et al, 2002).

### 3.2. Kode Genetik Protein

Ide dasar yang menghubungkan struktur DNA dengan struktur protein ialah bahwa urutan pasangan-pasangan basa pada suatu gen menentukan urutan asam-asam amino pada rantai polipeptida yang sesuai. Oleh karena itu struktur-struktur dan karenanya sifat-sifat semua enzim dan protein yang dapat dibuat oleh satu individu dianggap ditentukan oleh urutan pasangan basa gen-gennya (Harris, 1994).

Protein dihasilkan dari proses translasi. Proses ini terjadi ketika urutan nukleotida dari hasil transkripsi DNA menjadi mRNA, menjadi urutan asam amino pada protein. Setiap asam amino dikodekan oleh rangkaian 3 nukleotida yang disebut kodon. Penentuan awal dari rangkaian protein ditandai dengan kodon *start*

(AUG) dan diakhiri dengan kodon stop (UAA, UAG atau UGA) (Marks dkk., 2000).

Mutasi pada urutan DNA gen dapat mengubah urutan asam amino dari protein yang dikode oleh gen. Perubahan satu basa (*point mutation*) dapat berupa transisi atau transversi (Kurnianti, 2013). Transisi merupakan perubahan basa purin/pirimidin menjadi basa purin/pirimidin lainnya, sedangkan transversi berupa perubahan basa purin menjadi basa pirimidin dan sebaliknya (Prayuni, 2008).

#### 4. NAIVE BAYES

Bayes merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema Bayes (atau aturan Bayes) dengan asumsi independensi (ketidaktergantungan) yang kuat (naif). Dengan kata lain Naive Bayes model yang digunakan adalah “model fitur independen” (Prasetyo, 2012). Independensi yang kuat pada fitur yang dimaksudkan pada *Naive Bayes* adalah fitur pada sebuah data tidak berkaitan dengan ada atau tidaknya fitur lain dalam data yang sama (Prasetyo, 2012).

Prediksi Bayes didasarkan pada teorema Bayes dengan formula umum sebagai berikut (Prasetyo, 2012):

$$P(H | E) = \frac{P(E | H) \times P(H)}{P(E)} \dots\dots\dots(1)$$

Ide dasar dari aturan Bayes adalah bahwa hasil dan hipotesis atau peristiwa (H) dapat diperkirakan berdasarkan pada beberapa bukti (E) yang diamati. Ada beberapa hal penting dan aturan Bayes tersebut, yaitu (Prasetyo, 2012):

1. Sebuah probabilitas awal/priori H atau P(H) adalah probabilitas dari suatu hipotesis sebelum bukti diamati.
2. Sebuah probabilitas akhir H atau P(H|E) adalah probabilitas dari suatu hipotesis setelah bukti diamati.

Naive Bayes berkaitan dengan klasifikasi, korelasi hipotesis, dan bukti dengan klasifikasi. Hipotesis dalam teorema Bayes merupakan label kelas yang menjadi masukan dalam model klasifikasi. Persamaan Naive Bayes untuk klasifikasi adalah (Prasetyo, 2012):

$$P(Y | X) = \frac{P(Y) \prod_{i=1}^q P(X_i | Y)}{P(X)} \dots\dots\dots(2)$$

P(Y|X) : probabilitas data dengan vektor X pada kelas Y. P(Y) adalah probabilitas awal kelas Y.

$\prod_{i=1}^q P(X_i | Y)$  adalah probabilitas independen kelas Y dari semua fitur dalam vektor X. Nilai P(X) selalu tetap sehingga dalam perhitungan probabilitas akhir nantinya kita tinggal menghitung bagian  $P(Y) \prod_{i=1}^q P(X_i | Y)$  dengan memilih nilai terbesar sebagai kelas yang dipilih sebagai hasil prediksi. Sementara probabilitas independen  $\prod_{i=1}^q P(X_i | Y)$  tersebut merupakan pengaruh semua fitur dari data terhadap setiap kelas Y, yang dinotasikan dengan

$$P(X | Y = Y) = \prod_{i=1}^q P(X_i | Y = y) \dots\dots\dots(3)$$

Setiap set fitur  $X = \{X_1, X_2, X_3, \dots, X_q\}$  terdiri atas q atribut (q dimensi) (Prasetyo, 2012).

#### 5. METODE PENELITIAN

Data didapatkan dari situs [www.uniprot.org](http://www.uniprot.org) dan [p53.fr](http://p53.fr) yang berupa dataset sekuen protein *human TP53 isoform 1* yang merupakan gabungan 20 residu asam amino berupa gabungan string sepanjang 393. Data terdiri dari 4 kelas yang terdiri dari kelas *non-cancer*, *breast cancer*, *colorectal cancer* dan *lung cancer*. Jumlah seluruh data adalah 848 dataset. Langkah-langkah untuk melakukan klasifikasi dengan menggunakan algoritma naive bayes adalah sebagai berikut:

1. Masukan data latih yang berupa dataset protein. Setiap deret protein merupakan variabel yang menentukan apakah pasien menderita kanker atau tidak.
2. Hitung probabilitas fitur atau likelihood pada data uji. Likelihood didapatkan dari menghitung probabilitas fitur pada setiap kelas. Probabilitas fitur biasa ditulis dengan  $P(X_i|Y)$ . Persamaan dapat dilihat pada persamaan (3).
3. Hitung probabilitas awal kelas atau prior. Prior didapatkan dari menghitung probabilitas suatu kelas terhadap semua kelas yang ada.



4. Setelah probabilitas awal kelas dan probabilitas fitur ditemukan, kemudian hitung probabilitas akhir atau posterior pada tiap kelas. Posterior didapatkan dengan menggunakan persamaan (2).
5. Nilai posterior tertinggi ditetapkan sebagai penentuan hasil dari klasifikasi.

## 6. SKENARIO DAN HASIL

### 6.1. Uji Pengaruh Data Uji Terhadap Jumlah Data

Pengujian sistem ini dilakukan untuk mengetahui pengaruh data latih dan data uji terhadap tingkat akurasi. Pengujian dilakukan dengan menyamakan jumlah dataset pada setiap kelas. Dataset per kelas adalah 320, 400, 480, 588 dan 848 data dan jumlah seluruh data adalah 848 data. Dari dataset tersebut dibagi menjadi data latih dan data uji. Data uji diambil dengan persentase 10%, 20%, 30%, 40%, 50% dan 60%. Misalkan pada dataset 320 diambil data uji sebanyak 10% (32 data), maka jumlah data latih adalah 90% (288 data).

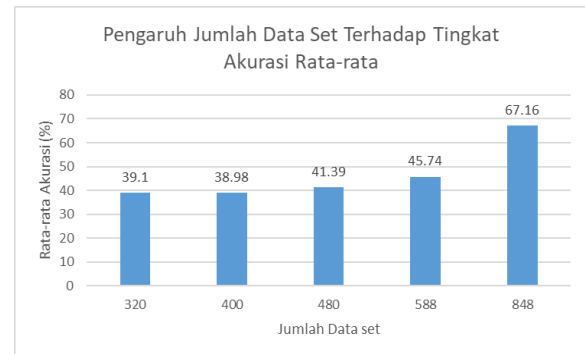
Pada pengujian persentase data uji terhadap jumlah data didapatkan akurasi yang variatif. Jumlah data uji yang menghasilkan akurasi terbaik adalah pada jumlah dataset 848. Pada jumlah data sebanyak 588 dataset didapatkan akurasi minimum sebesar 25% pada persentase data uji 10%. Akurasi maksimum sebesar 79,17% didapatkan pada persentase data uji 60% dan dengan jumlah 848 data.

Dari Tabel 1 dapat disimpulkan bahwa semakin tinggi persentase data uji belum tentu mendapatkan akurasi yang tinggi. Pengambilan dan pembagian data juga dapat berpengaruh pada tinggi dan rendahnya akurasi yang didapat. Menyeimbangkan jumlah data pada data pada dataset 320, 400, 480 dan 588 dapat menyebabkan rendahnya tingkat akurasi. Sedangkan ketika semua data digunakan pada dataset 848 didapatkan tingkat akurasi yang cukup tinggi dibandingkan dengan dataset lainnya

**Tabel 1** Pengaruh Data Uji Terhadap Jumlah Data

Persentase Data Uji	Akurasi per Jumlah Dataset (%)				
	320	400	480	588	848
10 %	28,13	30	27,09	25	47,68
20 %	48,44	46,25	50	51,67	59,83
30 %	52,09	53,34	56,25	61,37	67,4
40 %	37,5	35,63	41,15	48,31	72,57
50 %	35,63	34,5	36,67	44,87	76,32
60 %	32,82	34,17	37,16	43,19	79,17

### 6.2. Uji Pengaruh Jumlah Dataset Terhadap Tingkat Akurasi



**Gambar 1** Pengaruh Dataset Terhadap Tingkat Akurasi

Hasil dari pengujian dataset didapat rata-rata akurasi pada dataset sebanyak 320 adalah 39,1%. Pada dataset sebanyak 400 data didapat rata-rata akurasi sebesar 38,98%. Rata-rata akurasi yang didapat pada dataset sebanyak 480 data adalah 41,39%. Pada dataset sebanyak 588 data didapat rata-rata akurasi sebesar 45,74%. Rata-rata akurasi terbaik didapatkan pada dataset 848 yaitu sebesar 67,16%. Grafik hasil pengaruh jumlah dataset terhadap tingkat akurasi dapat dilihat pada Gambar 1.

Dari Gambar 1 dapat dilihat bahwa semakin banyak dataset maka semakin tinggi rata-rata akurasi yang didapat. Tetapi pada dataset 400 hasil yang didapat sebaliknya, rata-rata akurasi yang didapat lebih rendah dari pada dataset 320 meskipun selisihnya sedikit. Jadi semakin banyak jumlah dataset belum tentu rata-rata akurasi yang didapatkan semakin tinggi. Faktor yang mempengaruhi tinggi rendahnya akurasi adalah sebaran data pada dataset.

## 7. KESIMPULAN

Berdasarkan penelitian yang dilakukan didapatkan kesimpulan sebagai berikut:

1. Tahapan klasifikasi kanker berdasarkan struktur protein yaitu menghitung probabilitas fitur (likelihood), menghitung prior, dan menghitung probabilitas akhir (posterior). Probabilitas akhir didapatkan dari kelas yang mempunyai nilai probabilitas akhir tertinggi.
2. Nilai akurasi terkecil yang dihasilkan sistem adalah 25% pada pengujian 588 dataset pada persentase data uji 10% dari seluruh dataset. Sedangkan nilai akurasi terbesar yang dihasilkan sistem adalah

79,17% pada pengujian 848 dataset pada persentase data uji 60%. Nilai rata-rata akurasi terkecil yang dihasilkan sistem adalah 38,98% pada pengujian 400 dataset. Rata-rata akurasi terbesar yang dihasilkan sistem adalah 67,16% pada pengujian 848 dataset.

## 8. DAFTAR PUSTAKA

- Adelia, F. 2013. Penentuan Potensi Tsunami Akibat Gempa Bumi Bawah Laut Dengan Metode Modified K-Nearest Neighbor (MKNN). Universitas Brawijaya, Malang
- Harahap, Ikhsanudin Ahmad, 2004. Perawatan Pasien Dengan Kolostomi Pada Penderita Cancer Colorectal. Universitas Sumatera Utara, Medan
- Harris, Harry., 1994. Dasar-dasar Genetika Biokemis edisi ketiga. Gadjah Mada University Press, Yogyakarta
- Hermawati, Laily., 2015. Penggabungan Algoritme Backward Elimination Dan Naive Bayes Untuk Mendiagnosis Penyakit Kanker Payudara. Universitas Wahid Hasyim Semarang, Semarang
- Kelvin, Joanne Frankel dkk., 2011. 100 Tanya-Jawab mengenai Gejala Kanker dan Efek Samping Pengobatan Kanker Edisi Kedua. PT. Indeks, Jakarta Barat
- King, Roger J.B., 2000. Cancer Biology 2<sup>nd</sup> Edition. Pearson Education Limited, Inggris
- Kurnianti, Ria., 2013. Penggunaan Metode Pengelompokan K-Means Pada Klasifikasi KNN Untuk Penentuan Jenis Kanker Berdasarkan Susunan Protein, Universitas Brawijaya, Malang
- Marks, Dawn B. dkk., 2000. Biokimia Kedokteran Dasar: Sebuah Pendekatan klinis. EGC, Jakarta
- Mulyana, Sri., 2013. Penerapan Hidden Markov Model Dalam Clustering Sequence Protein Globin. Universitas Brawijaya, Malang
- Osterath, B., 2014. Beginilah Terjadinya Kanker. [online] Tersedia di: <<http://www.dw.com/id/beginilah-terjadinya-kanker/a-17407704>> [Diakses 3 September 2017]
- Prasetyo, Eko., 2012. Data Mining : Konsep dan Aplikasi menggunakan MATLAB. C.V ANDI OFFSET, Yogyakarta
- Prayitno, Adi dkk., 2005. Ekspresi Protein p53, Rb, dan c-myc pada Kanker Serviks Uteri dengan Pengecatan Immunohistokimia. Universitas Sebelas Maret, Surakarta
- Prayuni, Kinasih., 2008. Isolasi DNA Genom Padi (*Oryza Sativa L.*) Kultiviar, Rojolele, Nipponbare, dan Batutegi. Universitas Indonesia, Depok
- PUSDATIN Kementerian Kesehatan RI., 2015. InfoDATIN Pusat Data dan Informasi Kementerian Kesehatan RI: Situasi Penyakit Kanker. Kementerian Kesehatan RI, Jakarta Selatan
- Rahman, Dila., 2012, <<http://katakatanpabahasa.blogspot.com/2012/07/kode-genetik-sandi-3-huruf-yang.html>>
- S, Christine N.S, 2010. Hubungan Merokok Dengan Kanker Paru di RSUP Haji Adam Malik Tahun 2009. Universitas Sumatera Utara, Medan
- Strachan, T. dan Read, Andrew P., 1999. Human Molecular Genetics 2 2nd Edition. Wiley-Liss, New York
- Sudarka, Ir. Wayan., 2009. Pemuliaan Kelainan Genetik dan Sitogenesis Pada Tanaman. Universitas Udayana, Bali
- Warianto, Chaidar., 2011. Mutasi. Universitas Airlangga, Surabaya