CLUSTERING DATA *MICRORNA* KANKER PAYUDARA MENGGUNAKAN METODE *K-MEANS*

SKRIPSI

Disusun Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Komputer
pada Departemen Ilmu Komputer/Informatika

Disusun oleh:
DICKY FABRO SAHARA
24010313120057

DEPARTEMEN ILMU KOMPUTER/INFORMATIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO
2019

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan di bawah ini:

Nama : Dicky Fabro Sahara

NIM : 24010313120057

Judul : *Clustering* Data *MicroRNA* Kanker Payudara menggunakan Metode *K-Means*

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan di dalam daftar pustaka.

Semarang, 24 September 2019

Dicky Fabro Sahara
24010313120057

HALAMAN PENGESAHAN

Judul : *Clustering* Data *MicroRNA* Kanker Payudara menggunakan Metode *K-Means*

Nama : Dicky Fabro Sahara

NIM : 24010313120057

Telah diujikan pada sidang skripsi dan dinyatakan lulus pada tanggal 24 September 2019

Semarang, 27 September 2019

Mengetahui,
Ketua Departemen/Ilmu Komputer/Informatika
FSM UNDIP

Dr. Retno Kusumaningrum, S.Si., M.Kom.
NIP. 19810[illegible]2005012001

Panitia Penguji Skripsi,
Ketua

Drs. Suhartono, M.Kom.
NIP. 19550407198303100[illegible]

HALAMAN PENGESAHAN

Judul : *Clustering* Data *MicroRNA* Kanker Payudara menggunakan Metode *K-Means*

Nama : Dicky Fabro Sahara

NIM : 24010313120057

Telah diujikan pada sidang skripsi tanggal 24 September 2019

Semarang, 27 September 2019

Pembimbing,

Dr. Eng. Adi Wibowo, S.Si., M.Kom.

NIP. 198203092006041002

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI SKRIPSI UNTUK KEPENTINGAN AKADEMIS

Sebagai *civitas* akademik Universitas Diponegoro, saya yang bertanda tangan di bawah ini:

Nama : Dicky Fabro Sahara

NIM : 24010313120057

Program Studi : Informatika

Departemen : Ilmu Komputer/Informatika

Fakultas : Sains dan Matematika

Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan **Hak Bebas Royalti Non-eksklusif** ***(Non-exclusive Royalty Free Right)*** kepada Universitas Diponegoro atas karya ilmiah saya yang berjudul:

Clustering Data MicroRNA Kanker Payudara menggunakan Metode K-Means

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-eksklusif ini, Universitas Diponegoro berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Skripsi saya tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Semarang, 24 September 2019

Yang Menyatakan

Dicky Fabro Sahara

24010313120057

ABSTRAK

Kanker payudara merupakan penyakit yang disebabkan oleh pertumbuhan sel yang abnormal dimana sel tersebut menyerang sel normal di jaringan sekitarnya. Peranan *MicroRNA* pada kanker telah menjadi perkembangan penting dalam studi kanker sejak penemuannya pada tahun 2002. *MicroRNA* merupakan indikator pengukuran dari tingkat keberadaan suatu kanker serta mempunyai peranan penting dalam deteksi dan diagnosis khususnya kanker payudara. Pada penelitian ini, dilakukakan *clustering* data *MicroRNA* kanker payudara menggunakan metode *K-Means* dan mengetahui jumlah *cluster* terbaik serta pola pengelompokkan yang dipengaruh fitur pada data *MicroRNA*. Dari hasil penelitian ini, perbandingan pengujian nilai $k = 2$, nilai $k = 3$, nilai $k = 4$, dan nilai $k = 5$ menunjukkan bahwa nilai $k = 2$ merupakan *cluster* terbaik dengan nilai *silhouette* 0.454. Pola pengelompokan data menggunakan metode K-Means dipengaruhi oleh fitur ekspresi: *hsa-mir-10b*, *hsa-mir-21*, *hsa-mir-125b-1*, *hsa-mir-125b-2*, *hsa-mir-145*, *hsa-mir-200a*, dan *hsa-mir-200b*.

Kata Kunci: Kanker Payudara, *MicroRNA, K-Means Clustering*

ABSTRACT

Breast cancer is a disease caused by abnormal cell growth where the cells attack normal cells in the surrounding tissue. The role of MicroRNA in cancer has been an important development in cancer studies since its discovery in 2002. MicroRNA is an indicator of measurement of the level of the presence of a cancer and has an important role in the detection and diagnosis, especially breast cancer. In this study, clustering of breast cancer MicroRNA data was performed using the K-Means method and knowing the best number of clusters and grouping patterns affected by features in the MicroRNA data. From the results of this study, the comparison of testing the value of k = 2, the value of k = 3, the value of k = 4, and the value of k = 5 indicate that the value of k = 2 is the best cluster with a silhouette value of 0.454. The pattern of data grouping using the K-Means method is affected by the expression features: hsa-mir-10b, hsa-mir-21, hsa-mir-125b-1, hsa-mir-125b-2, hsa-mir-145, hsa-mir-200a, and hsa-mir-200b.

Keywords: Breast Cancer, MicroRNA, K-Means Clustering

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas berkat dan rahmat-Nya sehingga penulis dapat menyelesaikan skripsi dengan berjudul "*Clustering* Data *MicroRNA* Kanker Payudara menggunakan Metode *K-Means*" dengan baik dan lancar.

Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana strata satu pada Departemen Ilmu Komputer/Informatika Fakultas Sains dan Matematika Universitas Diponegoro. Dalam penyusunan skripsi ini penulis banyak mendapat bimbingan, bantuan, dan dukungan dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, penulis menyampaikan terimakasih kepada:

1. Dr. Retno Kusumaningrum, S.Si., M.Kom., selaku Ketua Departemen Ilmu Komputer/Informatika;
2. Panji Wisnu Wirawan, S.T., M.T., selaku Koordinator Skripsi;
3. Dr. Eng. Adi Wibowo, S.Si., M.Kom., selaku Dosen Pembimbing Skripsi;
4. Semua pihak yang telah membantu kelancaran dalam penyusunan Skripsi, yang tidak dapat disebutkan satu persatu.

Pada penyusunan laporan ini penulis menyadari bahwa masih terdapat banyak kekurangan baik dari segi materi maupun segi penyampaian materi tersebut. Hal tersebut dikarenakan keterbatasan kemampuan dan pengetahuan dari penulis. Oleh karena itu, kritik dan saran yang bersifat membangun sangat penulis harapkan. Semoga laporan skripsi ini dapat bermanfaat bagi semua pihak.

Semarang, 24 September 2019

Dicky Fabro Sahara

DAFTAR ISI

Hal

DAFTAR GAMBAR

Hal

DAFTAR TABEL

Hal

DAFTAR LAMPIRAN

Hal

BAB I
PENDAHULUAN

Bab ini membahas latar belakang masalah, rumusan masalah, tujuan dan manfaat, serta ruang lingkup penelitian.

1.1. Latar Belakang

Kanker adalah pertumbuhan sel yang abnormal di mana sel tersebut menyerang sel normal di jaringan sekitarnya. Penyebab kanker paling umum adalah perubahan (mutasi) pada gen dalam sel. Kanker payudara adalah kanker yang menyerang jaringan payudara. Beberapa faktor risiko terjadinya kanker payudara, antara lain usia di atas 50 tahun, jenis kelamin perempuan, riwayat keluarga, keturunan, obesitas *post-menopause* dan terapi hormonal. Paparan terhadap estrogen terkait erat dengan etimologi kanker payudara. Estrogen merupakan hormon seks wanita yang dibutuhkan dalam sejumlah proses dalam tubuh. Mekanisme estrogen terhadap proses karsinogenik masih terlalu kompleks, namun terdapat beberapa bukti penelitian bahwa estrogen menyebabkan *proliferasi* sel payudara normal maupun ganas (Vogel, 2018). Mutasi pada gen *breast cancer 1* (BRCA1) dan *breast cancer 2* (BRCA2) bertanggung jawab terhadap 3-8% kasus kanker payudara. Kedua gen ini dipercaya berperan sebagai *tumor supressor genes* yang berperan dalam mempertahankan integritas DNA dan regulasi transkripsi DNA (Suparman&Suparman, 2014). Ekspresi gen pada sebuah sel dikendalikan secara akurat dan didasarkan pada kondisi fisiologis pada tubuh dimana mereka berada. *MicroRNA* berukuran kecil memainkan peran penting dalam sebagian besar pengaturan sel (Sankar, et al., 2017).

MicroRNA (*MiRNA*) terdiri atas *RNA* endogen *non-coding* berukuran kecil, yaitu panjangnya 22-24 *nukleotida*. *MiRNA* bertindak sebagai *tuner* yang baik dari regulator gen pada eukariotik uniseluler, hewan, tumbuhan dan manusia. Secara umum, regulasi gen di dalam sel seluruhnya didasarkan pada jenis interaksi *MiRNA*-m*RNA* dan interaksinya dari satu sel ke sel lain bervariasi berdasarkan kondisi fisiologis (Sankar, et al., 2017). *MicroRNA* (mi*RNA*) memiliki potensi besar sebagai *biomarker* tumor dan target terapi. Seiring perkembangan pesat dari teknologi eksperimental, penelitian ekspresi gen menjadi semakin terspesialisasi dan beragam. Struktur data yang

kompleks setelah membawa tantangan besar untuk identifikasi *biomarker* (Yang, et al., 2017). Dalam dunia kedokteran, *biomarker* adalah indikator yang dapat diukur dari tingkat keparahan atau keberadaan beberapa keadaan penyakit. Perubahan ekspresi *MiRNA*s dalam berbagai jenis *neoplasia* membentuk pola spesifik *MiRNA signature* untuk jenis kanker tertentu (Lan, et al., 2014). Perubahan ekspresi *MiRNA* pada kanker, menimbulkan dugaan bahwa *MiRNA* dapat berperan sebagai onkogen atau tumor supresor gen (Juwita, 2017). *MicroRNA* berperan sebagai modulator ekspresi gen sehingga dapat dijadikan sebagai kandidat diagnostik dan indikator *prognostik* serta target terapi yang potensial (Heneghan, et al., 2009). Oleh karena itu, *MiRNA* menjadi salah satu alternatif *biomarker* kanker yang sangat menjanjikan dimasa mendatang.

Clustering data microRNA adalah proses pengelompokkan data menjadi beberapa *cluster*. Pengelompokkan data tersebut berguna untuk mengidentifikasi pola data pada setiap *cluster*. Pembahasan mengenai *clustering* data MicroRNA pernah dilakukan di beberapa penelitian. Contoh penelitian tersebut yaitu penelitian yang dilakukan oleh (Yang, et al., 2017) yang menunjukkan bahwa hasil dari metode berbasis pengelompokan dapat mengidentifikasi *biomarker* kombinasi *MicroRNA* memiliki akurasi dan efisiensi tinggi.

Dalam penelitian digunakan algoritma *K-Means* untuk mengelompokkan data *MicroRNA* menjadi *k cluster* berdasarkan karakteristiknya, sehingga objek yang mempunyai karakteristik yang sama dikelompokan dalam satu *cluster* yang sama dan objek yang mempunyai karakteristik yang berbeda dikelompokan ke dalam *cluster* yang lain. Metode *K-Means* merupakan algoritma yang paling banyak digunakan karena secara konseptual metode k-means sederhana, dapat dimodifikasi, dan mudah diimplementasikan (Sopian, et al., 2016). Algoritma *K-Means* mencoba menemukan variabilitas yang berpengaruh pada jumlah *cluster*. Algoritma *K-Means* mengelompokkan *n* objek berdasarkan atribut menjadi *k* partisi, dimana $k < n$. *Nilai k* akan menentukan *cluster* pada data *MicroRNA* sedangkan *n* adalah sejumlah data *MicroRNA*.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka dapat dirumuskan permasalahan yaitu: Bagaimana membuat aplikasi *clustering* data *MicroRNA* dengan

menggunakan metode *K-Means* untuk mengetahui pengelompokan data dan analisis pola fitur yang berpengaruh?

1.3. Tujuan dan Manfaat

Tujuan dari penelitian ini adalah sebagai berikut:

1. Menghasilkan aplikasi untuk peng-*cluster*-an data penyakit kanker payudara menggunakan metode *K-Means Clustering*.
2. Menganalisis hasil kerja dari metode *K-Means* untuk *clustering* data penyakit kanker payudara berdasarkan jumlah *cluster*.
3. Menganalisis pola fitur yang berpengaruh pada data *MicroRNA* kanker payudara.

Manfaat dari aplikasi *Clustering Data MicroRNA Penyakit Kanker Payudara* berguna untuk mengelompokkan data *MicroRNA* sehingga dapat membantu menemukan pola pengelompokkan dalam *cluster*.

1.4. Ruang Lingkup

Ruang lingkup dalam menerapkan *clustering* data *MicroRNA* menggunakan metode *K-Means Clustering* adalah sebagai berikut:

1. Data berupa *dataset MicroRNA* yang mempunyai satuan integer.
2. Aplikasi dapat menampilkan pengelompokan data *MicroRNA* dengan metode *K-Means.*
3. Nilai *k* pada *K-Means* adalah 2, 3, 4, dan 5.
4. Bahasa pemrograman yang digunakan dalam pengembangkan aplikasi adalah *python*.

1.5. Sistematika Penulisan

Sistematika penulisan yang digunakan dalam penyusunan laporan skripsi ini terdiri dari 5 bab, yaitu Pendahuluan, Landasan Teori, Metodologi Penelitian, Implementasi dan Pengujian, serta Penutup.

BAB I PENDAHULUAN

Bab pendahuluan berisi latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup masalah, serta sistematika penulisan dalam pelaksanaan penelitian *Clustering* Data *MicroRNA* Kanker Payudara menggunakan Metode *K-Means*.

BAB II LANDASAN TEORI

Bab ini berisi tinjauan pustaka yang berhubungan dengan skripsi sebagai landasan untuk merumuskan dan menganalisa permasalahan pada skripsi ini. Tinjauan pustaka yang digunakan meliputi kanker payudara, *MicroRNA*, metode *K-Means Clustering*, metode pengembangan, permodelan analisis, metode pengujian, bahasa pemrograman, dan *tools* yang digunakan.

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan mengenai tahapan yang dilakukan dalam penelitian tugas akhir. Tahapan tersebut meliputi garis besar penyelesaian masalah, pengumpulan data, normalisasi data, Perhitungan *K-Means*, evaluasi sistem, deskripsi umum aplikasi dan analisa kebutuhan sistem.

BAB IV IMPLEMENTASI DAN PENGUJIAN

Bab ini membahas mengenai tahapan implementasi dan analisis dari skenario pengujian aplikasi maupun metode. Tahapan tersebut meliputi implementasi dari aplikasi, pengujian aplikasi, dan pengujian metode *K-Means Clustering*.

BAB V PENUTUP

Bab penutup berisi kesimpulan dari penelitian dan uraian yang telah dibahas pada bab sebelumnya dan saran untuk pengembangan sistem lebih lanjut.