Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN: 2548-964X
Vol. 3, No. 8, Agustus 2019, hlm. 7353-7360 http://j-ptiik.ub.ac.id

Klasifikasi Fungsi Senyawa Aktif Berdasarkan Notasi *Simplified Molecular Input Line Entry System* (SMILES) Dengan Metode *K-Means Naïve Bayes* (KMNB)

Revi Anistia Masykuroh[1], Dian Eka Ratnawati[2], Syaiful Anam[3]

[1,2]Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
[3]Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Brawijaya
Email: [1]revianistia@gmail.com, [2]dian ilkom@ub.ac.id, [3]syaiful@ub.ac.id,

Abstrak

Indonesia merupakan negara tropis yang memiliki keanekaragaman hayati (biodiversitas) paling banyak di dunia. Hampir seluruh bagian dari tumbuhan baik daun, akar, batang, buah, bunga, rimpang, dan biji dapat diambil manfaatnya untuk kesehatan. Namun pada kenyataannya, pemanfaatan tumbuhan sebagai obat di Indonesia masih sangat terbatas. Oleh karena itu, dibutuhkan penelitian lebih lanjut dan berkesinambungan tentang tumbuhan obat atau obat herbal serta teknologi pengolahan untuk dapat memaksimalkan pemanfaatannya. Pada tahun 1980, David Weininger menemukan suatu notasi kimia untuk memproses informasi-informasi yang berkaitan dengan kimia modern yang diberi nama *Simplified Molecular Input Line System* (SMILES) khusus untuk penggunaan komputer. Pada penelitian ini digunakan metode gabungan *K-Means* dan *Naïve Bayes* karena metode ini dianggap mampu mengelompokkan data sesuai dengan kemiripannya dan proses pengklasifikasiannya lebih mudah dipahami. Berdasarkan hasil pengujian, metode *K-Means Naïve Bayes* mampu memberikan nilai rata-rata akurasi sebesar 85,45% dengan rasio data latih 80% dan data uji 20%. Sistem diuji dengan menggunakan pengujian *K-Fold Cross Validation* dengan *K-Fold* sebanyak 10 yang menghasilkan akurasi tertinggi sebesar 86,66% pada *fold* ke-9 dan terendah sebesar 70,37% pada *fold* ke-1. Rata-rata dari akurasi pengujian menggunakan *K-Fold Cross Validation* sebesar 82,6%.

Kata kunci: Senyawa Aktif, SMILES, K-Means, Naïve Bayes

Abstract

Indonesia is a tropical country that has the most biodiversity in the world. Almost all of the plants part like leaf, root, stem, fruit, flowers, seeds, and rhizome can be used for human health. In Indonesia the utilization of plants as medicine is so limited. Therefore, further research and continuous plant drugs or herbal remedies is really needed as well as the technologies are able to maximize the utilization. In 1980, David Weininger found a chemical notation for processing informations that related to a modern chemistry named Simplified Molecular Input Line System (SMILES) and that notation is specifically for computer used. On this research, K-Means Naïve Bayes methods are used for the classification of the functions of the active compounds because this methods are able to grouping data according to their similarity and the classification process is much easier to understand. Based on the test results, the K-Means Naïve Bayes are abled to give an accuracy system 85.45% with a 80% training data ratio and 20% testing data. The system also being tested using K-Fold Cross Validation with K-Fold as many as 10, the highest accuracy that can be given is 86.66% on 9th fold and the lowest is 70.37% on 1st fold. While the average of accuracy using the K-Fold Cross Validation is 82.6%.

***Keywords**: Active compounds, SMILES, K-Means, Naïve Bayes*

1. PENDAHULUAN

Seiring dengan munculnya berbagai penyakit baru tentu menuntut peneliti khususnya di bidang farmakologi untuk terus mengembangan risetnya. Namun kenyataannya, tingginya biaya serta minimnya instrumen penelitian membuat para peneliti enggan sehingga menyebabkan kebutuhan dan ketersediaan obat masih mengandalkan impor

dari India dan Cina.

Seharusnya potensi obat di Indonesia tidak perlu diragukan mengingat negara kita merupakan negara tropis yang memiliki keanekaragaman hayati (biodiversitas) paling banyak di dunia. Hal ini membuktikan bahwa pemanfaatan tumbuhan sebagai obat di Indonesia masih sangat terbatas.

Tumbuhan dapat dikatakan sebagai tumbuhan obat apabila tumbuhan tersebut mengandung senyawa aktif. Senyawa aktif merupakan senyawa yang memiliki efek fisiologis yang berguna untuk menyembuhkan atau mencegah penyakit pada organisme yang lain yang bisa didapatkan dari tumbuhan atau hewan. Masih begitu banyak senyawa aktif yang belum ditemukan fungsinya dan masih dalam tahapan penelitian. Jika ingin mengetahui fungsi dari suatu senyawa aktif tersebut dibutuhkan penelitian di laboratorium yang membutuhkan waktu sangat lama karena dalam melakukan penelitiannya peneliti harus mengekstraksikan senyawa kimia terlebih dahulu agar dapat diketahui fungsinya.

Senyawa aktif memiliki suatu struktur molekul yang menyusun senyawa aktif itu sendiri. Struktur molekul suatu senyawa bisa berupa satu, dua, atau tiga dimensi. *Simplified Molecular Input Line System* (SMILES) merupakan representasi dari struktur molekul senyawa aktif dengan satu dimensi. SMILES merupakan suatu notasi kimia paling baru dimana komputer mampu membaca semua unsur penyusun dari sebuah senyawa kimia (Weininger, 1988).

Dengan perbandingan banyaknya manfaat serta beberapa kendala yang dialami maka diperlukan suatu sistem kIasifikasi senyawa aktif untuk memudahkan proses pengidentifikasian. Sebelumnya penelitian tentang SMILES dengan 11 fitur yang terdiri dari *Boron "B", Carbon "C", Nitrogen "N", Oksigen "O", Fosfor "P", Belerang* atau *Sulfur "S", Fluor "F", Klorin "Cl", Bromin "Br", Yodium "I",* dan *hidroksida "OH",* telah dilakukan dengan menggunakan *Learning Vector Quantization* (LVQ) dimana sistem mampu mendapatkan nilai akurasi terbaik sebesar 76,34% (Ramzini, S., Ratnawati, D. E., Anam, S., 2018). Penelitian tentang klasifikasi lainnya juga telah dilakukan oleh Lestari, P., I., Ratnawati, D., E., Muflikhah, L. (2018) tentang klasifikasi diagnosa penyakit kucing. Peneliti menggunakan metode gabungan *K-Means* serta *Naïve Bayes* dan didapatkan nilai rata-rata akurasi sistem sebesar 90% (Lestari, P., I., Ratnawati, D., E., Muflikhah, L. 2018).

Berdasarkan penjelasan tersebut, penulis akan melakukan penelitian dengan menggabungkan teknik pengklasteran *K-Means* dan klasifikasi *Naïve Bayes.* Metode *K-Means* digunakan karena pengimplementasiannya mudah dan mampu mengelompokkan data sesuai dengan kemiripannya. *K-Means* dapat mengurangi kompleksitas data sehingga lebih cepat dalam melakukan proses komputasi (Prasetyo, 2012). Setelah itu, metode *Naïve Bayes* digunakan dalam proses pengklasifikasian selanjutnya karena lebih mudah dipahami, pengodeannya sederhana, waktu komputasi yang lebih sedikit (Xhemali, D., Hinde, C. J., Stone, R. G., 2009). Selain itu, penelitian ini akan dilakukan dengan penambahan empat fitur yakni *@, #, =,* dan tanda buka tutup kurung *()*. Dengan menggunakan 15 fitur tersebut diharapkan sistem dapat memberikan hasiI akurasi yang Iebih tinggi daripada penelitian sebelumnya.

2. LANDASAN KEPUSTAKAAN

2.1 Senyawa Aktif

Senyawa dihasilkan melalui reaksi kimia dari dua unsur atau lebih dan memiliki ikatan. Senyawa kimia yang digunakan untuk obat merupakan senyawa aktif yang bisa didapat dari tumbuhan atau hewan dan mampu memberikan efek fisiologis pada organisme yang lain (Marisa, H., Mukti, R. W. & Salni. 2011)

2.2 SMILES

SMILES ditemukan pada tahun 1980 oleh David Weininger. SMILES tersusun dari atom-atom yang membentuk senyawa kimia. Atom penyusun tersebut bernama atom kimia organik. Atom kimia organik terdiri dari *Boron "B", Carbon "C", Nitrogen "N", Oksigen "O", Fosfor "P", Belerang* atau *Sulfur "S", Fluor "F", Klorin "Cl", Bromin "Br",* dan *Yodium "I"* (Weininger, 1988).

2.3 *Preprocessing*

Preprocessing adalah mengubah bentuk data dari yang awaInya tidak terstruuktur lalu di proses sedemikian rupa hingga menjadi bentuk data yang terstruktur (Manning, C., P. Raghavan, dan H. Schutze. 2009). Pada tahap

preprocessing notasi SMILES penggunaaan *regular expression* (regex) diterapkan. *Regular expression* merupakan suatu fungsi untuk mengetahui atau mencari pola dari sebuah kalimat. *Preprocessing* pada notasi SMILES didapatkan dengan mencari nilai dari masing-masing atom penyusunnya kemudian hasilnya dibagi dengan jumlah panjang SMILES. HasiI pembagian tersebut merupakan input dari proses klasifikasi menggunakan metode KMNB.

2.4 Klasifikasi

KIasifikasi merupakan fungsi data mining yang berfungsi untuk menargetkan kategori atau keIas (Kusnawi, 2007). Secara umum, proses kIasifikasi ini ada dua tahap yaitu tahap *training* (pelatihan) serta tahap *testing* (pengujian). Tahap peIatihan merupakan proses untuk meIakukan pembangunan model, Pembuatan suatu model kIasifikasi diawali dengan meIihat dari data latih itu sendiri. Data latih digunakan untuk mengenaIi suatu data baru yang akan diklasifikasikan ke keIas yang sudah disimpan sebeIumnya. Sedangkan tahap pengujian adalah proses untuk meIakukan uji modeI yang teIah diIakukan pada proses pelatihan, sehingga data yang sebelumnya tidak memiliki kelas dapat diketahui kelasnya.

2.5 *K-Means*

K-Means adalah metode data klastering yang mengelompokkan data berdasarkan kemiripannya menjadi satu klaster sedangkan data yang mempunyai karakteristik berbeda atau tidak sama maka dikelompokkan ke klaster lainnya (Agusta, 2007). Berikut merupakan langkah-langkah yang digunakan pada metode *K-Means*:

1) Menetukan jumlah klaster atau *k*
2) Inisialisasi nilai *k* secara random
3) Menghitung nilai kemiripan masing-masing data ke klaster dengan menggunakan Persamaan (1):

$$D_{ij} = \sqrt{\sum_{k=1}^{n}(x_{ik} - y_{jk})^2} \quad (1)$$

Keterangan:
D_{ij} = jarak sebuah data i dengan data j
n = banyaknya jumlah dimensi
k = dimensi ke-k
X_{ik} = data i dimensi ke-k
C_{jk} = data j dimensi ke-k (merupakan data titik *centroid*)

4) Mencari nilai jarak jarak terdekat dari masing-masing data tiap klaster
5) Menghitung kembali nilai *centroid* dengan keanggotaan data sekarang dengan menghitung rata-rata data untuk tiap klaster
6) Ulangi langkah nomor 3-5 hingga keanggotaan data pada masing-masing klaster tidak berubah

2.6 *Naïve Bayes*

Metode *naïve bayes* yang diterapkan pada peneIitian ini menggunakan dua tahap, yaitu tahap peIatihan serta tahap pengujian. Langkah-langkah perhitungan *Naïve Bayes* pada tahap pelatihan dapat dilakukan dengan:

1) Melakukan perhitungan niIai rata-rata serta niIai standar deviasi atau simpangan baku tiap kelas
2) Melakukan perhitungan peluang dari tiap kelas menggunakan Persamaan (2):

$$\boldsymbol{P}\,(\boldsymbol{H} \mid \boldsymbol{X}) = \frac{\boldsymbol{P}\,(\boldsymbol{X} \mid \boldsymbol{H})\,\boldsymbol{P}\,(\boldsymbol{H})}{\boldsymbol{P}\,(\boldsymbol{X})} \quad (2)$$

Keterangan:
P(H) = prior probability
P(X) = peluang data sampel *X*
P(X|H) = posteriori probability

Pada tahap pengujian langkah-langkah perhitungan *Naïve Bayes* dapat dilakukan dengan:

1) Menghitung nilai distribusi gaussian menggunakan Persamaan (3):

$$\boldsymbol{P}\,(\boldsymbol{X_i} = \boldsymbol{x_i} | \boldsymbol{Y} = \boldsymbol{y_j}) = \frac{1}{\sqrt{2\pi\sigma^2_{ij}}}\boldsymbol{exp}^{-\frac{(x_i - \mu_{ij})^2}{2\sigma^2_{ij}}} \quad (3)$$

Keterangan:
$\boldsymbol{P}$ = peluang
$\boldsymbol{X_i}$ = atribut ke i
$\boldsymbol{Y}$ = kelas yang dicari
y_j = sub kelas Y yang dicari
μ = rata-rata dari seluruh atribut tiap kelas
σ = varian dari seluruh atribut tiap kelas

2) Menghitung *likelihood* yaitu dengan mengalikan seluruh nilai distribusi *gaussian*
3) Menghitung nilai *posterior* dengan Persamaan (4):

$$Posteriori = likelihood\ x\ prior/evidence \quad (4)$$

4) Setelah didapatkan nilai *posterior*, maka dicari nilai tertinggi diantara semua peluang dan klasifikasikan sesuai dengan kelas tersebut

2.7 Gabungan Metode *K-Means* dan *Naïve Bayes*

Metode gabungan klastering *K-Means* dan

klasifikasi *Naïve Bayes* dibentuk sesuai dengan langkah-langkah berikut (Asikin, M. F., Ratnawati, D. E., dan Fauzi, M. A., 2017):

1) Menentukan *dataset*
2) Pengelompokkan data dengan metode *K-Means* dengan klaster sebanyak jumlah kelas ditambah satu klaster *unknown*
3) Klaster yang masih tidak bisa ditentukan kelasnya (klaster *unknown*) akan dilakukan proses pengklasifikasian menggunakan metode *Naïve Bayes*

3. METODOLOGI

3.1 Deskripsi Umum Sistem

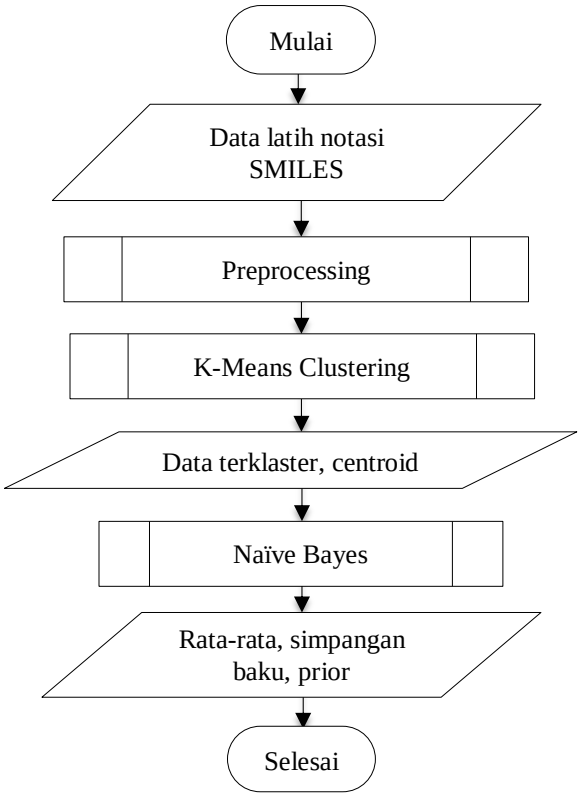
Penenlitian ini akan membangun sebuah sistem klasifikasi fungsi senyawa aktif dengan menggunakan notasi-notasi SMILES. Sistem diawali dengan peng*input*an notasi dan selanjutnya sistem akan menghitung nilai dari masing-masing fitur penyusunnya lalu dibagi panjang SMILES *(preprocessing)*. Sistem klasifikasi ini memiIiki dua tahap yaitu:

1) Tahap peIatihan
   Tahap peIatihan merupakan suatu proses untuk meIakukan pembangunan model menggunakan data latih. Data latih pada *K-Means* akan digunakan sebagai pembentuk nilai *centroid* dari masing-masing klaster yang telah ditentukan. Pada metode *Naïve Bayes* dilakukan perhitungan nilai rata-rata, simpangan baku, serta *prior probability* dari masing-masing kelas.
2) Tahap pengujian
   Tahap pengujian merupakan proses untuk meIakukan uji model. Pada tahap pengujian ini pengklasteran menggunakan *K-Means* dilakukan terlebih dahulu dengan mengambil nilai *centroid* dari tahap pelatihan sebelumnya. Selanjutnya nilai rata-rata, simpangan baku, serta *prior probability* yang telah didapatkan dari tahap pelatihan *Naïve Bayes* sebelumnya digunakan untuk perhitungan g*aussian*, *likelihood*, dan *posterior*. Tahap pengujian *Naïve Bayes* ini diIakukan dengan menggunakan data uji dari salah satu klaster hasil klasterisasi dari *K-Means*.

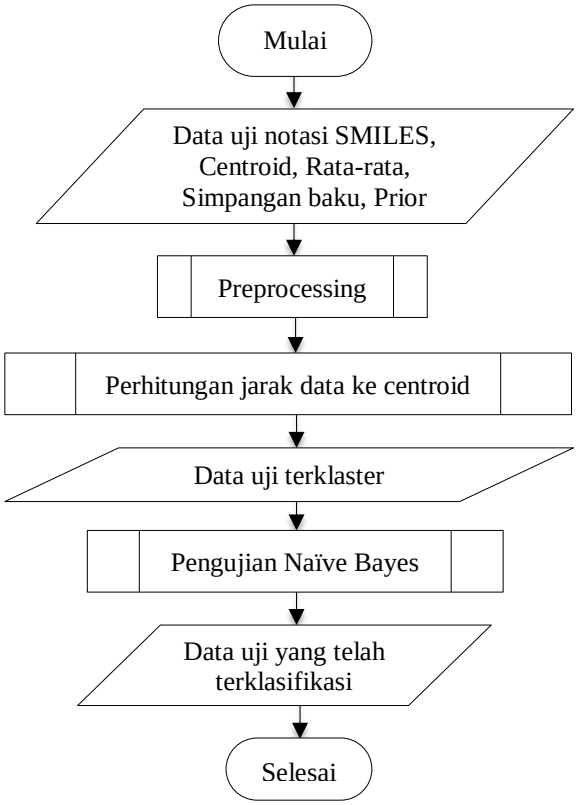
3.2 Perancangan Sistem

Sistem klasifikasi metode KMNB ini memiliki dua tahap yang pertama merupakan tahap pelatihan yang ditunjukkan Gambar 1 dan tahap pengujian yang ditunjukkan Gambar 2.

Tahap pelatihan diawali dengan *input* data latih. Setelah dilakukan perhitungan *preprocessing* langkah selanjutnya adalah proses pengelompokkan dengan metode *K-Means* sehingga didapatkan hasil datalatih terklaster dan nilai *centroid*. Setelah itu proses pelatihan *Naïve Bayes* diIakukan agar didapatkan nilai rata-rata, simpangan baku, dan *prior* yang digunakan untuk tahap pengujian selanjutnya.



Gambar 1. Diagram alir pelatihan sistem



Gambar 2. Diagram alir pengujian sistem

Sedangkan tahap pengujian diawali dengan proses *input* data uji, *centroid*, rata-rata, simpangan baku, dan prior. Setelah dilakukan perhitungan *preprocessing* langkah selanjutnya adalah proses perhitungan *euclidean distance* sehingga didapatkan hasil data uji terklaster. Setelah itu proses pengujian *naïve bayes* dilakukan sehingga sistem mampu menampilkan data uji yang telah terklasifikasi.

4. PENGUJIAN

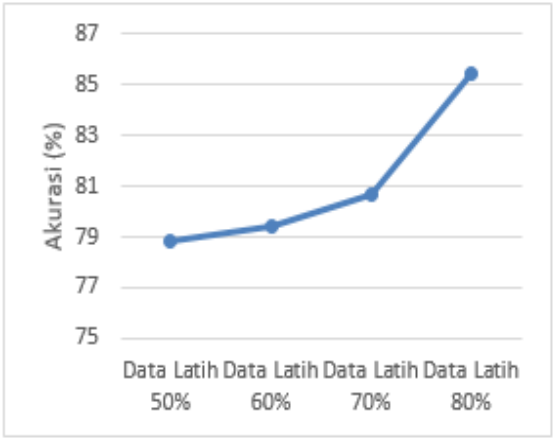
Seluruh pengujian yang akan dilakukan pada tahap ini menggunakan data sebanyak 818 yang terdiri dari tiga kelas yaitu: 302 data metabolisme, 402 data infeksi, dan 114 data anti-radang.

4.1 Pengujian Validasi Sistem

Hasil perhitungan manualisasi dan hasil klasifikasi dari sistem memiliki *output* yang sama yakni sebesar 90% sehingga dapat disimpuIkan jika sistem yang dibangun *valid*.

4.2 Pengujian Pengaruh Jumlah Data Latih dan Data Uji (*Hold Out Validation*)

Pengujian ini memiliki 4 skenario pengujian dengan rasio data Iatih dan data uji yaitu: 50%:50%, 60%:40%, 70%:30%, 80%:20%. Gambar 3 merupakan hasil dari pengujian *hold out validation*.

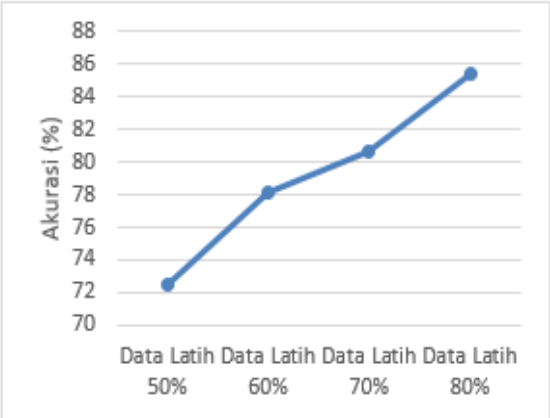


Gambar 3. Hasil pengujian *hold out validation*

Dari gambar tersebut dapat diketahui bahwa terdapat peningkatan rata-rata akurasi sistem pada rasio 50%:50% dan 60%:40%. Seiring dengan semakin banyaknya data Iatih maka nilai rata-rata akurasi juga semakin meningkat. NiIai rata-rata akurasi sistem terus meningkat hingga skenario pengujian ke-4 yaitu sebesar 85,45% dengan rasio data Iatih 80% dan data uji 20%.

4.3 Pengujian Pengaruh Jumlah Data Latih dan Data Uji Tetap

Pengujian ini diIakukan untuk memperkuat skenario pengujian *hold out validation* sebelumnya. Pada pengujian *hold out validation* nilai akurasi yang tertinggi didapat dengan menggunakan rasio data Iatih 80%dan data uji 20%. Maka dari itu, jika pada pengujian sebelumnya rasio data uji yang digunakan bervariasi maka pada pengujian kali ini rasio data uji yang digunakan tetap. Dengan adanya skenario ini dapat diketahui apakah benar jika rasio data Iatih semakin besar maka nilai akurasi yang diperoleh juga semakin meningkat. Gambar 4 merupakan hasiI dari pengujian jumlah data uji dan data uji tetap.

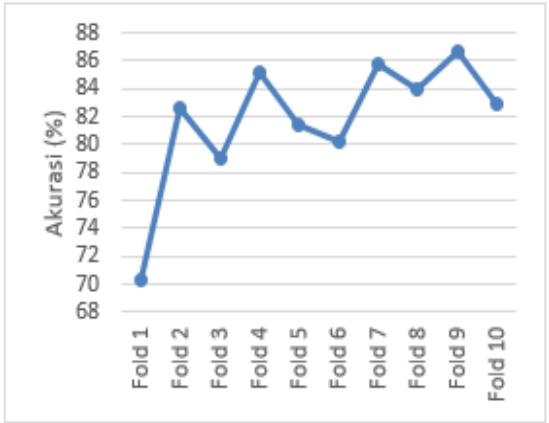


Gambar 4. Hasil pengujian jumlah data latih dan data uji tetap

Semakin bertambahnya data latih maka akurasi yang dapat diberikan sistem juga semakin meningkat. Kesimpulan dari pengujian ini adalah semakin banyak rasio data Iatih yang digunakan dalam proses pengklasifikasian menggunakan metode KMNB maka akurasi juga akan semakin meningkat karena sistem mampu melakukan pelatihan dengan data yang lebih bervariasi sehingga sistem mampu meminimalisir kesalahan saat melakukan penentuan kelas pada tahap klasifikasi.

4.4 Pengujian *K-Fold Cross Validation*

Pengujian selanjutnya diIakukan dengan membahi *dataset* sebanyak *k*. Pada tiap perulangan *k*, data yang digunakan akan selalu berubah fungsinya baik sebagai data Iatih atau sebagai data uji. Data yang digunakan pada peneIitian ini akan dibagi sebanyak sepuluh *fold*. Gambar 5 merupakan hasiI pengujian*K-Fold Cross Validation*.

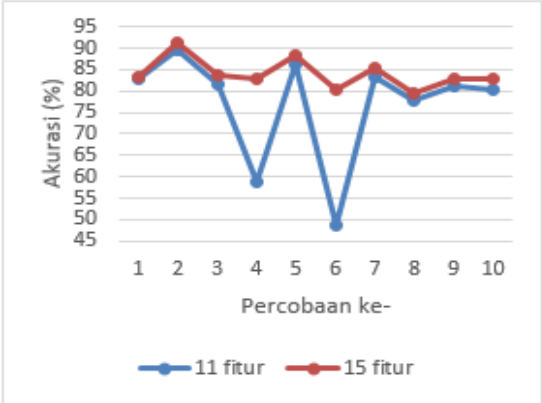


Gambar 5. Hasil pengujian *k-fold cross validation*

Pengujian ini menghasilkan nilai akurasi tertinggi sebesar 86,66% pada *foId* ke-9 dan terendah sebesar 70,37% pada *foId* ke-1. Sehingga dapat disimpulkan bahwa metode *K-Means Naïve Bayes* mampu mengklasifikasikan fungsi senyawa aktif yang telah dikategorikan dengan nilai rata-rata akurasi yang diberikan sistem sebesar 82,6%.

4.5 Pengujian jumlah fitur notasi SMILES

Pengujian jumIah fitur notasi SMILES diIakukan dengan membandingkan hasil akurasi 11 fitur dan 15 fitur yang diIakukan sebanyak 10 kaIi percobaan. 15 fitur yang digunakan terdiri dari atom *Boron "B", Carbon "C", Nitrogen "N", Oksigen "O", Fosfor "P", Belerang* atau *Sulfur "S", Fluor "F", Klorin "Cl", Bromin "Br", Yodium "I", hidroksida "OH", chirality "@",* rangkap dua *"=",* rangkap tiga *"#",* dan cabang *"()"* Gambar 6 merupakan hasiI pengujian jumlah fitur.

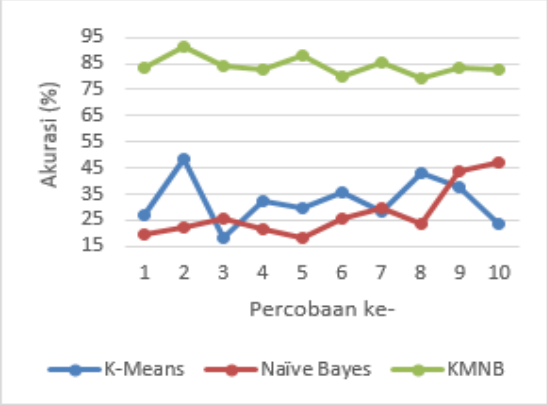


Gambar 6. Hasil pengujian jumlah fitur

Pada percobaan ke-4 dan ke-6 dapat diketahui bahwa dengan menggunakan 15 fitur akurasi bisa mencapai angka 80% keatas namun pada 11 fitur nilai akurasi yang diberikan hanya sekitar 50%. Hal ini dapat terjadi karena sistem mampu mengenali pola notasi SMILES lebih baik dengan 15 fitur. Ini membuktikan bahwa semakin banyak fitur yang digunakan saat proses kIasifikasi menggunakan KMNB maka niIai akurasi juga semakin meningkat karena sistem mampu mengenali karakteristik dari masing-masing senyawa aktif lebih kompleks.

4.6 Pengujian Metode KMNB dengan Metode Konvensional

Pengujian ini diIakukan sebanyak 10 kaIi percobaan untuk tiap skenario metode yang digunakan. Gambar 7 merupakan hasil perbandingan pengujian metode.



Gambar 7. Hasil pengujian metode

Dari Gambar 7 dapat diketahui bahwa pada pengujian metode KMNB mendapatkan hasil rata-rata paling tinggi yakni sebesar 84,81%. Hal ini menunjukkan bahwa dengan menggunakan metode KMNB sistem mampu melakukan proses pengklasifikasian lebih optimal karena data yang belum diketahui klasternya (klaster *unknown*) pada tahap *K-Means* akan diklasifikasikan kembali menggunakan metode *Naïve Bayes*.

5. KESIMPULAN

Dari peneIitian tersebut, dapat ditarik kesimpulan yaitu:

1) Pengimplementasian KMNB pada klasifikasi fungsi senyawa aktif diawali dengan *preprocessing*.
2) Setelah nilai *preprocessing* didapatkan maka data diproses dengan metode *K-Means* dengan penentuan klaster sebanyak jumlah kelas ditambah satu. Selanjutnya, mengelompokkan data sesuai dengan kelas mayoritasnya dan data yang belum dapat dikelompokkan (klaster *unknown*) akan dilakukan proses *Naïve* Bayes.
3) Hasil pengujian sistem dengan metode KMNB meliputi:
   a. Pengujian validasi sistem yang memperoleh akurasi sebesar 90% dan sesuai dengan perhitungan manual.
   b. Pengujian *hold out validation* mendapatkan nilai rata-rata akurasi tertinggi dengan rasio jumlah data Iatih 80% dan data uji 20% yang memperoleh akurasi sebesar 85,45%.
   c. Pengujian dengan *K-Fold Cross Validation* mendapatkan akurasi tertinggi 86,66% dan terendah 70,37% dengan nilai rata-rata 82,6%.
   d. Pengujian jumlah fitur yang memperoleh hasil rata-rata akurasi sebesar 84,81% untuk jumlah 15 fitur dan akurasi sebesar 77% untuk jumlah 11 fitur.
   e. Pengujian metode dimana KMNB mampu mendapatkan akurasi yang lebih baik daripada metode konvensionalnya.

Penambahan saran disertakan untuk memperbaiki kekurangan yang terjadi selama atau pada hasil penelitian agar untuk kedepannya sistem dapat dikembangkan. Saran yang bisa diambiI untuk peneIitian selanjutnya:

1) Menambah data Iatih dan data uji sehingga sistem dapat mengklasifikasikan fungsi senyawa aktif lebih dari 3 kelas.
2) Percobaan menggunakan metode lain.

6. DAFTAR REFERENSI

Agusta, Y., 2007. “K-Means penerapan permasalahan dan metode terkait”. Jurnal Sistem dan Informatika, Vol 3.

Asikin, M. F., Ratnawati, D. E., & Fauzi, M. A., 2017. “Penentuan Penerima Bantuan Ternak Menggunakan Algoritma K-Means & Naïve Bayes”. S1. Universitas Brawijaya.

Grainner. 1998. “Penyederhanaan Bayes”. Jakarta: Luhur.

Junaedi, H. 2011. “Penggambaran Rantai Karbon dengan Menggunakan Simplified Molecular Input Line System (SMILES)”. Prosiding Konferensi Nasional Inovasi dalam Desain dan Teknologi- IDeaTech 2011. Sekolah Tinggi Teknik Surabaya: 219-226.

Kusnawi, 2007. “Pengantar Solusi Data Mining”. Yogyakarta, Seminar Nasional Teknologi.

Lestari, P., I., Ratnawati, D., E., Muflikhah, L. 2018. “Implementasi Algoritme K-Means Clustering Dan Naive Bayes Classifier Untuk Klasifikasi Diagnosa Penyakit Pada Kucing”. S1. Universitas Brawijaya.

Manning, C., P. Raghavan, dan H. Schutze. 2009. “Introduction to Information Retrieval”. England: Cambridge University Press.

Marisa, H., Mukti, R. W. & Salni. 2011. “Isolasi Senyawa Antibakteri Dari Daun Jengkol (Pithecolobium lobatum Benth) dan Penentuan Nilai KHM-nya”. Jurnal Penelitian Sains, Volume 14, pp. 38-41.

Ramzini, S., Ratnawati, D. E., Anam, S. 2018. "Penerapan Metode Learning Vector Quantization (LVQ) untuk Klasifikasi Fungsi Senyawa Aktif Menggunakan Notasi Simplified Molecular Input Line System (SMILES)" . S1. Universitas Brawijaya.

Weininger, D. 1988. "SMILES, a Chemical Language and Information System". Medicinal Chemistry Project, Pomona College, Claremont, California 91 7 11.

Xhemali, D.,.Hinde, C. J., Stone, R. G., 2009. "Naïve Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages". IJCSI International Journal of Computer Science Issues, Vol. 4, No. 1, 2009.