Machine Learning

Kombinasi *K-Nearest Neighbor* (KNN) dan *Relief-F* Untuk Meningkatkan Akurasi Pada Klasifikasi Data

Rahmad Nurhadi Yusra 1, Opim Salim Sitompul 2, Sawaluddin 3

1,2Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Sumatera Utara, Medan, Indonesia

*3Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sumatera Utara, Medan, Indonesia*

|  |  |  |
| --- | --- | --- |
| Article Information |  | **ABSTRACT** |
| Received: Revised: Available online:  | Dalam penelitian ini, penulis mengusulkan proses peningkatan akurasi pada *K-Nearest Neighbor* (KNN) dengan kombinasi seleksi fitur menggunakan metode *Relief-F*. Adapun penyebab kurang maksimalnya akurasi pada *K-Nearest Neighbor* dibandingkan dengan metode klasifikasi lainnya disebabkan oleh pengaruh atribut yang kurang signifikan dan persentase pengaruh yang cenderung rendah dari suatu data dalam menentukan kelas pada data baru. Metode *Relief-F* digunakan untuk melakukan seleksi pada atribut yang korelasinya kurang baik dari data yang diujikan. Pengujian dari metode yang diusulkan yaitu membandingkan akurasi yang diperoleh dari metode KNN tanpa menggunakan seleksi fitur dengan KNN menggunakan seleksi fitur *Relief-F*. Hasil pengujian yang diperoleh yaitu metode yang diusulkan mampu meningkatkan akurasi klasifikasi dari KNN dengan peningkatan yang diperoleh yaitu sebesar 10.32% setelah dibandingkan dengan pengujian KNN tanpa seleksi fitur. |
| Keywords |
| Klasifikasi, K-Nearest Neighbor*, lRelief-F,* Seleksi Fitur, Peningkatan Akurasi |
| Correspondence |
| Phone: 081360331775E-mail: hadi2106yusra@gmail.com |

# PENDAHULUAN

Dalam ikajian i*Data* i*Mining* iataupun ijuga i*Machine* i*Learning*, suatu iteknik iyang idigunakan idalam imemilih iataupun menghilangkan ifitur iyang idianggap ikurang ipenting idengan menggunakan ikriteria itertentu idan itermasuk ijuga isalah isatu dari itahapan i*pre-processing* ipada idata, idan iteknik itersebut dikenal idengan isebutan iSeleksi Fitur(*Feature Selection*) [1].

*Feature Selection* merupakan cara yang efektif untuk melakukan ireduksi idata idan imenjadi ilangkah ipenting iyang perlu idilakukan iagar idapat imemperoleh ikinerja idan ihasil iyang baik ipada i*data* i*mining* [2].

Salah isatu imetode iyang isering idigunakan ipada iData Mining iataupun iMachine iLearning iyaitu i*K-Nearest lNeightbor* (KNN) yang merupakan metode klasifikasi yang paling sederhana dan paling sering digunakan untuk *l*klasifikasi*l*dengan menggunakan ipendekatan *l*iberdasarkan jarak tetangga terdekat dari data ipembelajaran ike idata ipengujian idalam menetapkan sebuah ikelas ipada idata iberdasarkan isuara iterbanyak [3].

Dari beberapa ipenelitian yang telah ada, akurasi yang diperoleh KNN kurang maksimal dibandingkan dari metode klasifikasi lainnya. Seperti pada penelitian dari [4], yang memperoleh perbandingan akurasi antara SVM dan iKNN dengan akurasi SVM sebesar 82.54% dan KNN sebesar 79,22%.

Kemudian pada penelitian [5], hasil perbandinganiantara *Naïve* i*Bayes* yang memiliki iakurasi iterbaik sebesar 73,7%, *Decision Tree* yaitu 58,9% dan KNN sebesar 56,7%.

Dari penjabaraniibeberapa penelitian tersebut dapat ditarik kesimpulanibahwaiakurasiiyangidiperolehiKNN cenderung lebih rendah yang disebabkan pengaruh atribut yang digunakan terhadap proses klasifikasi.

Pada penelitian [6] mengusulkan idengan imelakukan iseleksi fituriiuntukiimembuang atribut-atribut yang kurang relevan sebelum idata i idiklasifikasikan idengan K-NN. Dan berdasarkan hasil penelitian dari penelitianitersebut seleksi fituridapat menghasilkan akurasi iyang ilebih ibesar ipada iKNNisetelah dilakukannya seleksi ifitur.

Pada penelitian [7], yang melakukan penelitian dalam meningkatkan iakurasi untuk imemprediksi jenis kelamin (*Gender*) dengan seleksi fitur *Relief-F* denganihasil iyang diperoleh iyaitu iakurasi isebesar 93.35 % isetelah imenghapus fitur yang ikurang irelevan ipada idata iyang idiujikan.

Kemudian pada penelitian [8] yang melakukan kombinasi SVM dengan *Relief-F* dalam klasifikasi data yang menghasilkan peningkatan akurasi setelah SVM dikombinasikan dengan *ReliefF* yaitu peningkatan akurasi sebesar 24.37 %.

Maka pada penelitian ini penulis mengusulkan untuk mengkombinasikan imetode iKNN idengan iseleksi ifitur *Relief-F* dengan iharapan idapat imeningkatkan iperolehan iakurasi idari proses iklasifikasi idata ipada imetode iKNN idan idapat memaksimalkan iproses ipenentuan ikelas ipada isuatu idata.

**TINJAUAN PUSTAKA**

***Penelitian Terdahulu***

Penelitian dari [4], membandingkan akurasi antara *Support Vector Machine* (SVM) dan iKNN daniterlihat bahwa kinerja SVM lebih baik daripada KNN, dengan akurasi SVM sebesar 82.54% dan KNN sebesar 79,22%.

Penelitian [5], membandingkaniperforma idari i*Naïve* i*Bayes*, *Decision* i*Tree* dan *K-Nearest* i*Neighbor* (KNN). *Naïve* i*Bayes* memiliki iakurasi iterbaik dengan akurasi rata-rata 73,7%, *Decision Tree* yaitu 58,9% dan KNN sebesar 56,7%.

Penelitian [7], peningkatan akurasi dalam memprediksi jenis kelamin (*Gender*) dengan seleksi fitur menggunakan *Relief-F* denganhasil yang diperoleh menggunakan Relief-F yaitu sebesar 93.35 % setelah menghapus fitur yang kurang relevan dari kumpulan fitur yang ada pada data yang diujikan.

Penelitian [8] yang membahas tentang kombinasi metode *Support Vector Machine* (SVM) dengan *Relief-F* dalam proses klasifikasi. Dan terjadinya peningkatan akurasi setelah *Support Vector Machine* (SVM) dikombinasikan dengan *Relief-F* yaitu sebesar 24.37 %.

***Klasifikasi***

Klasifikasi umumnya dipergunakan untuk memprediksi atau mengelompokkan suatu ikelas idari isuatu idata berdasakan atribut. Adapun i*discription* idari isetiapi kelas idata itelah iditetapkan [9].

***K-Nearest Neighbor (K-NN)***

*K-Nearest lNeightbor* (KNN) imerupakan imodel iklasifikasi yang paling sederhana dengan menggunakan *l*pendekatan dari *k* tetangga terdekat *l*dan menetapkan sebuah kelas berdasarkan suara iterbanyak [10].

Langkah-langkah idalam iproses iklasifikasi ipada i*K-Nearest Neightbor* (KNN) isebagai iberikut:

1. Menentukan iparameter ik (jumlah itetangga ipaling idekat).
2. Menghitung ikedekatan iberdasarkan imodel ijarak i*Euclidean* terhadap idata ilatih iyang idiberikan, idengan ipersamaan (1) berikut.

 (1)

1. Mengurutkan ihasil ijarak iyang ididapatkan isecara i*ascending* (berurutan idari inilai itertinggi ike iterendah).
2. Menghitung ijumlah isetiap ikelas iberdasarkan *k* itetangga terdekat.
3. Kemudian, ikelas idengan inilai iterbesar idijadikan isebagai kelas iuntuk idata iuji.

***Seleksi Fitur***

Proses yang dilakukan untuk menentukan fitur-fitur yang signifikan idalam idataset yang isesuai untuk ipermasalahan iyang akan idipecahkan disebut idengan Seleksi iFitur. Jika semakin baik fitur yang dipilih, maka dapat meningkatkan nilai *accuracy* dari metode yang diuji. Seleksi fitur juga bermanfaat dalam mereduksi dimensi dari dataset idengan icara imembuang ifitur-fitur iyang tidak imemiliki ipengaruh iterhadap ipenentuan ikelas/label [1].

***Relief-F***

*Relief-F* merupakan metode untuk seleksi fitur yang dikembangkan oleh Kononenko ipada itahun 1994 idan merupakan pengembangan idari i*Relief*. *Relief-F* memiliki efisiensi itinggi dan tidak membatasi karakteristik tipe data. Langkah-langkah untuk imencari inilai ibobot ifitur idengan *Relief-F* iyaitu isebagai iberikut [11].

1. Inisialisasi inilai iawal iseluruh ibobot ifitur = 0 dan menentukan ijumlah iiterasi.
2. Memilih isebuah idata iyang iakan idijadikan isebagai ititik acak iatau ititik ipusat.
3. Mencari i*miss* idan i*hit* iterdekat idengan icara imenghitung jarak iantara ititik ipusat idengan idata iyang imemiliki ikelas yang isama. iJarak iterdekat iantara ititik ipusat idan idata ipada kelas ipositif idisebut i*hit*. iSedangkan, ijarak iterdekat iantara titik ipusat idengan idata iyang ipada kelas negatif disebut *miss*.
4. Melakukan i*update* ibobot iuntuk isetiap ifitur. Fitur dengan data ikategori idihitung menggunakan Persamaan (2) berikut.

 (2)

1. Sedangkan, ifitur idengan idata inumerik idihitung menggunakan ipersamaan i(3) iberikut.

 (3)

1. Sehingga irumus iperbaruan ibobot idihitung imenggunakan persamaan (4) iberikut.

 (4)

1. Selanjutnya, idilanjutkan idengan iiterasi iselanjutnya iyang dimulai idari ilangkah 1 ihingga ibobot ifitur iyang ibaru itelah didapat.

***Fold Cross Validation***

*Fold Cross Validation* digunakan iuntuk imengevaluasi idan membandingkan ialgoritma ipembelajaran idengan imembagi idata menjadi idua isegmen iyaitu imodel idata ipembelajaran (*training*) dan imodel idata ipengujian (*testing*) [12]. Tujuan iutama idari i*fold cross validation* iyaitu iuntuk imengukur ikinerja ipada imodel pembelajaran idari idata iyang itersedia idalam isatui metode. Selain iitu idapat ijuga iuntuk imembandingkan ikinerja idari idua atau ilebih ivariasi imodel iparameter.

Pada ipenelitian iini imenggunakan imetode ievaluasi i*10-fold cross validation* dengan membentuki10 i*subsets* dari *data set* yang digunakan. 9 *subsets* idigunakan isebagai i*training* idan 1 *subset* idigunakan isebagai i*testing*. iHasil ipengukuran akhiriiyaitu irata-rata idari i10 ikali ipengujian.

# METODOLOGI PENELITIAN

***Tahapan Penelitian***

Adapun untuk menguji dari metode yang diusulkan pada penelitian ini dilakukan dengan tahapan penelitian pada Gambar 1 berikut.



**Gambar 1.** Tahapan Penelitian

Adapun iketerangan itahapan ipenelitian ipada iGambar 1 menggambarkan ilangkah-langkah lyang lditerapkan ipada penelitian iini iyaitu isebagai iberikut:

***Dataset***

Adapun data yang digunakan pada penelitian yaitu *User Knowledge Modeling* yangimerupakan idata set iberisi itentang status ipengetahuan ipelajar itentang isubjek i*Electrical* i*DC Machines*. iAdapun Data set *User Knowledge Modeling* diperoleh dari *UCI* i*Machine* i*Learning* i*Repository*. Karakteristik dari data set tersebut yaitu Multivariate, dengan karakteristik atribut yaitu Integer. Jumlah idata ipada idata iset itersebut iyaitu 403 data dengan ibanyak iatribut isebanyak 5 atribut dan 1 kelas atribut yang iterdiri dari 4 kelas yaitu *Very Low, Low, Middle*, dan *High*.

Adapun rincian informasi atribut yang berhubungan pada *User Knowledge Modeling Dataset* yaitu pada Tabel 1 berikut.

Tabel 1. Atribut i*User Knowledge* i*Modeling* i*Dataset*

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Atribut | Nilai | Penanda |
| 1. | *STG* | 0 – 0.99 | X1 |
| 2. | *SCG* | 0 – 0.9 | X2 |
| 3. | *STR* | 0 – 0.95 | X3 |
| 4. | *LPR* | 0 – 0.99 | X4 |
| 5. | *PEG* | 0 – 0.99 | X5 |
| 6. | *UNS* (*Class*) | *Very Low, Low, Middle, High* | - |

Adapun iinformasi irincian idari idata iyang idigunakan iyaitu lpada iTabel l2 lberikut:

Tabel 2. lRincian iData *User* i*Knowledge* i*Modeling*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| lNo. | lX1 | X2 | lX3 | lX4 | lX5 | Class |
| 1 | 0 | 0 | 0 | 0 | 0 | ivery\_low |
| 2 | 0.08 | 0.08 | 0.1 | 0.24 | 0.9 | iHigh |
| 3 | 0.06 | 0.06 | 0.05 | 0.25 | 0.33 | iLow |
| 4 | 0.1 | 0.1 | 0.15 | 0.65 | 0.3 | iMiddle |
| 5 | 0.08 | 0.08 | 0.08 | 0.98 | 0.24 | iLow |
| 6 | 0.09 | 0.15 | 0.4 | 0.1 | 0.66 | iMiddle |
| 7 | 0.1 | 0.1 | 0.43 | 0.29 | 0.56 | iMiddle |
| 8 | 0.15 | 0.02 | 0.34 | 0.4 | 0.01 | ivery\_low |
| 9 | 0.2 | 0.14 | 0.35 | 0.72 | 0.25 | iLow |
| 10 | 0 | 0 | 0.5 | 0.2 | 0.85 | iHigh |
| 11 | 0.18 | 0.18 | 0.55 | 0.3 | 0.81 | iHigh |
| 12 | 0.06 | 0.06 | 0.51 | 0.41 | 0.3 | iLow |
| 13 | 0.1 | 0.1 | 0.52 | 0.78 | 0.34 | iMiddle |
| 14 | 0.1 | 0.1 | 0.7 | 0.15 | 0.9 | iHigh |
| 15 | 0.2 | 0.2 | 0.7 | 0.3 | 0.6 | iMiddle |
| 16 | 0.12 | 0.12 | 0.75 | 0.35 | 0.8 | iHigh |
| 17 | 0.05 | 0.07 | 0.7 | 0.01 | 0.05 | ivery\_low |
| 18 | 0.1 | 0.25 | 0.1 | 0.08 | 0.33 | iLow |
| 19 | 0.15 | 0.32 | 0.05 | 0.27 | 0.29 | iLow |
| 20 | 0.2 | 0.29 | 0.25 | 0.49 | 0.56 | iMiddle |
| 21 | 0.12 | 0.28 | 0.2 | 0.78 | 0.2 | iLow |
| 22 | 0.18 | 0.3 | 0.37 | 0.12 | 0.66 | iMiddle |
| 23 | 0.1 | 0.27 | 0.31 | 0.29 | 0.65 | iMiddle |
| 24 | 0.18 | 0.31 | 0.32 | 0.42 | 0.28 | iLow |
| 25 | 0.06 | 0.29 | 0.35 | 0.76 | 0.25 | iLow |
| 26 | 0.09 | 0.3 | 0.68 | 0.18 | 0.85 | iHigh |
| 27 | 0.04 | 0.28 | 0.55 | 0.25 | 0.1 | ivery\_low |
| 28 | 0.09 | 0.255 | 0.6 | 0.45 | 0.25 | iLow |
| 29 | 0.08 | 0.325 | 0.62 | 0.94 | 0.56 | iHigh |
| 30 | 0.15 | 0.275 | 0.8 | 0.21 | 0.81 | iHigh |
| … | **…** | **…** | **…** | **…** | **…** | **…** |
| 391 | 0.78 | 0.47 | 0.29 | 0.98 | 0.59 | High |
| 392 | 0.58 | 0.4 | 0.32 | 0.22 | 0.24 | Low |
| 393 | 0.68 | 0.43 | 0.6 | 0.47 | 0.55 | Middle |
| 394 | 0.57 | 0.37 | 0.75 | 0.27 | 0.32 | Low |
| 395 | 0.62 | 0.56 | 0.11 | 0.24 | 0.22 | Low |
| 396 | 0.64 | 0.58 | 0.14 | 0.32 | 0.21 | Low |
| 397 | 0.64 | 0.59 | 0.12 | 0.58 | 0.24 | Low |
| 398 | 0.68 | 0.61 | 0.34 | 0.31 | 0.23 | Low |
| 399 | 0.9 | 0.78 | 0.62 | 0.32 | 0.89 | High |
| 400 | 0.85 | 0.82 | 0.66 | 0.83 | 0.83 | High |
| 401 | 0.56 | 0.6 | 0.77 | 0.13 | 0.32 | Low |
| 402 | 0.66 | 0.68 | 0.81 | 0.57 | 0.57 | Middle |
| 403 | 0.68 | 0.64 | 0.79 | 0.97 | 0.24 | Middle |

Tabel 3. Jumlah Data Per Kelas

|  |  |  |
| --- | --- | --- |
| No. | Class Attribute | Jumlah Data |
| 1. | *Very Low* | 50 |
| 2. | *Low* | 129 |
| 3. | *Middle* | 122 |
| 4. | *High* | 102 |
|  | Jumlah | 403 |

# *Preprocessing*

*Preprocessing* idilakukan iuntuk imempersiapkan idata iyang benar-benar ivalid isebelum idiproses ipadai tahap berikutnya. Pada itahap iini idilakukan i*cleansing*, itransformasi, iseleksi atribut. iData iyang ididapat iakan idiolah iuntuk imendapatkan atribut iyang irelevan idan isesuai. iPada ipenelitian iini, normalisasi idata idilakukan idengan imenggunakan imetode i*Min-Max* idengan irumus iberikut.

 (5)

Keterangan:

*Newdata* : Data ihasil inormalisasi

*Min* : Nilai iminimum idari idata iperkolom

*Max* : Nilai imaksimum idari idata iperkolom

*Newmin* : Batas inilai iminimum iyang ikita iberikan

*Newmax* : Batas inilai imaksimum iyang ikita iberikan

***Klasifikasi K-Nearest Neighbor***

Setelah iihasil inormalisasi idata idiperoleh, iuntuk iselanjutnya dalam iproses iklasifikasi iterhadap idata set iyang idiujikan terlebih idahulu dilakukan pembagian data set menjadi dua bagian yaitu *Data Training* dan *Data Testing* dengan menggunakan metode evaluasi *10Fold Cross Validation* idengan ipembagiannya yaitu 9 subset iuntuk i*training* idan 1 subset iuntuk i*testing*.

Kemudian klasifikasi idilakukan dengan dua tahap yaitu pertama dengan iKNN isecara ikonvensional atau tanpa menggunakan seleksi fitur dan yang kedua yaitu KNN dengan seleksi ifitur iyang idiperoleh dari iproses iperhitungan *Relief-F*.

***Perbandingan Akurasi Klasifikasi***

Evaluasi ibertujuan iuntuk imengetahui itingkat iakurasi ihasil klasifikasi idataset iterhadap idata iuji. iEvaluasi idilakukan iuntuk menemukan isolusi ioptimaliyang idihasilkan idari iberbagai metode iklasifikasi iyang iikompleks idan idilakukaniberulang-ulang [13].

Oleh ikarena itu, ipada ipenelitian ini iakan imenggunakan *Accuracy* ikarena idianggap memadai. *Accuracy* dapat dirumuskan imenggunakan ipersamaan isebagai iberikut:

 (6)

# HASIL DAN PEMBAHASAN

## Hasil Penelitian

Untuk imempermudah ipengujian idari imetode iyang idiusulkan pada ipenelitian ini, ipenulis idibantu idengan imenggunakan *Tools* yaitu i*Jupyter* i*Notebook* idan i*Rapid* i*Miner* i*Studio* yang iberguna dalam imempersingkat iwaktu ipengujian iterhadap i*User Knowledge* i*Modeling* i*Dataset* iuntuk imelihat ihasil iperolehan akurasi idari iperbandingan iKNN isecara ikonvensional idengan kombinasi iKNN imenggunakan iseleksi ifitur i*Relief-F*.

Kemudian ilangkah ipertama iyang dilakukan yaitu melakukan i*Preprocessing* iuntuk imenormalisasi idata set iagar diperoleh idata iyang ivalid isebelum idilakukan iklasifikasi. Normalisasi idata idilakukan imenggunakan iperhitungan normalisasi i*min-max* isesuai idengan ipersamaan (5) iyang itelah dijelaskan ipada ibagian isebelumnya iAdapun ihasil inormalisasi *min-max* ipada i*User* i*Knowledge* i*Modeling* i*Dataset* iyaitu ipada Tabel 4 iberikut.

Tabel 4. Hasil iNormalisasi i*User* i*Knowledge* i*Modeling*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| lNo. | lX1 | X2 | lX3 | lX4 | lX5 | Class |
| 1 | 0 | 0 | 0 | 0 | 0 | ivery\_low |
| 2 | 0.0808 | 0.0889 | 0.1053 | 0.2424 | 0.9091 | iHigh |
| 3 | 0.0606 | 0.0667 | 0.0526 | 0.2525 | 0.3333 | iLow |
| 4 | 0.101 | 0.1111 | 0.1579 | 0.6566 | 0.303 | iMiddle |
| 5 | 0.0808 | 0.0889 | 0.0842 | 0.9899 | 0.2424 | iLow |
| 6 | 0.0909 | 0.1667 | 0.4211 | 0.101 | 0.6667 | iMiddle |
| 7 | 0.101 | 0.1111 | 0.4526 | 0.2929 | 0.5657 | iMiddle |
| 8 | 0.1515 | 0.0222 | 0.3579 | 0.404 | 0.0101 | ivery\_low |
| 9 | 0.202 | 0.1556 | 0.3684 | 0.7273 | 0.2525 | iLow |
| 10 | 0 | 0 | 0.5263 | 0.202 | 0.8586 | iHigh |
| 11 | 0.1818 | 0.2 | 0.5789 | 0.303 | 0.8182 | iHigh |
| 12 | 0.0606 | 0.0667 | 0.5368 | 0.4141 | 0.303 | iLow |
| 13 | 0.101 | 0.1111 | 0.5474 | 0.7879 | 0.3434 | iMiddle |
| 14 | 0.101 | 0.1111 | 0.7368 | 0.1515 | 0.9091 | iHigh |
| 15 | 0.202 | 0.2222 | 0.7368 | 0.303 | 0.6061 | iMiddle |
| 16 | 0.1212 | 0.1333 | 0.7895 | 0.3535 | 0.8081 | iHigh |
| 17 | 0.0505 | 0.0778 | 0.7368 | 0.0101 | 0.0505 | ivery\_low |
| 18 | 0.101 | 0.2778 | 0.1053 | 0.0808 | 0.3333 | iLow |
| 19 | 0.1515 | 0.3556 | 0.0526 | 0.2727 | 0.2929 | iLow |
| 20 | 0.202 | 0.3222 | 0.2632 | 0.4949 | 0.5657 | iMiddle |
| 21 | 0.1212 | 0.3111 | 0.2105 | 0.7879 | 0.202 | iLow |
| 22 | 0.1818 | 0.3333 | 0.3895 | 0.1212 | 0.6667 | iMiddle |
| 23 | 0.101 | 0.3 | 0.3263 | 0.2929 | 0.6566 | iMiddle |
| 24 | 0.1818 | 0.3444 | 0.3368 | 0.4242 | 0.2828 | iLow |
| 25 | 0.0606 | 0.3222 | 0.3684 | 0.7677 | 0.2525 | iLow |
| 26 | 0.0909 | 0.3333 | 0.7158 | 0.1818 | 0.8586 | iHigh |
| 27 | 0.0404 | 0.3111 | 0.5789 | 0.2525 | 0.101 | ivery\_low |
| 28 | 0.0909 | 0.2833 | 0.6316 | 0.4545 | 0.2525 | iLow |
| 29 | 0.0808 | 0.3611 | 0.6526 | 0.9495 | 0.5657 | iHigh |
| 30 | 0.1515 | 0.3056 | 0.8421 | 0.2121 | 0.8182 | iHigh |
| … | **…** | **…** | **…** | **…** | **…** | **…** |
| 391 | 0.7879 | 0.5222 | 0.3053 | 0.9899 | 0.596 | High |
| 392 | 0.5859 | 0.4444 | 0.3368 | 0.2222 | 0.2424 | Low |
| 393 | 0.6869 | 0.4778 | 0.6316 | 0.4747 | 0.5556 | Middle |
| 394 | 0.5758 | 0.4111 | 0.7895 | 0.2727 | 0.3232 | Low |
| 395 | 0.6263 | 0.6222 | 0.1158 | 0.2424 | 0.2222 | Low |
| 396 | 0.6465 | 0.6444 | 0.1474 | 0.3232 | 0.2121 | Low |
| 397 | 0.6465 | 0.6556 | 0.1263 | 0.5859 | 0.2424 | Low |
| 398 | 0.6869 | 0.6778 | 0.3579 | 0.3131 | 0.2323 | Low |
| 399 | 0.9091 | 0.8667 | 0.6526 | 0.3232 | 0.899 | High |
| 400 | 0.8586 | 0.9111 | 0.6947 | 0.8384 | 0.8384 | High |
| 401 | 0.5657 | 0.6667 | 0.8105 | 0.1313 | 0.3232 | Low |
| 402 | 0.6667 | 0.7556 | 0.8526 | 0.5758 | 0.5758 | Middle |
| 403 | 0.6869 | 0.7111 | 0.8316 | 0.9798 | 0.2424 | Middle |

Kemudian imelakukan itahapan idalam iseleksi ifitur ipada *User* i*Knowledge* i*Modeling* i*Dataset* imenggunakan iseleksi ifitur *ReliefF*. Dalam imelakukan iseleksi ifitur idengan i*Relief-F*, ipada penelitian iini idibantu idengan imenggunakan i*software* i*Rapid Miner* i*Studio* iuntuk imemperoleh inilai ibobot idari masing-masing iatribut ipada i*User* i*Knowledge* i*Modeling* i*Dataset*. iDan menghitung ipersentase idari ibobot idata iset idilakukan idengan menghitung inilai inormalisasi i*min-max* idengan irentang persentase idari i0 % isampai i100 %.

Hasil iperolehan ibobot iatribut iakan idiurutkan iberdasarkan urutan inilai ibobot iterbesar ihingga iurutan inilai ibobot iterendah yang ibertujuan iuntuk imelihat iseberapa ibesar itingkat ipengaruh dari imasing-masing iatribut iterhadap idata iberdasarkan ipersentase idari imasing-masing iatribut.

Tabel 5. Nilai Bobot *Relief-F* Data Set *User Knowledge Modeling*

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Atribut | lNilai Bobot | Persentase |
| 1. | X5 | 0.302 | 100 % |
| 2. | X4 | 0.037 | 10.7 % |
| 3. | X1 | 0.016 | 3.8 % |
| 4. | X2 | 0.007 | 0.6 % |
| 5. | X3 | 0.005 | 0 % |

Berdasarkan ipada iTabel 4 ibobot iyang idiperoleh idari 5 atributi*User* i*Knowledge* i*Modeling* i*Dataset*. Kemudian dilakukan seleksi fitur untuk membuang fitur dengan nilai persentase terendah. Atribut yang dihapus yaitu X1, X2, X3 karena memiliki ipersentase irendah iyang imenunjukkan ibahwa atribut itersebut imemiliki ipengaruh iyang ikurang isignifikan pada data iset.

Atribut X4 dan X5 imemiliki ipersentase iyang iterbesar sehingga idapat idiketahui ibahwa iatribut itersebut imemiliki pengaruh iyang ibesar ipada idata iset idan isebagai iatribut iyang terpilih iuntuk iproses iperhitungan iklasifikasi.

Tabel 6. Hasil Seleksi Fitur *User Knowledge Modeling Dataset*

|  |  |  |  |
| --- | --- | --- | --- |
| No. | Atribut | Persentase | Keterangan |
| 1. | X5 | 100 % | Terpilih |
| 2. | X4 | 10.7% | Terpilih |

Langkah selanjutnya yaitu menghitung klasifikasi KNN dengan dua tahap yang pertama yaitu menghitung KNN secara konvensional atau tanpa seleksi fitur, dan yang kedua yaitu KNN dengan iseleksi ifitur idari *ReliefF*. Adapun ihasil iakurasi idari KNN secara ikonvensional idan iKNN idengan seleksi ifitur idari i*ReliefF* dapat idilihat ipada iTabel 6 idan iTabel 7 isebagai iberikut.

Tabel 7. lHasil lAkurasi K-NN Secara Konvensional

|  |  |  |
| --- | --- | --- |
| *k* | Jumlah Data Benar | Akurasi (%) |
| 1 | 338 | 83.87 |
| 2 | 340 | 84.37 |
| 3 | 354 | 87.84 |
| 4 | 340 | 84.37 |
| 5 | 349 | 86.60 |
| 6 | 342 | 84.86 |
| 7 | 347 | 86.10 |
| 8 | 345 | 85.61 |
| 9 | 342 | 84.86 |
| 10 | 341 | 84.62 |
| Rata-Rata |  | 85.31 |

Tabel 8. Hasil Akurasi K-NN dengan *Feature Selection* *Relief-F*

|  |  |  |
| --- | --- | --- |
| *k* | Jumlah Data Benar | Akurasi (%) |
| 1 | 381 | 94.54 |
| 2 | 381 | 94.54 |
| 3 | 383 | 95.04 |
| 4 | 384 | 95.29 |
| 5 | 386 | 95.78 |
| 6 | 388 | 96.28 |
| 7 | 388 | 96.28 |
| 8 | 388 | 96.28 |
| 9 | 388 | 96.28 |
| 10 | 387 | 96.03 |
| Rata-Rata |  | **95.63** |

***Pembahasan***

Pada ibagian iini imelakukan ipembahasan imengenai ihasil pengujian iyang idiperoleh iuntuk imembuktikan ikinerja idari metode iyang idiusulkan. iUntuk ilebih ijelas iperbandingan irata-rata iakurasi idari iseluruh i*k* iyang idiperoleh iKNN ikonvensional dan iKNN + *ReliefF* ipada iseluruh idata iset iyang idigunakan dalam ipengujian idapat idilihat ipada iTabel 9 iberikut.

Tabel 9. Perbandingan Hasil Akurasi

|  |  |  |  |
| --- | --- | --- | --- |
| k | K-NN Konvensional(%) | K-NN +ReliefF(%) | Selisih Akurasi (%) |
| 1 | 83.87 | 94.54 | 10.67 |
| 2 | 84.37 | 94.54 | 10.17 |
| 3 | 87.84 | 95.04 | 7.2 |
| 4 | 84.37 | 95.29 | 10.92 |
| 5 | 86.60 | 95.78 | 9.18 |
| 6 | 84.86 | 96.28 | 11.42 |
| 7 | 86.10 | 96.28 | 10.18 |
| 8 | 85.61 | 96.28 | 10.67 |
| 9 | 84.86 | 96.28 | 11.42 |
| 10 | 84.62 | 96.03 | 11.41 |
| Rata-Rata | **85.31** | **95.63** | **10.32** |

Dari iTabel i4.9 idapat idilihat ibahwa ihasil iakurasi isetelah melakukan seleksi atribut pada KNN dengan *Relief-F* diperoleh peningkatan nilai akurasi dibandingkan nilai akurasi yang dihasilkan oleh KNN konvensional dengan selisih rata-rata akurasi sebesar 10.32 %. Pada KNN + *ReliefF* akurasi tertinggi diperoleh saat *k* bernilai 6, 7, 8 dan 9 yaitu sebesar 96.28 %, sedangkan nilai akurasi iterendah idiperoleh ipada isaat ik ibernilai 1 dan 2 iyaitu isebesar 94.54 %. Pada iKNN iKonvensional akurasi itertinggi idiperoleh isaat *k* ibernilai 3 iyaitu isebesar 87.84 %, sedangkan inilai iakurasi iterendah idiperoleh ipada isaat k bernilai 1 sebesar 83.87 %. Rata-rata akurasi dari seluruh *k* Pada KNN + *Relief-F* yaitu sebesar 95.63 % dan pada KNN konvensional sebesar 85.31%. Berdasarkan metode yang diusulkan iterlihat ibahwa iKNN imenggunakan i*Relief-F* imampu meningkatkan iakurasi iyang ilebih ibaik idibandingkan iakurasi dari iKNN itanpa imenggunakan iseleksi ifitur.

# KESIMPULAN

Berdasarkan ihasil ipengujian idan ievaluasi iterhadap imetode KNN dengan iseleksi iatribut imenggunakan i*Relief-F*, imaka didapatkan ikesimpulan iyaitu idengan imelakukan iseleksi iatribut pada iKNN imenggunakan i*Relief-F* iterbukti idapat imeningkatkan akurasi ipada iKNN.

Hasil irata-rata ipeningkatan iakurasi adalah isebesar 10.32 %. Hasil ipengujian iseleksi iatribut ipada KNN idengan i*Relief-F* mampu imeningkatkan inilai akurasi ipada i*User* i*Knowledge Modeling* i*Dataset* idari iseluruh i*k* iPada iKNN + i*Relief-F* iyaitu isebesar 95.63 % idan ipada iKNN ikonvensional isebesar 85.31%.

REFERENSI

[1] H. Liu, H. Motoda, R. Setiono, and Z. Zhao, "Feature Selection: Ever Evolving Frontier in Data Mining,". In *Feature selection in data mining,* pp. 4-13. 2010.

[2] D. A. Irawan, Z. A. Baizal, and E. G. Perdana, "Analisis dan Implementasi Algoritma Relieff untuk Feature Selection pada Klasifikasi Dataset Multiclass (Doctoral dissertation, MS thesis, Universitas Telkom, Jakarta, Indonesia,". 2011.

[3] Y. Chen, and Y. Hao, "A Feature Weighted Support Vector Machine and K-Nearest Neighbor Algorithm for Stock Market Indices Prediction," *Expert Systems with Applications* (2017), vol. 80, pp. 340-355, 2017.

[4] J. S. Raikwal, and K. Saxena, "Performance Evaluation of SVM and K-Nearest Neighbor Algorithm over Medical Data set," *International Journal of Computer Applications*. vol. 50, no. 14, pp. 35-39, 2012.

[5] A. Ashari, I. Paryudi, and A. M. Tjoa, "Performance Comparison between Naïve Bayes, Decision Tree and k-Nearest Neighbor in Searching Alternative Design in an Energy Simulation Tool," (IJACSA) *International Journal of Advanced Computer Science and Applications*. vol. 4, no. 11, pp. 33-39, 2013.

[6] M. Danil, S. Efendi, and R. W. Sembiring, "The Analysis of Attribution Reduction of K-Nearest Neighbor (KNN) Algorithm by Using Chi-Square," In *Journal of Physics: Conference Series,* vol. 1424, no. 1, pp. 012004, 2019.

[7] T. R. Reddy, B. V. Vardhan, M. GopiChand, and K. Karunakar, "Gender prediction in author profiling using ReliefF feature selection algorithm," In *Intelligent Engineering Informatics,* pp. 169-176, Springer, Singapore, 2018.

[8] J. Huang, J. Zhou, and L. Zheng, "Support Vector Machine Classification Algorithm Based on Relief-F Feature Weighting," In *2020 International Conference on Computer Engineering and Application (ICCEA)*, pp. 547-553, 2020.

[9] J. Han, J. Pei, and M. Kamber, "Data Mining Concept and Techniques, 3rd edition," *Morgan Kaufmann-Elsevier*. vol. **2**, no. 1, pp. 88-97, 2012.

[10] A. Danades, D. Pratama, D. Anggraini, and D. Anggriani, "Comparison of Accuracy Level K-Nearest Neighbor Algorithm and Support Vector Machine Algorithm in Classification Water Quality Status," *International Conference on System Engineering and Technology*, pp. 137-141, 2016.

[11] I. Kononenko, "Estimating Attributes: Analysis and Extensions of Relief," In *European conference on machine learning*, pp. 171-182, 1994.

[12] P. Refaeilzadeh, L. Tang, and H. Liu, "Encyclopedia of Database Systems," In *Cross-validation*, pp. 532-538, 2009.

[13] J. D. Novaković, A. Veljović, S. S. Ilić, Ž. Papić, and T. Milica, "Evaluation of Classification Models in Machine Learning," *Theory and Applications of Mathematics & Computer Science*, vol. 7, no. 1, pp. 39-46, 2017.