Perbandingan Klasifikasi Hasil Indeks Kepuasan Masyarakat Terhadap UPT Balai Latihan Kerja Surabaya Menggunakan Algoritma *Decision Tree*, *Random Forest*, *K-Nearest Neighbor*

Satria Adi Wicaksono[1], I Kadek Dwi Nuryana[2]
[1,2]Program Studi Teknik Informatika, Fakultas Teknik, Universitas Negeri Surabaya
[1]satriawicaksono1@mhs.unesa.ac.id
[2]dwinuryana@unesa.ac.id

Abstrak **–** Pelaksanaan evaluasi hasil indeks kepuasan masyarakat terhadap UPT Balai Latihan Kerja Surabaya yang dilakukan dengan cara memberikan kuisioner kepada masyarakat dan perserta pelatihan yang ada di UPT Balai Latihan Kerja Surabaya, kemudian berkas kuisioner diserahkan kepada pihak Staff Pengembangan dan Pemasaran.. Akan tetapi hasil kuisioner yang seharusnya diberikan kepada kepala UPT Balai Latihan Kerja Surabaya sebagai bahan evaluasi UPT Balai Latihan Kerja Surabaya seringkali hanya terbengkalai dan tidak diproses lebih lanjut oleh UPT Balai Latihan Kerja Surabaya. Sehingga evaluasi kuisioner khususnya pada sarana prasarana melalui selembaran kertas bisa dianggap tidak efektif dan tidak sesuai sasaran. Untuk mengatasi permasalah diatas penulis memberikan alternatif dengan cara memanfaatkan ilmu yang diperoleh selama perkuliahan dengan menggunakan teknik *data mining* untuk memprediksi kepuasan masyarakat menggunakan metode perbandingan 3 algoritma yaitu *Decision Tree, Random Forest, K-Nearest Neighbor.* Klasifikasi merupakan Teknik dalam *data mining* untuk mengklompokkan data berdasarkan data terhadap data sampel. Pada penelitian ini, penulis melakukan perbandingan 3 teknik klasifikasi untuk mengklasifikasikan respon kepuasan masyarakat pada *dataset UPT Balai Latihan Kerja Surabaya.* Perbandingan teknik klasifikasi ini dilakukan untuk mengetahui model dalam teknik klasifikasi mana yang paling efektif untuk mengklasifikasikan target pada dataset UPT Balai Latihan Kerja Surabaya. Proses klasifikasi diawali dengan *preprocessing* datauntuk melakukan penghilangan *missing value* dan pemilihan fitur pada dataset. Pada tahap evaluasi digunakkan teknik *confusion matrix* dan *Kfold cross validation*. Setelah dilakukan pengujian, didapatkan bahwa hasil klasifikasi menunjukan akurasi terbaik diperoleh oleh model *Random Forest memperoleh accuracy 1.00, Decision Tree memperoleh accuracy 0.91 , dan K-Nearest Neighbor 0.94*.

Kata Kunci : *Perbandingan, Klasifikasi, data mining, Decision Tree, Random Forest, K-Nearest Neighbor*

1. PENDAHULUAN

Selama beberapa tahun terakhir perkembangan teknologi informasi menjadi sangat maju dalam hal pengumpulan dan penyimpanan data, maka munculah suatu kebutuhan untuk dapat menghasilkan informasi dari data yang telah ada tersebut. Setiap informasi yang ada menjadi suatu hal yang penting untuk menentukan setiap keputusan dalam situasi tertentu. Hal ini menyebabkan penyedian informasi menjadi sarana untuk dianalisa dan diringkas menjadi suatu pengetahuan dari data yang bermanfaat ketika pengambilan suatu keputusan dilakukan. Data *mining* merupakan sebuah proses ekstraksi untuk mendapatkan suatu informasi yang sebelumnya tidak diketahui dari sebuah data. Data *mining* dapat menganalisa kasus lama untuk menentukan pola dari sebuah data menggunakan teknik pengenalan pola seperti statistic dan matematika (Written,2011).

Masalah yang dihadapi dalam UPT Balai Latihan Kerja Surabaya ini dalam mengevaluasi hasil indeks kepuasan masyarakat terhadap UPT Balai Latihan Kerja Surabaya yang dilakukan dengan cara memberikan kuisioner kepada masyarakat dan perserta pelatihan yang ada di UPT Balai Latihan Kerja Surabaya, kemudian berkas kuisioner diserahkan kepada pihak Staff Pengembangan dan Pemasaran.. Akan tetapi hasil kuisioner yang seharusnya diberikan kepada kepala UPT Balai Latihan Kerja Surabaya sebagai bahan evaluasi UPT Balai Latihan Kerja Surabaya seringkali hanya terbengkalai dan tidak diproses lebih lanjut oleh UPT Balai Latihan Kerja Surabaya. Sehingga evaluasi kuisioner khususnya pada sarana prasarana melalui selembaran kertas bisa dianggap tidak efektif dan tidak sesuai sasaran. Untuk mengatasi permasalah diatas penulis memberikan alternatif dengan cara memanfaatkan ilmu yang diperoleh selama perkuliahan dengan menggunakan teknik *data mining* untuk memprediksi kepuasan

masyarakat menggunakan metode perbandingan 3 algoritma yaitu *Decision Tree, Random Forest, K- Nearest Neighbor.*

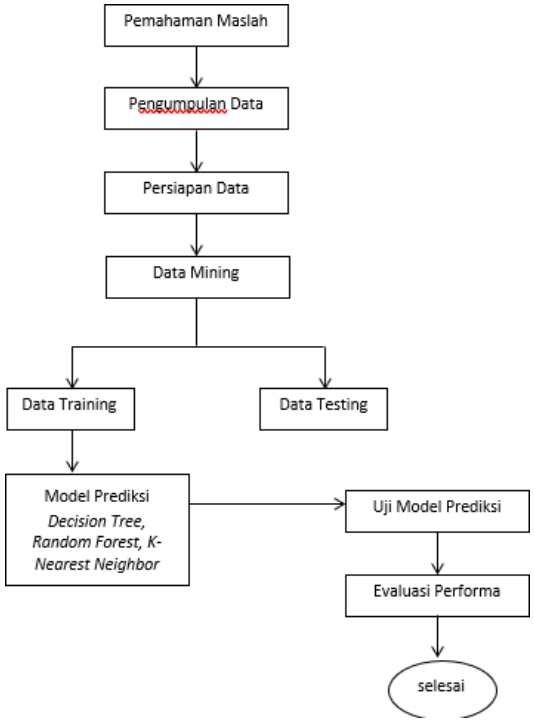
Data mining dapat diimplementasikan pada beberapa area seperti *Customer Relationship Management (CRM),* deteksi penipuan, evaluasi kredit, perawatan medis dan deteksi penyakit. Penelitian ini dilakukan dengan menganalisan perbandingan algoritma seperti *Decision tree, K- Nearest Neighbor dan Random Forest* dengan memperhatikan kriteria seperti keakuratan, kecepatan dan ketangguhan dalam penggunaan beberapa jenis data (Dogan & Tanrikulu, 2013). Pada penelitian ini menggunakan data yang diberikan oleh UPT Balai Latihan Kerja Surabaya dengan menggunakan metode klasifikasi dan membandingkan beberapa algoritma yaitu, Algoritma *Decision Tree*, Algoritma *K-Nearest Neighbor*, Algoritma *Random Forest* untuk menganalisis tingkat kepuasan masyarakat terhadap sarana prasarana UPT Balai Latihan Kerja Surabaya.

UPT Balai Latihan Kerja Surabaya merupakan salah satu Badan Usaha Milik Negara yang ditugaskan khusus oleh pemerintahan untuk membantu penyerapan lulusan oleh dunia kerja, UPT Balai Latihan Kerja Surabaya menyediakan layanan Kios 3in1, yaitu bentuk pelayanan untuk mengakses lowongan pekerjaan secara online dan Bursa Kerja Khusus (BKK) yang dapat dimanfaatkan oleh lulusan maupun perusahaan yang membutuhkan tenaga kerja. Pada tanggal 24 Desember 2010 lembaga pelatihan ini berhasil mendapatkan sertifikasi ISO 9001 : 2008 yang merupakan bukti pengakuan keberhasilan di bidang Manajemen Mutu.

UPT Balai Latihan Kerja (BLK) Surabaya dibangun pada tahun 1979 dengan dana bantuan dari Bank Dunia (World Bank) dan diresmikan oleh Menteri Tenaga Kerja Republik Indonesia (Dr. Harun Zein) pada 19 Maret 1980. Melalui berbagai pelatihan berbasis kompetensi yang telah diselenggarakan selama lebih dari 30 tahun, BLK Surabaya telah mencetak ribuan tenaga kerja terampil dan ahli yang bekerja di perusahaan- perusahaan nasional maupun multinasional. Kontribusi UPT Balai Latihan Kerja Surabaya sebagai lembaga pelatihan milik pemerintah dalam menghadapi tantangan saat ini adalah dengan meningkatkan daya saing tenaga kerja Indonesia melalui pelatihan berbasis kompetensi, uji kompetensi serta sertifikasi keahlian. Untuk menyelenggarakan uji kompetensi, UPT Balai Latihan Kerja Surabaya bekerja sama dengan Lembaga Sertifikasi Profesi (LSP), Badan Nasional Sertifikasi Profesi (BNSP).

2. METODE PENELITIAN

Metode penelitian yang digunakan di dalam penelitian ini mengikuti tahapan yang ada di *Knowledge Data Discovery* (KDD) dapat dilihat pada gambar 2.1.



Gambar 2.1 Alur Penelitian

2.1 Pemahaman masalah

Bagian ini mengharuskan peneliti untuk memahami masalah dan solusi apa yang mungkin diusulkan. Selain itu juga di tahap ini peneliti menentukan kesenjangan penelitian (*gap*) yang dicoba untuk dipecahkan. Tujuan dari penelitian ini adalah untuk menentukan apakah ada pola tersembunyi yang dapat di atasi dari data survey kepuasan masyarakat yang digunakan untuk memprediksi serta mencari tingkat akurasinya.

2.2 Pengumpulan Data

Data yang diporleh penulis merupakan data yang diberikan oleh pihak UPT Balai Latihan Kerja Surabaya, yang dimana data tersebut sebelumnya diperoleh dengan menyebarkan kuisioner kepada masyarakat dan perserta pelatihan di UPT Balai Latihan Kerja. Sehingga dapat digunakan sebagai acuan penelitian dalam menganalisa, membandingkan, serta memprediksi kepuasan masyarakat menggunak 3 perbandingan algoritma, yaitu *Decision Tree, Random Forest, K-Nearest Neighbor*. Tahapan ini juga merupakan tahapan yang sangat penting karena di dalamnya terdapat proses penentuan atribut apa saja yang akan digunakan sebagai data latih (*training data*).

2.3 Persiapan Data (*Data Preparation*)

Data *Preparatinon* (persiapan data) adalah proses dimana menyiapkan data mentah yang didapatkan sehingga layak untuk diproses dan dianalisa lebih lanjut. Ada pula langkah – langkah awal, yaitu melakukan pengumpulan data, pembersihan data (data *cleaning*) dan pelabelan data mentah ke dalam bentuk yang cocok sehingga dapat diproses oleh algoritma *machine learning.*

2.4 Data Mining

Data mining merupakan suatu langan dalam *Knowledge Discovery in Database (KDD). Knowledge Discovery* sebagai suatu proses terdiri atas pembersihan data (data *cleaning*), intergrasi data (data *intergation*), pemilihan data (data selection), transformasi data (*data transformation*), data *mining*, evaluasi pola (*pattern evaluation*) dan penyajian pengetahuan *(knowledge presentation*). Data mining mengacu pada proses untuk menambang (*mining*) pengetahuan dari sekumpulan data yang sangat besar. Data mining sendiri dibagi menjadi 2 yaitu data training, data testing.

1. Data Training

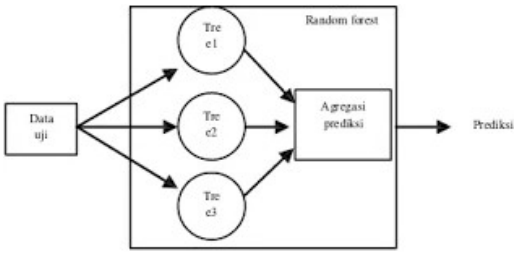
Data Training yang sudah ada akan digunakan untuk membuat prediksi atau menjalankan fungsi dari sebuah algoritma *Machine Learning.*

2. Data Testing

Data Testing adalah bagian dataset yang kita tesuntuk melihat keakuratannya, atau dengan kata lain melihat performanya.

2.5 MODEL KLASIFIKASI

Tahapan pemodelan ini dilakukan untuk membangun data training dengan menggunakan pelatihan algoritma sehingga menghasilkan sejumlah aturan. Pada penelitian



ini, menggunakan algoritma *Decision Tree*, selain itu, algoritma lain yang digunakan untuk melakukan perbandingan, yaitu algoritma *Random Forest, K-Nearest Neighbor*.

2.1 Decision Tree

Decision tree merupakan teknik model prediksi yang dapat digunakan untuk klasifikasi atau regresi model dalam bentuk struktur pohon. Proses pada decision tree adalah dengan cara memecah terus ke dalam himpunan bagian yang lebih kecil lalu pada saat itu juga sebuah pohon keputusan secara bertahab dikembangkan. Hasil dari proses tersebut adalah pohon dengan *node* keputusan dan *node* daun.

2.2 Random Forest

Random forest merupaka salah satu metode dalam *Decision tree. Decision Tree* atau pohon pengambil keputusan adalah sebuah diagram alir yang berbentuk seperti pohon yang memiliki sebuah *root node* yang berisi tentang pertanyaan, tentang data dan sebuah *leaf node* yang digunakan untuk memcahkan masalah serta membuat keputusan.

Random forest adalah kombinasi dari masing-masing tree yang baik kemudian dikombinasikakn ke dalam satu model. *Random Forest* bergantung pada sebuah nilai *vector random* dengan distribusiyang sama pada semua yangmasing-masing decision tree memiliki kedalmn yang maksimal. *Random forest* adalah classifier yang terdiri dari classifier yang berbentuk pohon {h(x,θ k), k = 1,) dimana θk adalah ramdom vector yang didistribusikan secara independen dan masing- masing tree pada sebuah unit kan memilih class yang paling popular pada input x.

3.1 Memusatkan *random forest*

Terdapat *classifier* h1(x), h2(x), …. hk (x) dan dengan *training set* dari distribusi random *vector* Y,X, berikut fungsi yang terbentuk

$$Mg(X,Y) - \max_{j \neq y} av_k I(h_k(X) = Y) - \max_k av\ I(h_k(X) = j) \quad (1)$$

Fungsi *error* yang digunakan

$$PE^* = P_{X,Y}(mg(X,Y) < 0) \quad (2)$$

Hasil dari penggabungan fungsi

$$P_{X,Y}(P\theta(h(X,\theta) = Y) - \max_{j \neq y} P\theta(h(X,\theta) = j) < 0) \quad (3)$$

Gambar 2.1 Proses Prediksi Algoritma RandomForest

2.3 K-Nearest Neighbor

K-Nearest Neighbor merupakan salah satu metode klasifikasi dalam data mining yang termaasuk ke dalam supervised learning. Pengklasifikasian yangdilakukan berdasarkan atribut dan data training, sehingga proses pengklasifikasian data baru dilakukan berdasarkan perbandingan kemiripan mayoritas pada data training. Dalam *K-Nearest Neighbor* nilai jarak ditentukan dengan pengujan data testing terhadap data training , kemudian menggunakan nilai terkecil dari nilai ketetanggan terdekat. Penghitungan jarak umumnya menggunakan jarak *Euclidean Distance* sebagai berikut.

$$d\ (x_i, x_j = \sqrt{\sum_{r=1}^{n} (a_r(x_r) - a_r((x_j))^2} \quad (4)$$

Dimana d($x_i x_j$) merupakan jarak Euclidean, x_i dan x_j adalah dua titik yang akan dihitung jaraknya dengan x_i merupakan *record* data i dan x_j adalah record data ke j, serta r adalah data ke r dengan i,j bernilai 1,2,3,n

2.5 Pengujian Model

Pola atau model yang telah didapatkan dan dikembangkan dalam penelitian ini akan diterapkan pada data hasil survey kepuasan masyarakat khususnya pada sarana prasarana yang ada di UPT Balai Latihan Kerja Surabaya menggunakan *Jupyter Annaconda 3.* Dataset terdiri dari *workshop bengkel, ruang teori, litrik, toilet, dan sarana penunjang.*

Sebanyak 70% data latih (data *training*) yang akan digunakan untuk proses prediksi. Sedangkan 30%dari *dataset* akan digunakan sebagai data uji (data *testing*)

2.6 EVALUASI

Pada tahapan evaluasi, dilakukan pengujian menggunakan *cross validation* dari algoritma yang digunakan untuk klasifikasi, hasil pengujian tersebut akn dibandingkan nili *accuracy, precision, dan recall* dari algoritma *Decision Tree, Random Forest, dan K-Nearest Neighbor*. Selain itu, pengujian juga dilanjutkan dengan metode *confusion matrix* untuk melihat mana algoritma yang terbaik.

a. Confusion Matrix

Confusion Matrix adalah pengukuran terhadap kinerja suatu sistem klasifikasi merupakan hal yang penting. Kinerja sistem klasifikasi menggambarkan seberapa baik sistem dalam mengklasifikasikan data. *Confusion Matrix* merupakan salah satu metode yang dapat digunakan untuk mengukur suatu metodeklasifikasi. Pada dasarnya *confusion matrix* mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya.

No	Nilai Interval	Klasifikasi
1	25 - 64	Tidak Baik
2	65 - 76	Kurang Baik
3	76 - 88	Baik
4	88 - 100	Sangat Baik

Berdasarkan jumlah keluaran kelasnya, sistem klasifikasi dapat dibagi menjadi 4 jenis yaitu *klasifikasi binary, multi-class, multi-label* dan *hierarchical.* Pada klasifikasi *binary*, data masukan dikelompokkan ke dalam salah satu dari dua kelas. Jenis klasifikasi ini merupakan bentuk klasifikasi yang paling sederhana dan banyak digunakan.

Sementara itu, pada bentuk klasifikasi *multi- class*, data masukan diklasifikasi menjadi beberapa kelas. Contohnya dalam penelitian ini, baik,tidak baik, sangat baik, dan kurang baik. Bentuk klasifikasi *multi-label* pada dasarnya hampir sama dengan klasifikasi *multi-class* dimana data dikelompokan menjadi beberapa kelas, namun pada klasifikasi *multi-label*, data dapat dimasuklan dalam beberapa kelas sekaligus. Bentuk klasifikasi yang terakhir adalah *hierarchichal*.

b. Cross Validation

Cross Validation adalah sebuah pengujian standar yang berfungsi untuk memprediksi *error rate*. Jumlah data training dan data testing yang mewakili setiap kelas harus memiliki jumlah yang sama, pembagian data dilakukan secar acak dengan perbandigan yang sama pada setiap kelasnya. Tingkat kesalahan pada setiap tingkat iterasi akan dihitung secara rata-rata sehingga dapat menghasilkan *error rate* keseluruhan.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Pengujian

Pengujian dilakukan terhadap UPT Balai Latihan Kerja Surabaya yang dinilai pelayanan dan sarana prasarana untuk mendapatkan tingkat akurasi yang terbaik dari keseluruhan data sampel yang di uji.

Agar penelitian ini mendapatkan hasil yang diharapkan maka penulis memutuskan untuk melakukan proses pengklasifikasian

data sampel menjadi 4 interval nilai dan data variabel yang digunakan adalah workshop bengkel, ruang teori, kelistrikan, kamar mandi, dan sarana prasarana penunjang maka dapat dilihat pada tabel 3 berikut ini.

No	WORKSHOP / BENGKEL						RUANG TEORI				LISTRIK		TOILET				SARANA PENUNJANG		
	1	2	3	4	5	6	1	2	3	4	1	2	1	2	3	4	1	2	3
1	3	3	4	3	2	4	4	4	4	4	4	4	4	2	2	3	4	4	4
2	5	5	4	4	4	4	5	4	5	4	5	5	4	5	4	4	4	4	4
3	5	5	5	5	5	5	5	5	5	5	5	5	4	3	4	3	5	4	5
4	5	5	5	5	5	5	5	5	5	4	5	5	3	3	3	3	5	5	5
5	5	5	5	4	4	5	5	4	5	4	4	4	4	2	3	2	4	4	3

Tabel 3 Dataset dan Klasifikai Nilai Kuisioner

Tabel 3 Dataset dan Klasifikai Nilai Kuisioner

3.2. Pengujian Data Dengan Metode *DecisionTree*

a. Confusion Matrix

Pengujian dilakukan terhadap UPT Balai LatihanKerja Surabaya yang dinilai pelayanan dan sarana prasarana untuk mendapatkan tingkat akurasi yang terbaik dari keseluruhan data sampel yang di uji.

Berikut *confusion matrix* yang diperoleh dari algoritma *Decision Tree* berdasarkan pengujian data lama dan data baru menggunakan tools python dapat dilihat pada gambar 3.1.

	precision	*recall*	*accuracy*
Tidak Baik	1.00	1.00	0.95
Kurang Baik	0.86	1.00	
Baik	1.00	0.86	
Sangat Baik	0.94	1.00	

Tabel 3.1 Hasil Keakuratan Prediksi

Pada gambar 3.1 diatas, Menghitung akurasi dari hasil prediksi klasifikasi dengan menggunakan *confussion matrix.* Oleh karena itu perlu melakukan import library *sklearn.metrics* untuk fungsi *classification_report*, *confusion_matrix, maka memperoleh hasil hasil accuracy 0.77*.

b. Cross Validation

Berikut *cross validation* yang diperoleh dari algoritma *Decision Tree* berdasarkan pengujian menggunakan tools python dapat dilihat pada gambar 3.2.

	precision	*recall*	*accuracy*
Tidak Baik	0.00	0.00	0.91
Kurang Baik	0.60	0.75	
Baik	0.92	0.92	
Sangat Baik	1.00	1.00	

Tabel 3.2 Hasil Prediksi Cross Validaiton DecisionTree

Pada gambar 3.2 diatas, Menghitung akurasi dari hasil prediksi klasifikasi dengan menggunakan *cross validation, maka memperoleh hasil hasil accuracy 0.91* .

3.3. Pengujian Data Dengan Metode *RandomForest*

a. Confusion Matrix

Pengujian dilakukan terhadap UPT Balai Latihan Kerja Surabaya yang dinilai pelayanan dan sarana prasarana untuk mendapatkan tingkat akurasi yang terbaik dari keseluruhan data sampel yang di uji.

Berikut *confusion matrix* yang diperoleh dari algoritma *Random Forest* berdasarkan pengujian data lama dan data baru menggunakan tools pythondapat dilihat pada gambar 3.3.

Tabel 3.3 Confusion Matrix Random Forest

	precision	*recall*	*accuracy*
Tidak Baik	0.50	0.50	0.77
Kurang Baik	0.60	0.60	
Baik	0.80	0.71	
Sangat Baik	0.82	0.93	

Pada gambar 3.3 diatas, Menghitung akurasi dari hasil prediksi klasifikasi dengan menggunakan *confussion matrix.* Oleh karena itu perlu melakukan import library *sklearn.metrics* untuk fungsi *classification_report*, *confusion_matrix, makamemperoleh hasil hasil accuracy 0.95*

b. Cross Validation

Berikut *cross validation* yang diperoleh dari algoritma *Decision Tree* berdasarkan pengujian menggunakan tools python dapat dilihat pada gambar 3.4

	precision	recall	accuracy
Tidak Baik	1.00	1.00	1.00
Kurang Baik	1.00	1.00	
Baik	1.00	1.00	
Sangat Baik	1.00	1.00	

Tabel 3.4 Hasil Prediksi Cross Validaiton RandomForest

Pada gambar 3.4 diatas, Menghitung akurasi dari hasil prediksi klasifikasi dengan menggunakan *cross validation,, maka memperoleh hasil hasil accuracy 1.00* .

3.4. Pengujian Data Dengan Metode *K-Nearest Neighbor*

a. Confusion Matrix

Metode	Precision				Recall				Accuracy
	TB	KB	B	SB	TB	KB	B	SB	
Decision Tree	0.50	0.60	0.80	0.82	0.50	0.60	0.71	0.93	0.77
Random Forest	1.00	0.86	1.00	0.96	1.00	1.00	0.86	1.00	0.95
K-Nearest Neighbor	0.67	1.00	0.86	0.78	1.00	0.80	0.75	0.88	0.82

Pengujian dilakukan terhadap UPT Balai Latihan Kerja Surabaya yang dinilai pelayanan dan sarana prasarana untuk mendapatkan tingkat akurasi yang terbaik dari keseluruhan

Metode	Precision				Recall				Accuracy
	TB	KB	B	SB	TB	KB	B	SB	
Decision Tree	0.00	0.60	0.92	1.00	0.00	0.75	0.92	1.00	0.91
Random Forest	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
K-Nearest Neighbor	0.67	0.80	0.75	1.00	0.67	0.67	1.00	0.94	0.88

data sampel yang di uji.

Berikut *confusion matrix* yang diperoleh dari algoritma *K-Nearest Neighbor* berdasarkan pengujian menggunakan tools python dapat dilihat pada gambar 3.5.

Tabel 3.5 Confusion Matrix K-Nearest Neighbor

Pada gambar 3.5 diatas, Menghitung akurasi dari hasil prediksi klasifikasi dengan menggunakan *confussion matrix.* Oleh karena itu perlu dilakukan import library *sklearn.metrics* untuk fungsi *classification_report*, *confusion_matrix, makamemperoleh hasil hasil accuracy 0.94* .

b. Cross Validation

Berikut *cross validation* yang diperoleh dari algoritma *Decision Tree* berdasarkan pengujian menggunakan tools python dapat dilihat pada gambar 3.6.

	Precision	Recall	Accuracy
Tidak Baik	0.67	0.67	0.88
Kurang Baik	0.80	0.67	
Baik	0.75	1.00	
Sangat Baik	1.00	0.94	

Tabel 3.6 Hasil Prediksi Cross Validaiton K-NearesNeighbor

Pada gambar 3.6 diatas, Menghitung akurasi dari hasil prediksi klasifikasi dengan menggunakan *cross validation, maka memperoleh hasil hasil accuracy 0.94.*

3.5 Pembahasan

Hasil penilitian dengan menggunkan *Jupyter Annaconda 3* ini menunjukan algoritma *Random Foret* dengan performa terbaik yang mendapat nilai akurasi sebesar 0.95. Algoritma *Decision Tree* dengan nilai akurasi sebesar 0.77. Algoritma *K-Nearest Neighbor* mendapat nilai akurasi sebesar 0.82. Nilai akurasi tersebut terdapat dengan menggunakan data Indeks Kepuasan Masyarakat terhadap sarana prasaran. Hasil dari penelitian ini dapat dilihat pada tabel 3.7 untuk melihat hasil perbandingan akurasi pada masing-masing algoritma.

Tabel 3.7 Hasil Pengujian Confusion Matrix

Keterangan :
B = Baik
KB = Kurang Baik
SB = Sangat Baik
TB = Tidak Baik

Tabel 3.8 Hasil Pengujian Cross Validation

Keterangan :
B = Baik
KB = Kurang Baik
SB = Sangat Baik
TB = Tidak Baik

	Precision	Recall	Accuracy
Tidak Baik	0.67	1.00	0.82
Kurang Baik	1.00	0.80	
Baik	0.86	0.75	
Sangat Baik	0.78	0.88	

Dari hasil tabel 3.8 diatas perbandingan performa akurasi dengan menggunakan *Cross Validation,* terlihat algoritma *Decision Tree* memiliki nilai persentase 0.91. Algoritma *K-Nearest Neighbor* memiliki nilai persentase 0.88. Sedangkan algoritma *Random Forest* memiliki nilai persentase paling tinggi yaitu 1.00. Dapat dikatakan penggunaan algoritma *Random Forest* lebih baik daripada *Decision tree*, *K-Nearest Neighbor* untuk jumlah data yang terbatas.

3. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian dan pengujian yang telah dilakukan oleh peneliti mengenai perbandingan antara algoritma *Decision Tree, Random Forest, dan K-Nearest Neighbor*, melakukan klasifikasi terhadap 128 data survey kepuasan masyarakat terhadap UPT Balai Latihan Kerja Surabaya, maka dapat diperoleh sebuah kesimpulan dan saran, sebagai berikut :

a. Hasil accuracy menunjukan hasil yang berbeda tergantung jenis data, nilai k-fold dan jumlah instance, label class, percentage split yang digunakan pada penelitian ini menggunakan percentage split, 70% data training dan 30% data testing.
b. Algoritma Klasifikasi Random Forest merupakan algoritma yang paling cocok digunakkan di anatara algoritma Decision
Tree dan algoritma K-Nearest Neighbor. Hasil akurasi Radom Forest 1.00, sedangkan algoritma Decision Tree 0.91 dan algoritmaK- Nearest Neighbor dengan hasil 0.94.
c. Saran dalam penelitian ini. Agar penulis lain melakukan pengujian dan perbandingan menggunakan algoritma yang lain untuk mendapatkan pengetahuan dan ilmu yang lebih luas lagi.

4. DAFTAR PUSTAKA

[1] Ardiansyah, Rahayuningsih, P.A., & Maulana, R.(2018). (Data Mining Untuk Dataset Blogger dengan Eapid Miner. *Jurnal Khatulistiwa Informatika, VI(1), 20-28*

[2] Mariana Windarti, Agustinus Suradi.(2018).Perbandingan Kinerja 6 Algoritme Klasifikasi Data Mining Untuk Prediksi Masa Studi Mahasiswa. Universitas Widya Dharma Klaten

[3] Yeyen Dwi Atma, Arif Setyanto. (2018). Perbandingan Algoritma c4.5 dan K-NN Dalam Identifikasi Mahasiswa Berpotensi Drop Out.Universitas Amikom Yogyakarta. METIK JURNAL Vol.2, No.2, 2018.

[4] Ahmad Syukron, Agus Subekti.(2018).Penerapan Metode Random Over-Under Sampling dan Random Forest Untuk Klasifikasi Penilaian Kredit: AMIK BSI Yogyakarta.

[5] Arie Yandi Saputra, Yogi Primadasa.(2018). Penerapan Teknik Klasifikasi Untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma K- Nearest Neighbor: STMIK Bina Nusantara Jaya Lubuklinggu.

[6] Giat Karyono.(2016). Analisa Teknik Data Mining ”Algoritma C4.5 Dan K-Nearest Neighbor” Untuk Mendiagnosa Penyakit Diabetes Mellitus: STMIK AMIKOM Purwokerto.

[7] Imam Sutoyo.(2018). Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Perserta Didik: Universitas Bina Sarana Informatika Jakarta.

[8] Danny Sebastian.(2019). Implementasi Algoritma K-Nearest Neighbor Untuk Melakukan klasifikasi Produk Dari Beberapa E-Marketplace: Universitas Kristen Duta Wacana Yogyakarta.

[9] Nur Khotimah, Deden Istiawan. (2018). Perbandingan Algoritma C4.5 dan K-Nearest Neighbor Untuk Prediksi Lahan Kritis di Kabupaten Pemalang. Akademi Statistika MuhamadiyahSemarang.

[10] Dogan, N. & Tanrikulu, Z. (2013). A comparative analysis of classification algorithms in data mining for accuracy, speed and robustness. *Information Technology an Management.*