POLA BERBASIS UNTUK MENENTUKAN JENIS DAN KEGUNAAN SPECIAL FAT DENGAN METODE NAIVE BAYES, DECISION TREE, NEURAL NETWORK DAN STANDARD PERUSAHAAN

Abdul Halim Anshor

Program Studi Teknik Informatika Sekolah Tinggi Teknologi Pelita Bangsa
abdulhalim.anshor15@gmail.com

Disetujui, 21 Agustus 2017

Abstrak

Bagi perusahaan PT. Wilmar Cahaya Indonesia, Tbk kualitas fat adalah yang paling utama. Faktanya bahwa pengecekan kualitas fat memerlukan waktu yang sangat lama, pengecekan kualitas fat melalui serangkaian tes uji dengan menggunakan parameter Kadar Air (KA) atau Moisture & Impurities selama 1 jam, FFA (Free Fatty Acids) selama 10 menit, IV (Iodine Value) selama 45 menit dan MPT (Melting Point) selama 1 jam. Serta pengecakan terakhir yaitu pengecekan SFC (Solid Fat Content) yang berfungsi sebagai final penecekan memerlukan waktu 5.35 jam. Untuk itulah penulis melakukan penelitian terhadap penerapan metode machine learning untuk menentukan kualitas fat tersebut. Maka dalam penelitian ini penulis akan mencari pola berbasis untuk menentukan jenis dan kegunaan special fat dengan metode machine learning (metode Naïve Bayes, Decision Tree dan Neural Netwok). Adapun metode yang digunakan pada penelitian kali ini antara lain, Naive Bayes Classifier, Decicion Tree dan Neural Network. Dari ketiga metode tersebut akan dibandingkan tingkat efektifitasnya untuk mendapatkan akurasi, efisiensi dan kecepatan waktu dalam menentukan keputusan. Penulis melakukan serangkaian uji coba dengan menggunakan tool WEKA 3.7.4, dimana dalam proses klasifikasi penulis menggunakan data training set, Supplied set (data training dan data uji), serta Cross Validation (data uji yang dijadikan data training dan data uji) untuk mendapatkan nilai akurasi terbaik.

Kata Kunci: *Naive Bayes Classifier*, *Decicion Tree* dan *Neural Network*.

Abstract

For PT. Wilmar Cahaya Indonesia, Tbk quality fat is the most important. The fact that fat quality checks take a long time, checking the quality of fat through a series of test tests using Moisture & Impurities for 1 hour, FFA (Free Fatty Acids) for 10 minutes, IV (Iodine Value) during 45 minutes and MPT (Melting Point) for 1 hour. And the last check is the SFC (Solid Fat Content) which serves as a final checking takes 5.35 hours. For that reason the authors do research on the application of machine learning method to determine the quality of fat.So in this study the authors will look for pattern-based to determine the type and usefulness of special fat by machine learning method (Naïve Bayes method, Decision Tree and Neural Netwok). The methods used in this study include, Naive Bayes Classifier, Decicion Tree and Neural Network. Of the three methods will be compared the level of effectiveness to obtain accuracy, efficiency and speed of time in determining decisions. The author performs a series of trials using the WEKA 3.7.4 tool, which in the classification process the author uses the data training set, Supplied set (training data and test data), and Cross Validation (test data used as training data and test data) to get the value best accuracy.

Keywords: *Naive Bayes Classifier*, *Decicion Tree* and *Neural Network.*

1. Pendahuluan

Permasalahan kualitas dan produktivitas merupakan hal yang sangat penting dalam setiap kegiatan perusahaan manufacturing. Hal ini tentu akan berpengaruh pada kelanjutan hidup dan perkembangan perusahaan tersebut. Sebagai contoh perkembangan perusahaan di PT. Wilmar Cahaya Indonesia Tbk ditandai dengan meningkatnya angka permintaan akan special fat. Special fat merupakan bahan baku yang terbuat dari minyak sawit, special fat juga terdiri dari berbagai macam jenis yang memiliki banyak kegunaan dalam kehidupan sehari-hari. Seperti pemakaian special fat pada berbagai jenis produk makanan dan minuman serta kebutuhan rumah tangga (seperti sabun, sampo) dan lain sebagainya. Dengan angka permintaan yang tinggi perusahaan perlu melakukan suatu langkah efektif untuk meningkatkan kualiatas dan produktivitasnya. Untuk hal itu perusahaan akan menciptakan standar nilai mutu special fat yang dapat menjadi basis atau dasar dalam menghasilkan special fat yang berkualitas, tentunya dengan kualitas yang baik akan membantu peningkatan produktivitas.

Kualitas suatu special Fat dapat ditentukan dengan menggunakan parameter kadar air (KA), free fatty acid (FFA), iodine value (IV), melting point (MPT) dan solid 30oC. Dengan menggunakan parameter tersebut perusahaan akan membuat standard baku dari kualitas special fat. Semakin tinggi tingkat produktivitas dan berjalannya waktu perusahaan akan mendapatkan beberapa masalah diantaranya kesulitan dalam menangani masalah relevansi dan akurasi kualitas produk yang sudah dibuat dengan produk-produk baru yang muncul dan potensi market yang baik. Kualitas produk yang baik tidak cukup hanya menggunakan parameter KA, FFA, IV dan MPT. Tetapi diperlukan sebuah test lagi yaitu tes SFC (Solid Fat Content) baik tempering ataupun non tempring dengan menggunakan alat yang disebut NMR. SFC merupakan suatu proses pengukuran mesin yang dilakukan dengan urutan langkah sebagai berikut, dimana special fat dimasukkan kedalam tube setinggi 5cm dan dimasukkan pada water bath dengan suhu 60oC selama 30 menit, kemudian diangkat dan dimasukkan pada water bath dengan suhu 0oC selama 90 menit kemudian diangkat dan dimasukkan lagi pada water bath dengan suhu 10 oC, 15 oC, 20 oC, 25 oC, 30 oC, 35 oC, 40 oC, 45 oC dengan interval waktu masing-masing suhu 30 menit dan setiap perpindahan suhu pada water bath dilakukan pengukuran, tiap sampel dimasukkan ke dalam mesin NMR pada tempat yang telah disediakan. Sebelumnya tiap sampel dilap dan dibersihkan. NMR akan mengeluarkan nilai prosentase solid fat content. NMR adalah mesin yang selain mengukur solid fat content juga menentukan tekstur, kesegaran (fresh), jenis dan kegunaan special fat. Dari paparan di atas Proses NMR (Non Tempring) memerlukan waktu 5.35 jam untuk setiap produk, tentu saja proses tersebut akan memakan waktu yang sangat lama. Selain itu dalam pengecekan dan nilai yang dihasilkan NMR (Non Tempring) tidak ada standar baku yang digunakan. Dari raw material fat yang sama kemudian melewati proses hydrogenasi dan mendapatkan nilai analisa KA, FFA, IV yang sama tetapi bisa mendapatkan nilai NMR (Non Tempring) yang berbeda. Sehingga penulis mencoba untuk menggunakan metode machine learning dengan tiga metode (Naive Bayes, Decision Tree, dan Neural network) dimana dengan metode ini akan dicari pola berbasis melalui data trening dan data uji untuk mentukan standar kualitas special fat, dan membandingkan ketiga metode machine learning tersebut, serta menentukan metode yang tepat digunakan. Diharapkan dengan metode yang penulis usulkan diperoleh perbaikan yang signifikan dalam akurasi data, kecepatan dan ketepatan data uji. Penulis berharap metode yang diusulkan akan menghasilkan hasil keputusan yang tepat dan cepat..

2. Tinjauan Pustaka

2.1. *Machine Learning*

Machine Learning adalah metode analisis data yang mengotomatisasi bangunan model analisis. Menggunakan algoritma yang iteratif belajar dari data, pembelajaran mesin memungkinkan komputer untuk menemukan wawasan tersembunyi tanpa eksplisit diprogram mana mencarinya. Teknologi Machine Learning memiliki banyak kegunaan diantaranya untuk mendetesi spam email, rekomendasi buku, pendeteksian wajah dan ekspresi lainnya serta menentukan pola suara. Pada media social digunakan untuk deteksi pencurian online. Menentukan prediksi akurasi kualitas produk, prediksi kebutuhan market dan banyak lagi (Tom Mitchell, 1997).

Ada beberapa metode yang digunakan dalam teknologi Machine learning diantaranya :

1.Naive Bayes classifier
2.Neural Networks
3.Decision Tree

2.2. *Naive Bayes Classifier*

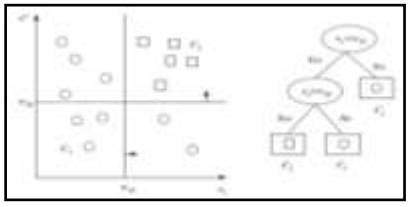
Naive Bayes merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan toerema *Bayes* dengan asumsi independensi (ketidaktergantungan) yang kuat.

Dengan kata lain , *Naive Bayes*, model yang digunakan adalah "model fitur independen".
Dalam metode *Naive Bayes*, maksud independensi yang kuat pada fitur adalah bahwa sebuah fitur pada sebuah data tidak berkaitan dengan ada atau tidaknya fitur lain dalam data yang sama (Tom Mitchell,1997). Prediksi Bayes didasarkan pada toerema bayes dengan formula umum sebagai berikut :

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

2.3. *Decision Tree*

Decision Tree merupakan salah satu metode klasifikasi yang menggunakan representasi struktur pohon (tree) dimana setiap node merepresentasikan atribut. Cabangnya merepresentasikan nilai dari atribut, dan daun merepresentasikan kelas. Node yang paling atas dari Decision Tree disebut sebagai root. Decision Tree merupakan metode klasifikasi yang paling populer digunakan. Selain karena pembangunannya relatif cepat, hasil dari model yang dibangun mudah untuk dipahami (Tem Alpadyne, 1995).
Pada Decision Tree terdapat 3 jenis node, yaitu:
a. Root Node, merupakan node paling atas, pada node ini tidak ada input
dan bisa tidak mempunyai output atau mempunyai output lebih dari
satu.
b. Internal Node , merupakan node percabangan, pada node ini hanya
terdapat satu input dan mempunyai output minimal dua.
c. Leaf node atau terminal node , merupakan node akhir, pada node ini
hanya terdapat satu input dan tidak mempunyai output.



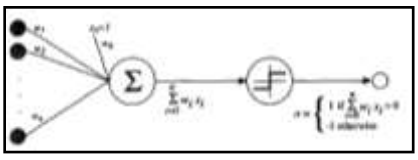
Gambar 1. Contoh *Dataset* dan *Decesion Tree*nya (Sumber : Alpadyne Etem 1995."*Introduction to Machine Learning Second Edition" London: The MIT Press*, pp.186.)

Decision Tree dapat berkembang menjadi besar dalam analisa proses . Jika akurasi dari pengetesan tidak akurat tree bisa besar. Untuk itu diperlukan pemilihan atribut yang paling berguna dalam klasifikasi sampel, ketika atribut yang penting terpilih setiap node perlu di tes. Hal ini memerlukan perhitungan gain pada setiap node. Informasi gain mengukur seberapa baik atribut memberikan pengaruh pada sampel training menurut target klasifikasi, kemudian hasil pengukuran ini digunakan untuk memilih atribut pada setiap pertumbuhan tree. Untuk menghitung informasi gain untuk setiap node, kita harus tahu entropi setiap node, di mana entropi adalah pengukuran homogenitas set sampel. Mengingat sampel beberapa sasaran konsep (dua masalah, entropi set s relatif terhadap klasifikasi biner ini adalah sebagai berikut (Etem Alpadyne, 1995) :
Penjelasan dari formula tersebut adalah sebagai berikut :
Entropy (S)=$-p_+ \log_2 p+ - p_- \log_2 p_-$ ……...[2.5]

Keterangan :
S : adalah ruang (data) sample yang digunakan untuk training.
P+ : adalah jumlah yang bersolusi positif (mendukung) pada data sampel
untuk kriteria tertentu.
P- : adalah jumlah yang bersolusi negative (tidak mendukung) pada data
sampel untuk kriteria tertentu.

2.4. *Neural Network*

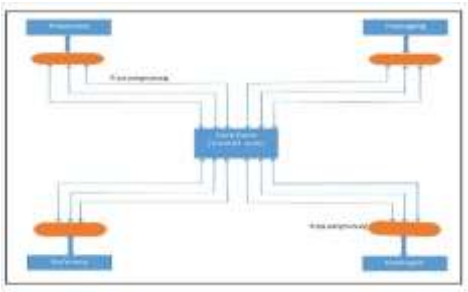
Pembuatan jaringan saraf tiruan terbentuk dari jutaan (bahkan lebih) dasar neuron yang terinterkoneksi dan terintegrasi antar satu dengan yang lain sehingga dapat melaksanakan aktifitas secara teratur dan terus menerus sesuai kebutuhan. Tiruan neuron dalam struktur jaringan saraf tiruan adalah sebagai elemen pemroses seperti gambar 2.2 yang dapat berfungsi seperti halnya sebuah neuron. Sejumlah sinyal masukkan x dikalikan dengan masing-masing penimbang yang bersesuaian w. kemudian dilakukan penjumlahan dari seluruh hasil perkalian tersebut keluaran yang dihasilkan dilakukan ke dalam fungsi pengaktif untuk mendapat tingkatan derajat sinyal keluarannya F(x,w). Walaupun masih jauh dari sempurna, namun kinerja dari tiruan neuron ini identik dengan kinerja dari sel biologi yang kita kenal saat ini (Tom Mitchell, 1997).



Gambar 2. Unit Perceptron
(Sumber : *Mitchell, Tom. M. 1997."Machine Learning". Portland. McGraw-Hill Science Engineering/Math.pp.96)*

2.5. *Standard* Perusahaan Pengolahan Special Fat

Special fat ditampung pada tanki-tanki besar yang disebut dengan Tank Farm yang berjumlah 62 tanki. Tanki tersebut terhubung dengan pipa – pipa untuk memasuki proses pembuatan bahan fat sampat menjadi produk jadi. Proses pembuatan dan pengukuran kualitas fat dapat diilustrasikan seperti gambar berikut



Gambar 3. Jalur Proses Antar Produksi

Secara garis besar ada lima departemen yang mengolah dan memproses jalannya produksi di PT. Wilmar Cahaya Indonesia, Tbk yaitu :

1.Departemen Tank Farm

Adalah departemen yang bertugas melayani semua departemen proses produksi kemudian menerima kembali hasil proses produksi dan menyimpan serta menampung pada tanki-tanki yang telah disediakan. Pada proses inilah terjadi kerentanan atau kemungkinan terjadi kesalahan tercampurnya fat dengan fat yang lain, terkontaminasi fat dengan fat yang lain sehingga mengurangi serta merusak kualitas hasil produksi.

2.Departemen Fraksinasi

Adalah departemen yang memproses dan mengolah crude fat menjadi dua bagian dengan cara dipres. Hasil dari proses pengepresan yang cair disebut olein dan yang padat disebut stearin. Kemudian hasil tampungannya dikirim kembali ke Tank Farm untuk disimpan dan akan digunakan oleh departemen lain.

3.Departemen Refenery

Adalah departemen yang memproses dan mengolah fat untk merubah warna, menghilangkan bau, menurunkan kadar air serendah mungkin dan menghilangkan gum/ getah dan minyak asam. Hasil produksinya akan dikembalikan kedepartemen tank farm untuk disimpan serta digunakan oleh departemen lain.

4.Departemen Hydrogen

Adalah departemen yang mengolah dan memproses fat untuk menurunkan iodine value yaitu nilai kekerasan fat secara kimiawi, kemudian hasil produksinya dikirim kembali ke departemen tank farm untuk keperluan departemen lain.

5.Departemen Packaging

Adalah departemen yang mengolah dan memproses fat yang telah menjadi akhir proses atau finishing produk untuk selanjutnya dilakukan pengemasan atau pengepakan untuk dikirim ke gudang pendingin dan siap dijual. Namun ada beberapa jenis fat yang menjadi finish produk tetap disimpan di departemen tank farm yang kemudian langsung ditransfer ke mobil tanki atau yang biasa disebut bulky untuk dikirim ke customer.

Dari penjelasan diatas bisa dibayangkan proses transfer fat ke departemen tank farm atau departemen tank farm ke departemen produksi intensitasnya begitu tinggi setiap harinya maka begitu rentan terjadi kesalahan transfer fat. Kemungkinan human error lah yang biasa terjadi :

1.Kelalaian dan kelupaan dalam membuka dan menutup valve.

2.Kelalaian dan kelupaan dalam informasi tujuan transfer fat.

3.Kelalaian dalam mengawasi tanki saat transfer sehingga fat tumpah karena kepenuhan.

Perusahaan Wilmar Cahaya Indonesia, Tbk memiliki beberapa standar dalam menentukan kualitas produk special fat diantaranya sebagai berikut :

1.Tekstur Special Fat :

Tekstur merupakan kondisi fisik dari special fat yang dapat diketahui secara visual, tekstur juga menyangkut masalah tingkat kekentalan, kekerasan dan kelembutan fat. Untuk mengetahui tekstur fat ada beberapa kondisi yang harus dipenuhi sebagai berikut :

a.Jika fat diukur pada suhu 30oC memiliki nilai N30 sama dengan 0 sampai dengan 20 persen maka fat tergolong dalam kategori Liquid Fat.

b.Jika fat diukur pada suhu 30oC memiliki nilai N30 sama dengan 21 sampai dengan 100 persen maka fat tergolong dalam kategori Solid Fat.

2.Jenis Special Fat

PT. Wilmar Cahaya Indonesia,Tbk yang ada di Cikarang mengolah dua jenis fat yaitu Louric dan Palmitic,

a)Louric merupakan jenis fat yang berasal dari Crude Coconut Oil (CCNO) dan fat yang berasal dari sawit bagian intinya yang disebut Crude Kernel Oil (CKO).

b) Palmitic merupakan fat yang berasal dari kelapa sawit bagian serabutnya
yang disebut Crude Palm Oil (CPO).

Dari kedua jenis ini dihasilkan banyak produk kurang lebih 33 jenis dan nama produk yang dijual oleh perusahaan.

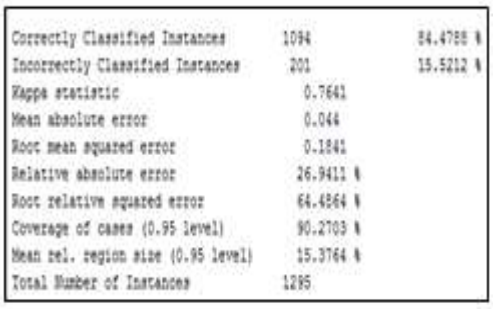
Hasil turunan dari CCNO menghasilkan produk :

1.RCNO (Refine Coconut Oil) dihasilkan dari 100 persen CCNO yang melewati proses refenery (proses menghilangkan bau dan merubah warna lebih jernih).

2.Fonta Cream dihasilkan dari 100 persen CCNO yang melewati proses Hydro (proses untuk menurunkan Iodine Value) dan prose refinery.

2.6. *Pembacaan* Data Hasil Analisa WEKA

2 Data yang dihasilkan dari proses klasifikasi disajikan dalam bentuk tabel , diagram dan gambar. Berikut adalah salah satu contoh dari data output WEKA



Correctly Classified Instances	1094	84.4788 %
Incorrectly Classified Instances	201	15.5212 %
Kappa statistic	0.7641	
Mean absolute error	0.044	
Root mean squared error	0.1841	
Relative absolute error	26.9411 %	
Root relative squared error	64.4864 %	
Coverage of cases (0.95 level)	90.2703 %	
Mean rel. region size (0.95 level)	15.3764 %	
Total Number of Instances	1295	

Gambar 4. *Stratified cross-validation*

Pada data diatas ada beberapa item pokok yang harus diperhatikan sebagai berikut :

1. Correctly Classified Instances
Merupakan jumlah sampel yang memiliki nilai benar. *Correctly Classified Instances* disebut juga sebagai disebut juga sebagai *accuracy*. Persamaan untuk nilai *accuracy* adalah :

Accuracy =(True Positif +True Negatif)/Total sampel .....[2.12]

Dari gambar 2.7 kita dapat menghitung nilai *accuracy*-nya adalah :

Accuracy= (TP+TN)/total = (100+50)/165 = 0.91, dalam persen 91 %

2. Incorrectly Classified Instances
Merupakan jumlah sampel yang memiliki nilai salah. Persamaan untuk *Incorrectly Classified Instances adalah :*
Incorrectly Classified Instances = (Total sampel – Total *correctly Classified Instances*)*/Total sampel* .....
.[2.13]
Dari gambar 2. 7 kita dapat menghitung nilai *Incorrectly Classified Instances* -nya adalah :
Incorrectly Classified Instances = Total sampel - Accuracy = (165 – 150)/165=0.09, dalam persen 9%.

3. Metodologi

Metodologi yang digunakan dalam melakukan penelitian ini adalah :

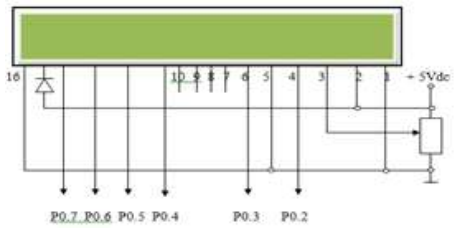
3.1 Observasi Awal

Observasi awal ditujukan untuk mendapatkan gambaran umum informasi dan situasi, Proses dan kebutuhan perusahaan yang diteliti. Observasi langsung dilakukan dengan cara melakukan serangkaian test pengukuran sebagai data awal yakni pengukuran beberapa parameter yang dibutuhkan seperti pengujian KA, IV, FFA, MPT dan Special content N30.
Untuk memahami proses dari data uji yang diambil penulis juga melakukan wawancara dengan beberapa orang staf ahli di departemen terkait, departemene R&D dan departemen QC.
Dalam rangka untuk membuat pembelajaran yang akurat dan ilmiah, beberapa data yang dibutuhkan untuk dikumpulkan itu adalah dibagi menjadi dua tahap selama pengumpulan data dan perhitungan. pengumpulan data dan observasi, tabulasi data dan perhitungan. Tahap pertama pengumpulan data dan perhitungan bertujuan sebagai pedoman dalam memilih topik. Data yang dikumpulkan dalam tahap pertama akan melanjutkan ke beberapa perhitungan data yang akan diperlukan untuk menyelesaikan tesis ini. Tahap berikutnya adalah tabulasi data dan kesimpulan, fase ini bertujuan untuk menganalisis data set disiapkan sebelum dilakukan percobaan. Fase ini terdiri dari kualitas data masing-masing atribut untuk mencegah perhitungan salah perhitungan data selanjutnya, mengkonversi format data dimana data akan dikonversi ke dalam format yang sesuai aplikasi, pembatasan data terus menerus di mana untuk mencegah ribuan data set yang masuk, dan yang terakhir adalah pengaturan kelas di mana akan dipilih dari daftar atribut.
Selama perhitungan data, peneliti menggunakan alat pendukung yang disebut Weka versi 3.7.4

3.2 Eksperimen Klasifikasi Data

Langkah pertama data yang telah disiapkan yang pertama data training, kedua data uji dalm bentuk xls akan dirubah ke dalam bentuk .csv kemudian di running. Data hasil running di rubah kedalam format .arff untuk mengklasifikasi data adalah dengan memilih jenis klasifikasi yang terdapat pada Generic Object Editor dialog box. Klasifikasi yang digunakan pada penelitian ini menggunakan metode supplied data set dimana data training dan data uji berada pada file terpisah. Klasifikasi pada machine learning merupakan pendekatan angka nominal dan numerik. Sedangkan implementasi skema dengan menggunakan grafik .

Langkah selanjutnya melakukan klasifikasi dengan metode :



Gambar 5. Klasifikasi Data

- *Naive Bayes*

Klasifikasi akan menggunakan algoritma klasifikasi naive bayes. Dimana parameter dataset akan diklasifikasikan. Peneliti akan memilih persentasi sebanyak 1295 , 800, 500 data training dan 500, 200 data uji valid.

- *Decision tree*

Analisa *decision tree* akan menggunakan algoritma J48. Hasil analisa akan ditunjukan dengan grafik decision tree

- *Neural Network*

Analisa Neural Network akan menggunakan algoritma klasifikasi kelas *Multilayer percepteron* dan hasil analisa akan ditunjukan dengan grafik jaringan syaraf tiruan.

4. Pembahasan dan Hasil

5.1. Pembahasan

Pada eksperimen tahap 1 peneliti melakukan proses klasifikasi dengan metode supplied test , dimana dalam uji coba data training dan data uji dibuat secara terpisah atau terdiri dari dua file.

4.1.1 *Naive Bayes Classifier*

Data itu dilanjutkan dengan melakukan eksperimen 1295 dari data yang ditetapkan sebagai data training, sedangkan 500 dataset dan 200 dataset sebagai uji validasi, hal ini berarti agar data uji memiliki tingkat akurasi yang lebih tinggi. Output di bawah ini adalah hasil dari model Naive Bayes dengan menggunakan algoritma klasifikasi Bayes.

Correctly Classified Instances	1094	84.4788 %
Incorrectly Classified Instances	201	15.5212 %
Kappa statistic	0.7641	
Mean absolute error	0.044	
Root mean squared error	0.1841	
Relative absolute error	26.9411 %	
Root relative squared error	64.4864 %	
Coverage of cases (0.95 level)	90.2703 %	
Mean rel. region size (0.95 level)	15.3764 %	
Total Number of Instances	1295	

Gambar 6. Stratified Cross-validation Naive Bayes Algorithm Data Training

Hasil dari startified Cross-validation di atas menunjukkan dataset training setelah proses klasifikasi dengan algoritma Naive bayes menghasilkan nilai benar atau akurasi 84.48% (jumlah sampel benar 1094), nilai salah 15.52% (jumlah sampel salah 201). Nilai Kappa yang dihasilkan 0.76 poin, nilai ini menunjukan hasil mendekati nilai sempurna mendekati 1.

Untuk mengetahui nilai error dari dataset training di atas kita dapat melihat dari nilai Mean absolute error (MAE), Root mean square error (RMSE), Relative absolute error (RAE), Root relative squared error (RRSE). Nilai error tersebut mengukur perbedaaan nilai sampel yang diprediksi oleh estimator dengan nilai sampel yang diamati. Pada dataset di atas nilai MAE adalah 0.04, nilai RMSE adalah 0.18, nilai RAE adalah 26.94%, nilai RRSE adalah 64.49 %.

Suatu dataset dapat diestimasi apakah data tersebut terdiri atau mengandung parameter dari populasi dengan mengetahui nilai coverage probability-nya, nilai coverage probabilty diatas menggunakan 0.95 level (95%) yakni level untuk dataset dengan distribusi standar atau dataset dalam jumlah besar. Nilai coverage probabilty dataset di atas adalah 90.27%, berarti dataset memiliki nilai estimasi yang baik.

Correctly Classified Instances	437	87.4 %
Incorrectly Classified Instances	63	12.6 %
Kappa statistic	0.801	
Mean absolute error	0.0382	
Root mean squared error	0.157	
Relative absolute error	24.4013 %	
Root relative squared error	56.318 %	
Coverage of cases (0.95 level)	94.2 %	
Mean rel. region size (0.95 level)	16.428 %	
Total Number of Instances	500	

Gambar 7. *Stratified Cross-validation Naive Bayes Algorithm* Data Uji 1.

Correctly Classified Instances	159	79.5 %
Incorrectly Classified Instances	41	20.5 %
Kappa statistic	0.7046	
Mean absolute error	0.0579	
Root mean squared error	0.190	
Relative absolute error	35.3826 %	
Root relative squared error	69.6645 %	
Coverage of cases (0.95 level)	93 %	
Mean rel. region size (0.95 level)	17.9375 %	
Total Number of Instances	200	

Gambar 8. Stratified Cross-validation Naive Bayes Algorithm Data Uji 2

Hasil dari startified Cross-validation di atas menunjukkan dataset uji 1 (sebanyak 500 sampel) dan dataset uji2 (sebanyak 200 sampel), setelah proses klasifikasi dengan algoritma Naive bayes menghasilkan nilai benar atau akurasi 87,4% (jumlah sampel benar 437) dan 79.5% (jumlah sampel benar 159), nilai salah 12,6% (jumlah sampel salah 63) dan 20.5% (jumlah sampel salah 41). Nilai Kappa yang dihasilkan 0.80 poin dan 0.70, nilai ini menunjukan hasil mendekati nilai sempurna mendekati 1.

Untuk mengetahui nilai error dari dataset uji di atas kita dapat melihat dari nilai Mean absolute error

(MAE), Root mean square error (RMSE), Relative absolute error (RAE), Root relative squared error (RRSE). Nilai error tersebut mengukur perbedaaan nilai sampel yang diprediksi oleh estimator dengan nilai sampel yang diamati. Pada dataset di atas nilai MAE adalah 0.38 dan 0.58, nilai RMSE adalah 0.16 dan 0.19, nilai RAE adalah 24.4% dan 35.38%, nilai RRSE adalah 56.52 % dan 69.66%. Suatu dataset dapat diestimasi apakah data tersebut terdiri atau mengandung parameter dari populasi dengan mengetahui nilai coverage probability-nya, nilai coverage probabilty diatas menggunakan 0.95 level (95%) yakni level untuk dataset dengan distribusi standar atau dataset dalam jumlah besar. Nilai coverage probabilty dataset di atas adalah 94.2% dan 95%, berarti dataset memiliki nilai estimasi yang baik.

Tabel 1. Detailed Accuracy by Class Naive Bayes Data Training

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.742	0.039	0.932	0.742	0.826	0.966	Laurie, Coating
0.934	0.014	0.952	0.934	0.943	0.997	Pelattie, Moulding
0.986	0.006	0.948	0.986	0.967	0.999	Pelattie, Coating
1	0	1	1	1	1	Pelattie, Spread
0.912	0.177	0.783	0.912	0.843	0.951	Laurie, Moulding
0.815	0.02	0.468	0.815	0.595	0.947	Laurie, Spread
1	0	1	1	1	1	Laurie, Bahan pembantu
1	0	1	1	1	1	Pelattie, Bahan pembantu
0.845	0.09	0.862	0.845	0.846	0.964	

Pada tabel di atas dapat diketahui hasil perhitungan dari Detailed Accuracy dataset training didapat nilai TP rate (0.845), nilai FP rate (0.09), nilai precision (0.862), nilai recall (0.845), nilai F-measure (0.846) dan ROC area (0.964). Dataset memiliki nilai ROC area yang baik mendekati 1.

Tabel 2. Detailed Accuracy by Class Naive Bayes Data Uji 1

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.742	0.026	0.948	0.742	0.833	0.981	Laurie, Coating
0.5	0.04	0.091	0.5	0.154	0.97	Laurie, Spread
0.974	0.112	0.883	0.974	0.926	0.975	Laurie, Moulding
0.959	0.032	0.968	0.959	0.957	0.999	Pelattie, Moulding
1	0.002	0.667	1	0.8	1	Laurie, Coating
0.947	0.006	0.957	0.947	0.9	0.998	Pelattie, Coating
1	0	1	1	1	1	Pelattie, Bahan pembantu
1	0	1	1	1	1	Pelattie, Spread
0.974	0.063	0.909	0.974	0.894	0.98	

5. Kesimpulan

Dari penelitian yang telah dilakukan penulis dapat menarik beberapa kesimpulan sebagai berikut :

1. Penggunaan metode *machine learning* untuk menentukan jenis dan kegunaan *fat* memiliki berbagai macam keunggulan bila dibandingkan dengan metode yang sedang berjalan. Dari faktor akurasi penggunaan metode yang sedang berjalan masih adanya tingkat kesalahan yang ditemui dilapangan karena perhitungan dilakukan secara manual dengan menggunakan standar pengecekan, sedangkan dengan metode *machine learning* memiliki nilai akurasi yang lebih baik seperti dengan metode Naive Bayes 87.4 %, *Decision Tree* 96.4% dan *Neural Network* 9.62%. Ketiga metode hampir mendekati akurasi 100%, ini menunjukan metode machine learning bisa dikatakan lebih valid dari metode yang sedang berjalan.

2. Metode machine learning dapat mengatasi masalah pencapaian waktu yang optimal. Seperti diketahui sebelumnya bahwa waktu yang dibutuhkan untuk menentukan jenis dan kegunaan fat sengan metode yang sedang berjalan di perusahaan membutuhkan waktu 5.35 jam untuk per item. Sedang kan dengan metode *machine learning* untuk metode *Naive Bayes* 0.01 detik, *Decision Tree* 0.02 detik, *Neural Network* 9.27 detik. Melihat perbandingan yang sangat signifikan penulis menyimpulkan bahwa metode machine learning lebih efisien dari metode yang sedang berjalan diperusahaan.

3. Salah satu inti dari pengukuran dengan metode klasifikasi *machine learning* adalah untuk mengetahui seberapa besar peluang data uji atau sampel yang mendekati dengan data sebenarnya. Untuk ini dapat kita ketahui dengan melihat nilai ROC area yang dihasilkan oleh aplikasi machine learning dengan tool WEKA 3.7.4. Nilai ROC dikatakan optimal dengan angka 1. Hal ini berarti sampel dikatakan akurat jika memiliki nilai ROC area yang mendekati 1. Dari penelitian di atas ketiga metode memiliki nilai ROC area yang mendekati 1 yakni metode Naive Bayes 0.989 poin untuk data training, 0.999 poin untuk data uji 1, 0.998 poin untuk data uji 2. Metode Decision Tree 0.995 poin untuk data training, 0.977 poin untuk data uji 1, 0.968 untuk data uji 2. Metode neural Network 0.989 poin untuk data training, 0.999 untuk data uji 1, 0.998 untuk data uji 2. Data tersebut menunjukan metode machine learning dapat dijadikan acuan dalam menentukan validasi yang lebih baik.

4. Dari ketiga metode *machine learning*, metode *Decision Tree* dengan algoritma J48 mampu mengatasi masalah akurasi dan waktu yang terbaik, yakni nilai akurasi 96.4 % dan waktu 0.02 detik.

5. Proses klasifikasi dengan menggunakan k=fold cross validasi dilakukan untuk menentukan keakurasian dari data uji, yaitu dengan membagi data uji kedalam beberapa bagian, serta dibandingkan dengan bagian lainnya sebanyak k=fold. Dari uji coba yang dilakukan peneliti menyimpulkan dengan k=fold *cross validation*, maka nilai k=10 fold akan memberikan hasil yang optimal dalam menentukan akurasi data uji.

6. Dengan metode *machine learning* kita dapat mengetahui nilai kontribusi atribut dalam database yakni dengan mengetahui nilai informasi gain yang dihasilkan, dengan ini dapat diketahui atribut yang paling dominan. Pada penelitian ini MPT adalah atribut yang memiliki nilai tertinggi. Dengan mengetahui hal ini maka pengetesan MPT akan lebih ditingkatkan lagi kualitasnya, begitu juga dengan atribut yang lain.

Daftar Pustaka

[1] Alpadydin, Etem. (1995) : Introduction to Machine learning Second Edition. The MIT Press, London.11, 1-6

[2] Bhuvaneswari,R, Kalaiselvi, K. (2012) : Naïve Bayesian Classification Approach in Healthcare Application, International Journal of Computer and telecommunications in Saveetha Engineering Collage Chennai India, 3 R,, 1-7.

[3] Mitchell, Tom. M. (1997) :Machine Learning, McGraw-Hill, Portland, 112-130.

[4] Siltepavet, Anotai. (2012) :Improving Quality of Products in Hard Drive Manufacturing by Decision Tree Technique, IJCSI International Journal of Computer Science Issues, 9, 1-6.

[5] Welling, Max. (2011) :*A First Encounter with Machine Learning*, University of California Irvine, California, 78-88, 215-218.

[6] Standard Operation Procedure PT. WICA INDONESIA, Tbk No.SOP-Q&A-040.

[7] Standard Operation Procedure PT. WICA INDONESIA, Tbk No.SOP-Q&A-03.