

Perbandingan Performansi Teknik Klasifikasi *Breakdown* Mesin pada Proses Produksi Pembuatan *Battery* Mobil

Iveline Anne Marie^{1a}, Lukmanul Hakim^{1b}, Dedy Sugiarto^{2c}, Winnie Septiani^{1d}

Abstract. *Data mining is useful in finding interesting patterns of hidden information in a database with specified algorithms. Management of uncertainty in the automotive industry supply chain, with case data at PT QQQ that produce car batteries, classification techniques are used to manage uncertainty in the case of engine breakdown. Based on the utilization of classification techniques, performance comparison analysis was carried out from several methods, namely Decision Tree, Bagging, Boosting and Random Forest. The research data is divided into testing data (75%) and training data (25%). This study uses Software R for analysis needs. The need for testing the goodness of the model uses package (caret) help to see the value of accuracy, sensitivity and specificity. The analysis shows that the Random Forest and Bagging method is superior compared to the Decision Tree and Boosting methods based on accuracy criteria, while the sensitivity criteria, Bagging and Boosting methods are superior to Random Forest and DecisionTree. The lowest sensitivity value is owned by the Decision Tree Method, which indicates that the ability of the method is weak in predicting very few classes.*

Keywords: *classification technique, accuracy, sensitivity, specificity.*

Abstrak. *Data mining bermanfaat dalam menemukan pola yang menarik dari informasi yang tersembunyi didalam suatu database dengan algoritma-algoritma yang ditentukan. Pengelolaan ketidakpastian pada rantai pasok industri otomotif, dengan data kasus pada PT QQQ yang menghasilkan battery mobil, digunakan teknik klasifikasi untuk mengelola ketidakpastian pada kasus Breakdown mesin. Berdasarkan pemanfaatan teknik klasifikasi, dilakukan analisis perbandingan performansi dari beberapa metode, yakni Decision Tree, Bagging, Boosting dan Random Forest. Data penelitian dibagi menjadi data testing (75%) dan data training (25%). Penelitian ini menggunakan Software R untuk kebutuhan analisis. Kebutuhan pengujian kebaikan model menggunakan bantuan package (caret) untuk melihat nilai akurasi, sensitifitas dan spesifisitas. Hasil analisis menunjukkan bahwa metode Random Forest dan Bagging unggul dibandingkan metode Decision Tree dan Boosting berdasarkan kriteria akurasi, sedangkan untuk kriteria sensitifitas, metode Bagging dan Boosting lebih unggul dibandingkan Random Forest dan DecisionTree. Nilai sensitifitas terendah dimiliki oleh Metode Decision Tree, yang menandakan bahwa kemampuan metode tersebut lemah dalam menduga kelas yang jumlahnya sangat sedikit.*

Kata Kunci: *teknik klasifikasi, akurasi, sensitifitas, spesifisitas.*

I. PENDAHULUAN

Kerangka kerja konseptual risiko rantai pasokan dikembangkan berdasarkan sintesis dari

berbagai sudut pandang literatur, dimana risiko rantai pasok dibagi menjadi dua kategori yaitu risiko makro (bencana yang dapat disebabkan oleh manusia ataupun dikarenakan alam) dan risiko mikro (operasional). Sesuai dengan kerangka kerja konseptual tersebut, risiko rantai pasok didefinisikan sebagai kemungkinan dan dampak dari kejadian atau kondisi yang tidak diduga secara makro dan/atau mikro yang mempengaruhi berbagai bagian dari rantai pasok yang menyebabkan terjadinya kegagalan pada tingkat operasional, taktis, ataupun strategis. Risiko mikro dapat dibagi menjadi empat sub kategori, yaitu: risiko permintaan (*supply side*), risiko manufaktur (*manufacture side*) dan risiko pasokan (*supply side*), serta risiko infrastruktur (*infrastructure side*). Risiko manufaktur mengacu pada kejadian atau situasi buruk pada perusahaan

¹ Jurusan Teknik Industri, Fakultas Teknologi Industri, Universitas Trisakti, Kampus A, Jl. Kyai Tapa No.1, Grogol, Jakarta Barat 11440, Indonesia.

² Jurusan Sistem Informasi, Fakultas Teknologi Industri, Universitas Trisakti, Kampus A, Jl. Kyai Tapa No.1, Grogol, Jakarta Barat 11440, Indonesia.

^a email: iveline.annemarie@trisakti.ac.id

^b email: lukmanul.hakim@trisakti.ac.id

^c email: dedy@trisakti.ac.id

^d email: winnie_septiani@trisakti.ac.id

yang mempengaruhi kemampuan internal mereka untuk menghasilkan barang dan jasa, kualitas dan ketepatan waktu produksi, serta profitabilitas (Ho, 2015).

Data mining adalah sebuah proses pencarian secara otomatis informasi yang berguna dalam tempat penyimpanan data berukuran besar. Teknik-teknik *data mining* dapat digunakan untuk meningkatkan kemampuan sistem dalam menemukan informasi (*information retrieval*). Peran utama dari *data mining* yaitu menemukan pola yang menarik dari informasi yang tersembunyi didalam suatu *database* dengan algoritma-algoritma yang ditentukan. Saat ini, teknik data *mining* banyak digunakan untuk mengumpulkan, mengesktrak data serta menganalisis data dalam jumlah yang sangat besar (Sulastri & Gufroni, 2017). Pemanfaatan teknik-teknik data *mining* makin banyak diminati oleh para peneliti. Beberapa teknik-teknik data *mining* antara lain klasifikasi dan *clustering* (Petree, 2013).

Teknik klasifikasi merupakan proses menemukan model (fungsi) yang mampu menjelaskan dan membedakan kelas-kelas atau konsep, dengan tujuan agar model yang diperoleh dapat digunakan untuk memprediksikan kelas atau objek yang memiliki label kelas tidak diketahui. Model yang diturunkan didasarkan pada analisis dari pelatihan data, yaitu objek data yang memiliki label kelas yang diketahui serta dapat direpresentasikan dalam berbagai bentuk, seperti aturan *if-then* ataupun pohon keputusan (Tan, 2006). Teknik klasifikasi yang saat ini banyak digunakan antara lain adalah *decision tree*. Teknik klasifikasi dalam data *mining* dikategorikan sebagai *supervised learning* dengan tujuan memprediksi kategori yang tidak diketahui dalam data. Metode ini adalah metode prediksi dalam *data mining* yang seringkali digunakan karena mudah dalam interpretasi model yang dihasilkan (Tan, 2006). Fokus dari penelitian ini adalah pada penerapan teknik klasifikasi.

Dengan mempertimbangkan dinamika dan ketidakpastian pasar yang tinggi pada perusahaan industri otomotif, kajian dilakukan mengenai gangguan pada sistem produksi

perusahaan otomotif, PT XYZ.

Dengan menggunakan teknik diketahui bahwa gangguan dominan dalam proses produksi pada PT XYZ berasal dari proses *casting*. Rancangan model pengendalian gangguan untuk perusahaan terdiri dari sub-model tingkat risiko gangguan untuk mengetahui tingkat risiko gangguan rendah, sedang, atau tinggi. Sub-model status aksi yang dilakukan untuk mengetahui apakah gangguan tersebut membutuhkan tindakan segera atau tidak atau hanya pemantauan. Sub-model tindakan pengendalian gangguan untuk mengetahui tindakan pengendalian yang akan digunakan dan sub-model perhitungan *safety stock* untuk mengetahui kebijakan persediaan pengaman berdasarkan gangguan yang terjadi (Marie dkk., 2017).

Terkait pemanfaatan teknik klasifikasi, telah dilakukan juga penelitian yang menghasilkan rancangan sistem aplikasi dengan model keputusan matriks risiko yang memanfaatkan algoritma *decision tree* dan algoritma *naive Bayes* untuk membantu pengambil keputusan untuk mengetahui tingkat risiko kegagalan sistem transformator supaya dapat menentukan solusi lebih lanjut terkait pemeliharaan transformator pada Divisi X PT Y (Marie dkk., 2016).

Teknik *decision tree* termasuk dalam kelompok teknik klasifikasi tunggal. Kelemahan dari teknik klasifikasi tunggal adalah sangat rentan pada data yang tidak seimbang, dikarenakan teknik klasifikasi tunggal lebih memfokuskan pada kelas mayoritas dan cenderung akan mengabaikan informasi yang ada pada kelas minoritas. Berikutnya, para peneliti mencoba untuk menemukan teknik baru, salah satunya adalah teknik *ensemble* yang dipercaya mampu menangani berbagai macam tipe data. Teknik *ensemble* yang banyak digunakan adalah metode *bagging* dan metode *boosting*.

Metode *bagging* diperkenalkan oleh Breiman pada tahun 1996, memiliki dua tahapan yaitu *bootstrap* untuk pengambilan data dan *aggregating* untuk menggabungkan banyak nilai dugaan menjadi satu dugaan (Sartono & Syafitri, 2010). Metode *bagging* merupakan metode *ensemble* kegunaannya masih banyak diminati peneliti, khususnya kasus ketidakseimbangan

data.

Algoritma *decision tree* adalah lemah terhadap data dengan tingkat ketidakseimbangan yang tinggi, sementara pemanfaatan metode berbasis *bagging* memiliki nilai akurasi dan nilai sentifititas relatif sama. Hal itu menunjukkan metode berbasis *bagging* kuat terhadap data dengan ketidakseimbangan yang tinggi. Selain itu, algoritma *decision tree* cenderung membutuhkan proses komputasi yang lebih cepat jika dibandingkan dengan metode berbasis *bagging* (Hakim dkk., 2017).

Kemampuan metode *bagging* adalah untuk meningkatkan akurasi klasifikasi dengan baik serta dapat menangani kasus dengan tingkat kompleksitas yang tinggi dengan melengkapi penelitian sebelumnya (Fernanda & Otok, 2012). Disimpulkan bahwa penerapan *bagging* dan *boosting* mampu meningkatkan hasil klasifikasi pada berbagai iterasi.

Perkembangan metode *boosting* sudah mengalami banyak modifikasi yang membuktikan bahwa menggunakan metode *boosting* pada berbagai ilmu makin banyak diminati. Selain metode *bagging* dan *boosting*, akhir-akhir ini muncul teknik yang cukup populer yaitu *random forest*.

Random forest merupakan salah satu teknik klasifikasi yang masih terbilang baru, jika dibandingkan dengan metode *bagging* dan metode *boosting*. Metode *random forest* bertujuan untuk memperbaiki pendugaan dari metode *bagging* (Breimen, 2001). Perbedaan metode *random forest* dan metode *bagging* pada dasarnya tidak terlalu jauh, hanya berbeda pada tahapan *random sub-setting*-nya. Perbedaan *sub-setting* pada metode *random forest* tersebut menghasilkan pohon yang berbeda-beda, sedang pada metode *bagging* menghasilkan pohon yang relatif sama. Tujuan dari Breimen yaitu ingin memperkecil korelasi antar pohon (Hastie dkk., 2008).

Perbedaan metode *boosting* dengan metode *bagging* dan metode *random forest* yang terdiri dari banyak pohon dari masing-masing gugus data adalah pada metode *bagging* dan metode *random forest*, untuk mendapatkan hasil akhir akan dilakukan prediksi gabungan, sedangkan

pada metode *boosting* cukup menggunakan satu pohon akan tetapi setiap ada kesalahan dari masing-masing pohon tersebut akan diberikan bobot untuk mendapatkan hasil yang lebih optimal sebanyak n kali.

Terkait penggunaan metode *random forest* juga telah dilakukan beberapa penelitian. Penerakan metode *random forest* dan metode *decision tree* pada 20 data set mendapatkan hasil, bahwa secara keseluruhan metode *random forest* lebih baik jika dibandingkan dengan metode *decision tree* pada data dengan jumlah observasi yang banyak, sedangkan pada jumlah observasi yang sedikit. Kemampuan metode *random forest* tidak jauh berbeda dengan metode *decision tree* (Ali dkk., 2012),

Selanjutnya, untuk menangani permasalahan pada teknik klasifikasi tunggal, telah dikembangkan teknik *resampling* meliputi teknik *over sampling* (menambahkan/mereplikasi contoh pada kelas minoritas) dan teknik *under sampling* (mengeliminasi contoh pada kelas mayoritas) atau gabungan dari kedua metode tersebut, untuk mengurangi rentan pada data yang tidak seimbang. Kedua teknik tersebut juga memiliki kekurangan. Teknik *oversampling* dapat mengakibatkan terjadinya *overfitting*, sedangkan teknik *undersampling* mengakibatkan hilangnya beberapa informasi dari data (Elrahman & Abraham, 2013).

Selanjutnya, penerapan metode *ensemble bagging* dan metode *boost* serta algoritma pengklasifikasi *naive Bayes* untuk menangani kelas tidak seimbang mendapatkan hasil bahwa teknik *ensemble* dengan algoritma *bagging* mampu meningkatkan sensitivitas secara signifikan (Saifuddin & Wahono, 2015). Namun berdasarkan penerapan metode *random forest* untuk menganalisis data terkait perbandingan beberapa metode didapatkan hasil bahwa secara keseluruhan metode *random forest* mampu mengungguli ketiga metode lainnya khususnya pada jumlah *training* data yang sangat banyak (Lowe & Kulkari, 2015).

Penerapan metode *bagging* dan *boosting* pada beberapa algoritma klasifikasi diantaranya: *decision tree* dan *random forest* mendapatkan hasil bahwa pada jumlah sampel yang relatif

besar, *decision tree* lebih baik jika dibandingkan dengan *random forest*. Metode *random forest* bertindak sebagai rata-rata hasil dari kedua *data sets*, dimana pada tahapan penggabungan kedua dengan metode *bagging* dan metode *boosting*, metode *random forest* lebih unggul dibandingkan dengan metode *decision tree* pada jumlah data yang sedikit. Untuk jumlah data yang banyak, *decision tree* dengan gabungan *bagging* lebih unggul jika dibandingkan dengan metode lainnya, yang menandakan bahwa metode *bagging* dan *boosting* mampu meningkatkan hasil klasifikasi (Singh & Kumar, 2016).

Penerapan metode *bagging* dalam bidang ilmu *bioinformatics* yang dilakukan membahas tentang metode *bagging* yang digabungkan *fuzzy rule-based classification* untuk menangani data *multiclass*, baik itu dari segi peningkatan akurasi maupun penanganan data yang berdimensi tinggi. Dari penelitian tersebut didapatkan kesimpulan bahwa secara keseluruhan metode *bagging* mampu meningkatkan akurasi klasifikasi dengan baik serta dapat menangani kasus dengan tingkat kompleksitas yang tinggi (Cordon dkk., 2008).

Hasil evaluasi metode *boosting* dan *bagging* mendapatkan hasil bahwa secara umum kekurangan dari metode *boosting* yaitu pada data dengan kompleksitas yang tinggi, metode *boosting* akan cenderung terjadi *overfitting* (Ganatra & Kosta, 2010).

PT QQQ adalah perusahaan penghasil komponen otomotif yaitu *battery* pada kendaraan bermotor. Perusahaan memproduksi dua jenis varian utama aki, yaitu: aki mobil (AMB) dan aki motor (MCB). Aki mobil terbagi menjadi tiga tipe, yaitu tipe MFDS (*maintenance free*), tipe premium dan tipe *hybrid*. Sedangkan untuk jenis aki motor, memiliki dua jenis varian utama yaitu: tipe premium dan tipe VRLA (*valve regulated lead acid*) atau lebih akrab disebut aki kering. *Breakdown* mesin menjadi bagian dari risiko manufaktur yang mempengaruhi kinerja PT. QQQ. Untuk membantu perusahaan meminimasi risiko manufaktur yang terjadi, dilakukan pemanfaatan teknik klasifikasi untuk data *breakdown* mesin pada PT. QQQ, sehingga dapat membantu perusahaan mengeksplorasi data *breakdown*

mesin dan berikutnya dapat dilakukan analisis model yang diperoleh untuk memprediksikan kelas atau objek yang memiliki label kelas tidak diketahui.

Pada artikel ini, beberapa teknik klasifikasi akan dicobakan untuk melihat teknik yang paling efektif dalam melakukan proses klasifikasi yang juga akan dilengkapi pembahasannya dengan beberapa kajian literatur untuk membuktikan kekurangan dari masing-masing teknik.

Teknik-teknik klasifikasi yang akan digunakan adalah metode *decision tree*, metode *bagging*, metode *boosting*, dan metode *random forest*, yang dibandingkan efektivitasnya berdasarkan kasus pada PT QQQ, perusahaan penghasil komponen otomotif.

II. METODE PENELITIAN

Penelitian ini bertujuan untuk mengetahui faktor-faktor pendorong berinvestasi dari para investor di galeri investasi. Eksplorasi terhadap faktor-faktor ini dilakukan dengan menggunakan kombinasi dari data sekunder dan data primer. Data sekunder diperoleh dari penelitian-penelitian terdahulu yang relevan. Sementara data primer diperoleh dari wawancara dengan pihak investor galeri investasi dan pihak BEI yang kemudian diolah dengan menggunakan metode penelitian kualitatif. Hal ini dilakukan untuk menjamin bahwa faktor-faktor yang dihasilkan memang sudah benar-benar mencakup hal-hal krusial untuk peningkatan minat berinvestasi di galeri investasi.

Terkait kebutuhan pengelolaan ketidakpastian pada rantai pasok industri otomotif, dengan menggunakan data kasus pada PT QQQ yang menghasilkan *battery* mobil, pada penelitian ini digunakan teknik klasifikasi untuk mengelola ketidakpastian dengan memanfaatkan kasus *breakdown* mesin. Berdasar pemanfaatan teknik klasifikasi ingin diketahui hasil analisis perbandingan performansi dari beberapa teknik klasifikasi yakni *decision tree*, *bagging*, *boosting*, dan *random forest*. Tahapan penelitian dimulai dengan kegiatan pengumpulan data yang dilanjutkan dengan pengelompokan untuk data atribut yang dianalisis. Berikutnya, data hasil

Tabel 1. Contoh data kerusakan mesin pada bulan Januari 2018

Tanggal	Shift	What?	Tingkat Risiko	Durasi (menit)	Jenis Kerusakan
3-Jan	1	Kabel auto conveyor m/c 49-50	Rendah	30	Mekanik
4-Jan	1	Belt conveyor kepanjangan	Rendah	30	Mekanik
4-Jan	2	Belt panjang mc 36 putus	Rendah	30	Mekanik
5-Jan	1	M/c 50 belt pendek sobek	Rendah	20	Mekanik
8-Jan	2	Heater leadel tidak ada ampere	Rendah	20	Mekanik
8-Jan	3	Heater POT trip	Rendah	30	Mekanik
9-Jan	2	M/c 29 belt panjang putus	Rendah	15	Mekanik
10-Jan	1	Peluncur dan pendorong mold m/c 53 rusak	Sedang	60	Mekanik
11-Jan	1	Kawat lifter putus	Rendah	30	Mekanik
11-Jan	2	Roll bekt pendek lepas	Rendah	30	Mekanik
11-Jan	3	M/c 15 temp top mould ampere error	Rendah	15	Mekanik
12-Jan	3	As auto conveyor patah	Tinggi	600	Mekanik

pengumpulan data *breakdown* mesin dibagi menjadi dua bagian yaitu data *testing* dan data *training*. Data *training* yang digunakan untuk membuat model sebanyak 75%, sedangkan data *testing* digunakan untuk melihat kebaikan model sebanyak 25%. Metode ini memudahkan para peneliti dalam mengambil keputusan baik itu dalam proses prediksi maupun melihat pola yang terjadi. *Software* yang digunakan pada analisis ini yaitu *software* R dengan bantuan beberapa *package*, seperti *package rpart* untuk teknik klasifikasi *decision tree*, *package adabag* untuk teknik klasifikasi *bagging* dan *boosting*, dan terakhir *package randomforest* untuk teknik klasifikasi *random forest*. Pengujian kebaikan model pada penelitian ini menggunakan bantuan *package* untuk melihat nilai akurasi, sensitifitas dan spesifisitas atau *confusion matrix*.

III. HASIL DAN PEMBAHASAN

PT QQQ memiliki KPI untuk Divisi *Maintenance* sebagai berikut:

1. *Mean time to repair* (MTTR) dapat diturunkan dari 0,53 jam menjadi 0,34 jam.
2. *Breakdown time* dapat diturunkan dari 148 jam menjadi 108 jam.

Untuk mencapai KPI Divisi *Maintenance*, perusahaan melakukan pencatatan kerusakan mesin yang terjadi. Data pada Tabel 1 adalah data contoh kerusakan mesin yang terjadi pada bulan Januari 2018.

Pada penelitian ini dilakukan analisis deskriptif untuk melihat sebaran data apakah data yang

akan dianalisis memiliki kelas yang seimbang atau tidak. Gambar 1 dan Gambar 2 merupakan gambaran mengenai analisis deskriptif dari data *breakdown* yang terjadi pada Divisi *Maintenance*.

Jika diperhatikan grafik pada Gambar 1, diketahui bahwa lama perbaikan kerusakan mesin dapat diselesaikan dengan waktu relatif singkat, diikuti dengan lama perbaikan sedang dan yang terakhir yang paling sedikit adalah lama perbaikan kerusakan mesin yang lama.

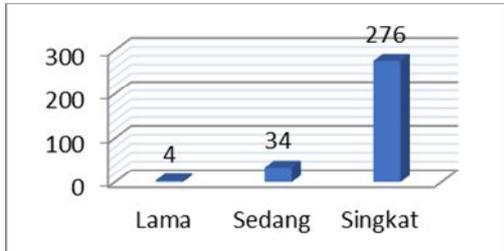
Gambar 2 menjelaskan distribusi waktu terjadinya kerusakan. Perusahaan menerapkan sebanyak 3 *shift* kerja. Kerusakan mesin yang terjadi tertinggi berlangsung pada *shift* 1 dengan jumlah kerusakan sebanyak 138, kemudian *shift* 3 dengan jumlah 92, dan yang terakhir yaitu *shift* 2 dengan jumlah kerusakan 84.

Selain itu, berdasarkan hubungan antara kerusakan mesin yang terjadi dengan *shift* terjadinya kerusakan berikut ini adalah informasi mengenai distribusi jenis kerusakan yang terjadi yang dapat dilihat pada Gambar 3.

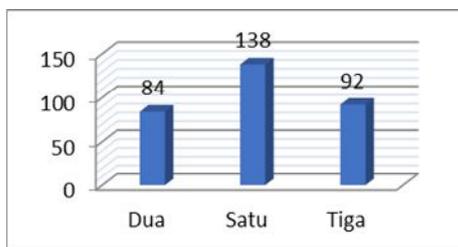
Pada Gambar 3 terlihat bahwa jenis kerusakan tertinggi terdapat pada mekanik sebanyak 171, listrik sebanyak 104, *pneumatic* sebanyak 31, *hydraulic* sebanyak 6, dan motor sebanyak 2. Informasi yang terakhir yaitu mengenai tingkat resiko pada data *breakdown*.

Gambar 4 adalah grafik yang menampilkan informasi mengenai tingkat resiko pada data *breakdown*. Pada Gambar 4 terlihat bahwa kerusakan mesin dengan tingkat resiko yang tinggi masih mendominasi yaitu sebesar 226 dan

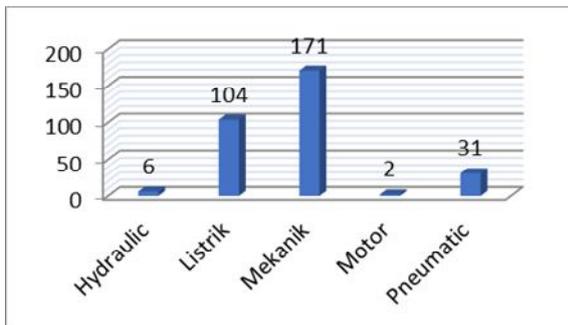
sisanya adalah kerusakan dengan tingkat resiko yang rendah sebesar 88.



Gambar 1. Lama perbaikan kerusakan



Gambar 2. Shift kerja saat terjadinya kerusakan



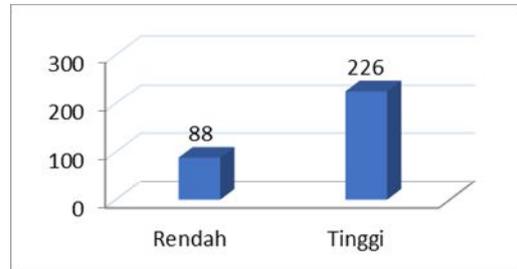
Gambar 3. Jenis kerusakan

Berikutnya, dilakukan pengolahan data kerusakan mesin dengan menggunakan *Software R* dan menghasilkan pohon keputusan sesuai dengan Gambar 5.

Pohon klasifikasi pada Gambar 5 memberikan beberapa informasi yang menarik, diantaranya:

1. Kerusakan dengan waktu perbaikan lama dan sedang biasanya tingkat resiko yang dihasilkan tinggi yaitu sebesar 100%.
2. Kerusakan dengan waktu perbaikan singkat dan jenis kerusakan *Hydraulic* memiliki tingkat resiko tinggi dengan probabilitas 100%.
3. Kerusakan dengan waktu perbaikan singkat, jenis kerusakan listrik, mekanik, *pneumatic* memiliki tingkat resiko tinggi dengan probabilitas sebesar 67,96%.

4. Kerusakan dengan waktu perbaikan singkat, jenis kerusakan listrik, mekanik dan *pneumatic*,



Gambar 4. Tingkat resiko kerusakan



Gambar 5. Pohon klasifikasi/decision tree

bila terjadi pada *shift* satu maka tingkat resiko yang dihasilkan tinggi dengan probabilitas 73%.

5. Kerusakan dengan waktu perbaikan singkat, jenis kerusakan listrik, mekanik dan *pneumatic* bila terjadi pada *shift* dua dan tiga maka tingkat resikonya adalah tinggi dengan probabilitas sebesar 63.63%.
6. Kerusakan dengan waktu perbaikan singkat, jenis kerusakan *pneumatic* yang terjadi pada *shift* dua dan tiga akan memiliki tingkat resiko yang tinggi dengan probabilitas sebesar 59%.
7. Kerusakan dengan waktu perbaikan singkat dan jenis kerusakan mekanik yang terjadi pada *shift* tiga memiliki tingkat resiko tinggi dengan probabilitas sebesar 69.23%.
8. Kerusakan dengan waktu perbaikan singkat dan jenis kerusakan mekanik bila terjadi pada *shift* dua dan tiga maka tingkat resiko rendah dengan probabilitas sebesar 75%.

Untuk dapat memenuhi KPI Divisi *Maintenance*, dengan mempertimbangkan aturan yang dihasilkan dari Pohon Keputusan,

Tabel 2. Perbandingan metode

Metode	Akurasi	Sensitifitas	Spesifisitas
<i>Decision Tree</i>	93.67	80	96.88
<i>Bagging</i>	96.2	97.1	90
<i>Boosting</i>	93.67	75	98.41
<i>Random Forest</i>	96.2	88.89	97.14

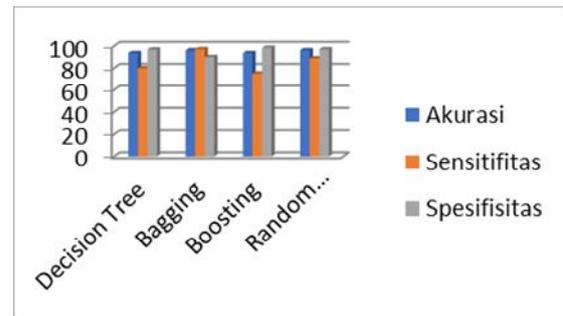
perusahaan dapat memanfaatkan informasi yang dihasilkan dengan mencari alternatif-alternatif strategi kebijakan perawatan mesin.

Tahapan penelitian selanjutnya menghasilkan data perbandingan kinerja untuk beberapa metode teknik klasifikasi yang digunakan, yaitu metode *decision tree*, metode *bagging*, metode *boosting* dan metode *random forest*. Hasilnya dapat dilihat pada Tabel 2.

Metode yang memiliki akurasi tertinggi adalah metode *random forest* dan metode *bagging*. Kedua metode tersebut memiliki nilai akurasi yang sama. akan tetapi nilai sensitivitas dan spesifisitasnya menunjukkan hasil yang berbeda.

Selanjutnya, kemampuan metode *bagging* menebak kelas yang resiko tinggi dengan benar sebesar 97,1%, sementara metode *random forest* sebesar 88,89%. Metode *random forest* unggul dalam menebak kelas dengan resiko rendah (spesifisitas) sebesar 97,14% berbeda dengan metode *bagging* yang hanya mampu menebak sebesar 90%.

Berikutnya nilai akurasi metode *decision tree* dan metode *boosting* memiliki nilai yang sama yaitu sebesar 93,67%. Perbedaan yang mencolok terletak pada nilai sensitifitas. Nilai sensitifitas terendah diantara semua metode yaitu metode *boosting* sebesar 75%. Hal ini



Gambar 6. Grafik perbandingan metode

menandakan bahwa kemampuan menebak kelas yang jumlahnya sangat kecil masih terbilang rendah. Kemudian yang terakhir nilai spesifisitas kedua metode tersebut tidak berbeda jauh yaitu metode *decision tree* sebesar 96,88% dan metode *boosting* sebesar 98.41%.

Gambar 6 menggambarkan tingkat akurasi, sensitifitas dan spesifisitas pada masing-masing metode. Diagram batang berwarna biru memberikan informasi mengenai nilai akurasi dari masing-masing metode. Secara global terlihat, bahwa keempat metode tersebut memiliki nilai akurasi yang terbilang tinggi yaitu semuanya berada diatas nilai 90%. Akan tetapi jika diperhatikan diagram batang yang berwarna merah atau nilai sensitifitas, metode *bagging* mengungguli metode yang lainnya dan yang paling kecil nilainya adalah metode *boosting*.

Kemudian diagram batang berwarna hijau memberikan informasi mengenai nilai spesifisitasnya atau kemampuan metode tersebut menebak kelas dengan tingkat resiko rendah dengan benar. Hasilnya menunjukkan bahwa metode *decision tree*, metode *boosting* dan

Tabel 3. Perbandingan kekurangan dan kelebihan masing-masing metode

Metode	Kekurangan	Kelebihan
<i>Decision tree</i>	Lemah pada data yang tidak seimbang (Hakim dkk, 2017)	Kuat pada kelas yang banyak (Singh & Kumar, 2016)
	Proses komputasi yang lama (Hakim dkk, 2017)	Proses komputasi yang cepat (Hakim dkk, 2017)
<i>Bagging</i>	Lemah pada data yang terdapat <i>missing value</i> (Cordon dkk, 2008)	Kuat pada data dengan kelas yang tidak seimbang (Hakim dkk, 2017)
		Mampu meningkatkan akurasi pada kelas minoritas (Saifuddin & Wahono, 2015)
<i>Boosting</i>	Lemah pada data dengan kompleksitas yang tinggi (Ganatra & Kosta, 2010)	Mampu meningkatkan hasil klasifikasi (Fernanda & Otok, 2012)
<i>Random Forest</i>	Lemah pada jumlah data yang sedikit (Ali dkk, 2012)	Unggul pada jumlah data yang banyak (Lowe & Kulkari, 2015)

metode *random forest* memiliki kemampuan yang hampir sama, dan nilai yang paling kecil terdapat pada metode *bagging*.

Tabel 3 menunjuk perbandingan kekurangan dan kelebihan pada masing-masing metode didapatkan hasil yang sama dengan penelitian yang dilakukan saat ini. Diketahui bahwa metode *bagging* dan metode *boosting* serta metode *random forest* memiliki komputasi yang lebih lama, jika dibandingkan dengan metode *decision tree*. Selain itu, secara keseluruhan metode yang berbasis *ensemble* seperti metode *bagging*, metode *boosting* dan metode *random forest* secara keseluruhan lebih unggul jika dibandingkan dengan metode *decision tree*, khususnya dalam memprediksi kelas dengan jumlah yang sedikit atau minoritas

Hasil perbandingan metode *decision tree*, metode *bagging*, metode *boosting* dan metode *random forest* menunjukkan kelebihan dan kekurangan untuk masing-masing metode sesuai dengan Tabel 3.

IV. SIMPULAN

Berdasarkan analisis perbandingan dapat ditarik kesimpulan bahwa metode *random forest* dan metode *bagging* secara keseluruhan adalah unggul dibandingkan dengan metode *decision tree* dan metode *boosting* berdasarkan kriteria akurasi. Sedangkan untuk nilai sensitifitas, metode *bagging* dan metode *boosting* menghasilkan kinerja lebih baik dibandingkan metode *random forest* dan metode *decision tree*. Untuk kriteria sensitifitas metode *decision tree* memiliki nilai terendah menandakan bahwa kemampuan metode ini lemah dalam menebak kelas yang jumlahnya sangat sedikit yaitu kelas dengan resiko yang tinggi. Berikutnya, diketahui bahwa nilai spesifisitas keempat metode adalah tinggi > 90% (kemampuan keempat metode menebak jumlah banyak, yaitu resiko rendah adalah sangat baik).

Untuk penelitian yang selanjutnya, agar lebih menguatkan informasi mengenai kekurangan serta kelebihan dari masing-masing metode, dapat digunakan beberapa *data set* dalam melakukan perbandingan metode.

UCAPAN TERIMA KASIH

Makalah ini merupakan hasil penelitian hibah yang dibiayai oleh Direktorat Riset dan Pengabdian Masyarakat Direktorat Jenderal Penguatan Riset dan Pengembangan Kementerian Riset, Teknologi, dan Pendidikan Tinggi, sesuai dengan Kontrak Penelitian Nomor: 011/KM/PNT/2018. Penulis sangat berterima kasih atas semua bantuan data dan dukungan dari pimpinan dan staf PT. GS *Battery* dan Lembaga Penelitian Universitas Trisakti atas terselenggaranya penelitian ini.

DAFTAR PUSTAKA

- Ali, J.; Khan, R.; Ahmad, N.; Magsood, I. (2012). "Random forests and decision trees". *International Journal of Computer Science Issue*, 9 (3), 272-278.
- Breiman, L. (1996). "Bagging predictors". *Machine Learning*, 24, 123-140.
- Breiman, L. (2001). "Random Forests". *Machine Learning*, 45 (1), 5-32.
- Córdon, O.; Quirin, A.; Sanchez, L. (2008). "A first study on bagging fuzzy rule-based classification systems with multicriteria genetic selection of the component classifiers". *International Workshop on Genetic and Evolving Fuzzy Systems*, 11-16.
- Elrahman, S.M.; Abraham, A. (2013). "A review of class imbalance problem". *Journal of Network and Innovative Computing*, 1, 332-340.
- Fernanda, J.W.; Otok, B.W. (2012). "Boosting neural network dan boosting cart pada klasifikasi diabetes militus tipe II". *Jurnal Matematika*, 2 (2).
- Hakim, L.; Sartono, B.; Saefuddin, A. (2017). "Bagging based ensemble classification method on imbalance datasets". *IJCSN - International Journal of Computer Science and Network*, 6 (6).
- Hastie, T.J.; Tibshirani, R.J.; Friedman, J.H. (2008). *The Elements of Statistical Learning: Data-mining, Inference and Prediction*. Second edition. New York: Springer-Verlag
- Ho, W.; Zheng, T.; Yildiz, H.; Talluri, S. (2015). "Supply chain risk management: A literature review". *International Journal of Production Research*, 53 (16), 5031-5069.
- Marie, I.A.; Sari, D.K.; Astuti, P.; Teorema, M. (2017). "Design of disturbances control model at automotive company". *IOP Conf. Series: Materials Science and Engineering*, 277, #012020
- Marie I.A.; Ariwibowo, A.B.; Saraswati; Witonohadi, A. (2016). *Determination of Failure Risk for*

- Transformer System Based on Classification Technique*. Proceeding of 9th International Seminar on Industrial Engineering and Management.
- Lowe, B.; Kulkarni, A. (2015). "Multispectral image analysis using random forest". *International Journal on Soft Computing (IJSC)*. 6 (1), 1-14.
- Petre, P. (2013). "Data mining solutions for the business environment". *Database Systems Journal*, 4 (4), 21-29.
- Saifuddin, A.; Wahono, R.S. (2015). "Penerapan teknik ensemble untuk menangani ketidakseimbangan kelas pada prediksi cacat software". *Journal of Software Engineering*. 1 (1), 28-37.
- Sartono, B.; Syafitri. (2010). "Metode pohon gabungan: solusi pilihan untuk mengatasi kelemahan pohon regresi dan klasifikasi tunggal". *Forum Statistika dan Komputasi*, 15 (1), 1-7.
- Singh, A.; Kumar, A. (2016). "Application of bagging and boosting for all the classification algorithms". *International Journal of Pharmacy and Technology*, 8 (3).
- Sulastri, H.; Gufroni, A.I. (2017). "Penerapan data mining dalam pengelompokan penderita thalassaemia". *Jurnal Nasional Teknologi dan Sistem Informasi*, 3 (2), 299-305.
- Tan, P.-N.; Steinbach, M.; Kumar, V. (2006). *Introduction to Data Mining*. USA: Pearson Education, Inc.