



Penggunaan Teknik *Data Mining* untuk Memprediksi *Financial Distress* pada Pemerintah Daerah Kabupaten/Kota di Indonesia

Randi Hermawan¹⁾ **Khamami Herusantoso**²⁾

¹⁾ Direktorat Transformasi Perbendaharaan, Kementerian Keuangan
Gedung Prijadi Praptosuhardjo IIIB Jl. Wahdin 2 No 3,
Jakarta Pusat 10710 Indonesia
email : randhermawan@gmail.com

²⁾ Pusdiklat Keuangan Umum, Kementerian Keuangan
Jl. Pancoran Timur 2 No 1, Jakarta 12780 Indonesia
email : khamami2005@gmail.com

(Diterima 27 November 2015; Diterbitkan 04 Desember 2015)

Abstrak: Penelitian ini bertujuan untuk membuat model prediksi financial distress pada pemerintah daerah kabupaten/kota di Indonesia dengan teknik data mining. Financial distress pada pemerintah daerah merupakan suatu kondisi di mana pemerintah daerah mengalami defisit selama tiga tahun berturut-turut dengan kumulatif defisit mencapai lebih dari 5% dari total pendapatan selama tiga tahun. Teknik data mining yang digunakan adalah teknik klasifikasi Decision Tree. Model prediksi financial distress yang dihasilkan dengan teknik klasifikasi Decision Tree memiliki akurasi keseluruhan sebesar 92,3387% dengan nilai True Positive Rate pada kelas Distress sebesar 0,496 atau 49,6% dan ukuran pohon sebesar 21 dengan 11 daun.

Keywords: Data mining, financial distress, pemerintah daerah, rasio keuangan.

Corresponding author: Khamami Herusantoso, E-mail : khamami2005@gmail.com.

A. Pendahuluan

1. Latar Belakang

Pelaksanaan otonomi daerah di Indonesia yang dilakukan sejak tahun 2001 telah memberikan ruang bagi pemerintah daerah untuk mengatur rumah tangga daerahnya secara mandiri. Hal ini merupakan langkah reformasi dan proses demokratisasi sistem pemerintahan yang ada di Indonesia, yang sebelumnya bersifat sentralistik. Pelaksanaan otonomi daerah ini diatur dalam UU Nomor 22 Tahun 1999 tentang Pemerintahan Daerah, yang kemudian digantikan oleh UU Nomor 32 Tahun 2004.

Namun, program otonomi daerah mulai menunjukkan berbagai permasalahan. Menurut Komite Pemantauan Pelaksanaan Otonomi Daerah (2014), setidaknya terdapat 276

kabupaten/kota yang terancam bangkrut akibat keborosan pemerintah daerah. Keborosan tersebut akan menyebabkan pemerintah daerah kekurangan dana untuk pembangunan fasilitas publik, pelayanan yang berkualitas, percepatan pembangunan, serta pemerataan kesejahteraan rakyat, sesuai dengan amanah kebijakan otonomi daerah (Syurmita, 2014). Apabila kondisi ini terus berlangsung, pemerintah daerah akan mengalami kondisi *financial distress*. *Financial distress* merupakan kondisi di mana pemerintah daerah mengalami kesulitan dalam mendanai operasinya.

Penelitian mengenai prediksi *financial distress* pada pemerintah daerah kabupaten/kota di Indonesia telah dilakukan oleh Sutaryo (2013) dan Syurmita (2014). Hasil penelitian Sutaryo (2013) menunjukkan bahwa rasio keuangan atas laporan keuangan yang disusun dengan dasar cash modified mempunyai kemampuan untuk memprediksi status *financial distress* pemerintah daerah. Sedangkan berdasarkan hasil penelitian Syurmita (2014), prediksi kondisi *financial distress* pemerintah daerah kabupaten/kota di Indonesia dipengaruhi variabel keuangan berupa tingkat kemandirian keuangan daerah, dan variabel non-keuangan berupa derajat desentralisasi, populasi penduduk, dan pemekaran wilayah.

Sementara itu, di negara lain juga terdapat beberapa penelitian mengenai *financial distress* pada pemerintah lokal. Trussel dan Patrick (2009) melakukan penelitian terkait dengan prediksi *financial distress* pada pemerintah lokal di negara bagian Pennsylvania, Amerika Serikat. Hasil penelitian mereka menunjukkan bahwa prediksi *financial distress* pemerintah lokal dipengaruhi oleh konsentrasi pendapatan pemerintah (*revenue concentration*) dan penggunaan utang (*debt usage*).

Dari beberapa penelitian mengenai prediksi *financial distress* pemerintah daerah di atas, belum ada yang menggunakan teknik *data mining* sebagai metode penelitiannya. Penelitian tentang penggunaan teknik *data mining* dalam memprediksi kondisi *financial distress* masih relatif terbatas baik di luar negeri maupun di dalam negeri. Teknik *data mining* lebih banyak digunakan dalam penelitian terkait dengan prediksi status *financial distress* pada perusahaan komersial. Oleh karena itu, pada penelitian ini penulis mencoba untuk menerapkan teknik *data mining* dalam membangun model prediksi *financial distress* pada pemerintah kabupaten/kota di Indonesia.

2. Tujuan penelitian.

Penelitian ini bertujuan untuk membangun model prediksi *financial distress* pada pemerintah daerah kabupaten/kota di Indonesia dengan teknik *data mining*, serta mengetahui akurasi model prediksi yang dibangun. Selain itu, penelitian ini juga bertujuan untuk mengetahui variabel-variabel yang mempengaruhi prediksi *financial distress* pada pemerintah daerah kabupaten/kota di Indonesia.

3. Rumusan masalah.

Berdasarkan latar belakang dan tujuan penelitian di atas, dapat dirumuskan permasalahan sebagai berikut.

- a. Variabel apa sajakah yang mempengaruhi *financial distress* pada pemerintah daerah kabupaten/kota di Indonesia?
- b. Bagaimana model prediksi *financial distress* pada pemerintah daerah kabupaten/kota di Indonesia dengan menggunakan teknik *data mining*?
- c. Bagaimana akurasi model prediksi *financial distress* pada pemerintah daerah kabupaten/kota di Indonesia dengan menggunakan teknik *data mining*?

B. Landasan Teori

1. *Financial distress*.

Financial distress merupakan istilah yang umumnya digunakan dalam keuangan perusahaan untuk menggambarkan kondisi kesulitan keuangan perusahaan. Kesulitan keuangan ini berupa ketidakmampuan perusahaan dalam membayar kewajibannya, kegagalan perusahaan dalam membayar dividen, kekurangan modal kerja, dan ketidakcukupan dana untuk membayar sejumlah biaya perusahaan. Kondisi financial distress ini merupakan gejala awal kondisi kebangkrutan perusahaan. Namun, apabila tidak segera ditindaklanjuti, maka kondisi financial distress dapat mengantarkan pada kebangkrutan perusahaan.

Geng et al. (2014, 240) menyatakan bahwa financial distress merupakan konsep luas yang terdiri dari beberapa kondisi. Kondisi financial distress pada perusahaan umumnya digambarkan dalam beberapa istilah, seperti kebangkrutan, kegagalan, insolvency, dan default. Altman dalam Geng et al. (2014, 240) menyimpulkan bahwa kondisi kebangkrutan adalah istilah yang paling dekat dengan definisi hukum dari financial distress.

Kegagalan perusahaan merupakan kondisi di mana perusahaan tidak dapat membayar kreditur, pemegang saham preferen, supplier, dan lainnya, atau memiliki rekening yang negatif atau mengalami kebangkrutan (Geng et al. 2014). Insolvency merupakan keadaan di mana perusahaan memiliki kekayaan bersih negatif, sedangkan default merupakan kondisi di mana perusahaan melanggar perjanjian dengan kreditur yang mengakibatkan munculnya tindakan hukum (Sutaryo, 2013).

Dalam konteks keuangan pemerintah daerah, Kloha et al. (2005, 314) mendefinisikan financial distress sebagai kegagalan pemerintah dalam memenuhi standar dalam bidang posisi operasi, utang, serta kebutuhan dan sumber daya masyarakat selama beberapa tahun berturut-turut. Menurut Kloha et al. (2005, 314), kondisi financial distress yang terjadi pada pemerintah daerah disebabkan oleh pergeseran pasar tenaga kerja dan populasi, pertumbuhan pemerintah, tuntutan kepentingan kelompok tertentu dalam pemerintahan, serta pengelolaan keuangan daerah yang buruk.

Sementara itu, Trussel dan Patrick (2009, 590) mendefinisikan financial distress sebagai suatu kondisi di mana pemerintah daerah mengalami defisit operasi yang persisten selama tiga tahun berturut-turut. Pemerintah daerah yang mengalami kondisi financial distress harus memenuhi dua kriteria. Pertama, pemerintah daerah tersebut harus mengalami defisit operasi selama tiga tahun berturut-turut. Kedua, jumlah kumulatif defisit operasi selama periode tiga tahun tersebut harus lebih dari 5% dari total pendapatan. Defisit operasi merupakan suatu

kondisi ketidakseimbangan antara realisasi pendapatan pemerintah daerah dengan realisasi pengeluarannya.

2. Data mining.

a. Konsep data mining.

Hand et al. (2001, 9) menyatakan bahwa data mining adalah analisis sekumpulan data amatan (observational data) yang seringkali berjumlah besar untuk menemukan hubungan yang tidak terduga sebelumnya dan untuk meringkas data dalam cara-cara yang mutakhir yang dapat dimengerti dan berguna bagi pemilik data. Sementara itu, Witten et al. (2011, 8) menyatakan bahwa data mining merupakan suatu proses yang bersifat otomatis atau semi otomatis yang dilakukan untuk menemukan pola-pola tertentu dalam data.

Menurut Gorunescu (2011, 14) tujuan utama proses *data mining* terdiri dari dua bagian, yaitu tujuan prediktif dan tujuan deskriptif. Tujuan prediktif dalam data mining dicapai dengan menggunakan bagian dari variabel untuk memprediksi satu atau lebih variabel lainnya. Tujuan deskriptif dalam data mining dicapai dengan mengidentifikasi pola-pola yang menggambarkan data dan yang dapat dimengerti dengan mudah oleh pengguna data.

b. Proses data mining.

1) Eksplorasi data.

Eksplorasi data terdiri dari pembersihan data, transformasi data, normalisasi data, penanganan data yang salah, pengurangan dimensi data, pemilihan subset fitur, dan sebagainya.

2) Membangun model dan validasinya.

Membangun model dan melakukan validasi terhadap model merujuk pada analisis terhadap berbagai macam model dan memilih model yang memiliki performa prediksi terbaik. Metode-metode yang dapat digunakan untuk membangun dan melakukan validasi model adalah metode klasifikasi, regresi, analisis cluster, deteksi anomali, analisis asosiasi, analisis pola sekuensial, dan sebagainya.

3) Menerapkan model

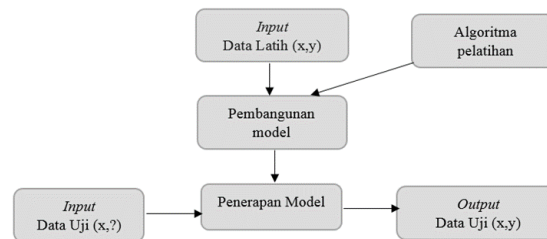
Langkah terakhir dalam proses data mining adalah menerapkan model pada data baru untuk menghasilkan peramalan atau estimasi yang cocok terhadap masalah-masalah yang diinvestigasi.

c. Konsep klasifikasi.

Menurut Fu (1997, 3), klasifikasi dalam data mining adalah turunan dari suatu fungsi atau model yang menggambarkan kelas dari suatu objek berdasarkan atributnya. Atribut adalah field data yang menggambarkan karakteristik atau fitur dalam suatu objek data (Han et al. 2012, 17). Fungsi atau model klasifikasi dibangun dengan menganalisis hubungan antara atribut dan kelas objek dalam set pelatihan (training set). Fungsi atau model klasifikasi seperti ini dapat digunakan untuk mengelompokkan objek-objek di masa depan dan mengembangkan

pemahaman yang lebih baik terhadap kelas-kelas objek dalam basis data (Fu 1997, 3). Gorunescu (2011, 15) menyatakan bahwa komponen dasar dari proses klasifikasi terdiri dari empat komponen, yaitu kelas (class), prediktor (predictors), set data pelatihan (training dataset), dan set data pengujian (testing dataset).

Gambar 1. Proses Klasifikasi



Sumber: Prasetyo, Eko. 2014. *Data Mining: Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Penerbit Andi

d. **Klasifikasi Decision Tree.**

Han (2012, 330) menyatakan bahwa Decision Tree merupakan salah satu teknik klasifikasi yang berupa suatu flowchart yang menyerupai pohon, di mana masing-masing “batang” (internal node) menunjukkan pengujian terhadap atribut, masing-masing “cabang” menggambarkan hasil dari pengujian, serta “daun” menunjukkan kelas data. Menurut Prasetyo (2014, 57), decision tree dapat memberikan keuntungan berupa visualisasi saran yang membuat prosedur prediksinya dapat diamati.

Dalam decision tree dikenal dengan adanya tree pruning. Pruning merupakan proses pemangkasan cabang-cabang pohon pada model decision tree untuk menghindari bias informasi yang dihasilkan, serta menyederhanakan bentuk pohon agar lebih mudah dipahami. Witten et al. (2011, 32) menyatakan bahwa proses pruning terdiri dari dua jenis, yaitu prepruning dan post pruning. Post pruning dilakukan dengan menghilangkan pohon kecil (subtree) yang kurang andal dan menggantinya dengan daun setelah proses klasifikasi, sedangkan prepruning dilakukan dengan menghilangkan pohon kecil berdasarkan nilai minimal objek pada setiap daun yang ditentukan pada saat proses klasifikasi berjalan (Nurchahyo 2014, 13). Drazin dan Montag (2012, 2) menyatakan bahwa prepruning dapat disebut juga dengan istilah on-line pruning. On-line pruning berbeda dengan post pruning di mana proses ini berjalan pada saat decision tree disusun.

e. **Pengukuran kinerja klasifikator.**

1) Akurasi.

Semua algoritma klasifikasi selalu berusaha untuk membangun model yang memiliki akurasi yang tinggi dan laju error yang rendah. Namun, ketika dihadapkan pada data pengujian, suatu model tidak dapat selalu memprediksi dengan benar. Oleh karena itu, perlu dilakukan pengukuran akurasi terhadap suatu model klasifikasi. Akurasi adalah

perbandingan antara jumlah data yang diprediksi secara benar dengan jumlah prediksi yang dilakukan.

2) *Precision dan Recall.*

Dalam bidang pencarian informasi, precision atau positive prediction value merupakan ukuran untuk mengukur kinerja model dalam mendapatkan data yang relevan. Sedangkan recall atau sensitivitas merupakan ukuran kinerja model dalam mendapatkan data relevan yang terbaca. Dalam data mining, precision adalah jumlah data yang true positive (data yang positif dikenali secara benar sebagai positif) dibagi dengan jumlah data yang dikenali secara positif. Sedangkan recall adalah jumlah data yang true positive dibagi dengan jumlah data yang sebenarnya positif (true positive + true negative).

3) *Area Under ROC Curve (AUC)*

Receiver Operating Character (ROC) Curve merupakan suatu kurva dua dimensi yang menggambarkan trade-off antara True Positive (TP) rate dan False Positive (FP) rate, di mana TP rate bertindak sebagai sumbu Y, dan FP rate sebagai sumbu X. Pencatatan dalam ROC dinyatakan dalam sebuah klausa yaitu semakin ke kiri (1,0), maka dinyatakan sebagai klasifikasi prediksi yang mendekati negatif, sedangkan apabila semakin ke kanan (0,1), maka dinyatakan sebagai klasifikasi prediksi yang mendekati positif. Titik (0,1) menggambarkan klasifikasi yang sempurna, karena semua kelas positif diklasifikasikan secara benar sebagai positif, sedangkan titik (1,0) menggambarkan klasifikasi yang gagal, karena semua kelas positif diklasifikasikan sebagai kelas negatif (Bisri dan Wahono 2015, 29). AUC merupakan penilaian kinerja yang dapat diterima untuk kurva ROC. Nilai AUC berkisar antara 0 hingga 1, dengan pembagian klasifikasi sesuai dengan Tabel 1. (Gorunescu 2011, 326).

Tabel 1. Nilai AUC dan Implikasinya.

AUC	Keterangan
0.90 – 1.00	<i>Excellent Classification</i>
0.80 – 0.90	<i>Good Classification</i>
0.70 – 0.80	<i>Fair Classification</i>
0.60 – 0.70	<i>Poor Classification</i>
< 0.60	<i>Failure</i>

Sumber: Gorunescu, Florin. 2011. *Data Mining: Concept, Models, and Techniques*. New York: Springer-Verlag.

C. Penelitian Terkait

Penulis menemukan beberapa penelitian yang membahas mengenai prediksi kondisi financial/fiscal distress pada pemerintah daerah di Indonesia dan local government di

beberapa negara lain, seperti Amerika Serikat dan Australia. Namun, sejauh pengetahuan penulis, belum ada yang menggunakan teknik data mining dalam penelitiannya. Teknik data mining telah digunakan dalam beberapa penelitian terkait prediksi financial distress pada sektor swasta.

Sutaryo (2013) menggunakan nilai relevan informasi keuangan pada laporan keuangan pemerintah daerah dengan cash modified basis berupa rasio-rasio keuangan sebagai variabel penelitian. Hasil penelitian Sutaryo (2013) menunjukkan bahwa rasio keuangan atas laporan keuangan yang disusun dengan dasar cash modified mempunyai kemampuan untuk memprediksi status financial distress pemerintah daerah. Sutaryo menggunakan variabel independen berupa rasio-rasio keuangan entitas swasta yang disesuaikan dengan entitas pemerintahan. Rasio-rasio ini dikelompokkan menjadi profitability ratio, liquidity ratio, capital structure ratio, dan performance ratio. Tingkat keberhasilan model dalam memprediksi kondisi financial distress pada analysis sample mencapai 73,7%, dan pada holdout sample mencapai 68%.

Sedangkan Syurmita (2014) menggunakan variabel keuangan dan non-keuangan dalam memprediksi kondisi financial distress. Berdasarkan hasil penelitian Syurmita (2014), prediksi kondisi financial distress pemerintah daerah kabupaten/kota di Indonesia dipengaruhi variabel keuangan berupa tingkat kemandirian keuangan dan solvabilitas daerah, dan variabel non-keuangan berupa derajat desentralisasi, populasi penduduk, dan pemekaran wilayah. Ketepatan prediksi model yang dihasilkan dalam penelitian ini mencapai 88,7%.

Sementara itu, di negara lain juga terdapat beberapa penelitian mengenai financial distress pada pemerintah lokal setempat. Jones dan Walker (2007) melakukan penelitian untuk menjelaskan variabel-variabel penjelas yang mempengaruhi financial distress pemerintah lokal di negara bagian New South Wales, Australia. Mereka menemukan bahwa tingkat financial distress pada pemerintah lokal dipengaruhi oleh jumlah populasi dan komposisi pendapatan.

Selain itu, Trussel dan Patrick (2009) juga melakukan penelitian untuk mengembangkan model prediksi financial distress pada pemerintah lokal di negara bagian Pennsylvania, Amerika Serikat. Penelitian ini menggunakan rasio-rasio keuangan yang dikelompokkan dalam lima kategori yaitu revenue concentration, administrative expenditures, debt usage, dan entity resources. Model regresi yang dihasilkan dalam penelitian ini memiliki tingkat akurasi prediksi mencapai 91%.

Sedangkan penelitian terkait dengan prediksi financial distress dengan menggunakan teknik data mining lebih banyak dilakukan pada sektor swasta. Berikut ini adalah beberapa penelitian terkait prediksi financial distress yang menggunakan teknik data mining.

Moradi et al. (2013) melakukan penelitian tentang prediksi financial distress pada perusahaan yang terdaftar pada Tehran Stock Exchange dengan menggunakan teknik data mining. Dalam penelitiannya, mereka menggunakan teknik Support Vector Data Description (SVDD) dan Fuzzy C-Means (FCM) Clustering. Hasil penelitian menunjukkan bahwa

dengan SVDD, model yang dibangun memiliki tingkat akurasi mencapai 92% dan dengan FCM, model yang dibangun memiliki tingkat akurasi mencapai 97%.

Penelitian lainnya dilakukan oleh Wei-Sen Chen dan Yin-Kuan Du (2008) dengan sampel penelitian perusahaan yang terdaftar pada Taiwan Stock Exchange Corporation. Teknik yang digunakan adalah Artificial Neural Networks (ANN) dan K-Means Clustering. Hasilnya menunjukkan bahwa dengan ANN mampu membangun model prediksi dengan akurasi mencapai 82,14% dan dengan K-Means Clustering mencapai 73,81%.

Di samping itu, terdapat penelitian lainnya oleh Geng et al. (2014) dengan sampel penelitian perusahaan yang terdaftar pada Shanghai Stock Exchange dan Shenzhen Stock Exchange. Teknik yang digunakan dalam penelitian ini adalah Artificial Neural Network (ANN), Decision Tree (DT), dan Support Vector Machine (SVM). Hasil penelitian menunjukkan bahwa model prediksi dengan menggunakan teknik Artificial Neural Network mampu memiliki kinerja yang lebih baik daripada Decision Tree dan Support Vector Machine.

D. Metode Penelitian

1. Data yang digunakan.

Data yang digunakan dalam penelitian ini merupakan data keuangan seluruh pemerintah daerah kabupaten dan kota di Indonesia. Data ini diperoleh dari situs DJPK, Kementerian Keuangan. Data keuangan yang dikumpulkan dalam penelitian ini berupa data keuangan yang terdapat pada Laporan Realisasi Anggaran (LRA) dan Neraca yang disampaikan oleh Pemerintah Daerah Kabupaten dan Kota pada periode tahun anggaran 2008 sampai dengan 2013.

Data keuangan yang dikumpulkan berbentuk file Excel yang merupakan gabungan dari laporan keuangan seluruh pemerintah daerah di Indonesia, baik pemerintah daerah provinsi maupun pemerintah daerah kabupaten/kota, pada periode tahun anggaran 2008-2013. File Excel yang didapatkan terbagi menjadi dua jenis untuk setiap tahun, yaitu file yang berisi gabungan data keuangan yang berasal dari LRA dan yang berasal dari Neraca. Data LRA dan Neraca yang diperoleh merupakan data yang telah di-update terakhir pada 21 November 2014. Data laporan keuangan Pemerintah Daerah yang berhasil dikumpulkan oleh penulis berjumlah 5706 laporan keuangan, sebagaimana dirinci pada Tabel 2.

Tabel 2. Jumlah Laporan Keuangan Pemerintah Daerah Periode 2008-2013

No.	Tahun Anggaran	Jumlah Laporan	
		LRA	Neraca
1	2008	484	439
2	2009	509	431
3	2010	524	462
4	2011	523	505
5	2012	524	491
6	2013	501	313
Jumlah		3065	2641

Sumber: Direktorat Jenderal Perimbangan Keuangan (diolah)

Data laporan keuangan yang digunakan dalam penelitian ini hanya laporan keuangan pemerintah daerah kabupaten/kota saja, sedangkan data laporan keuangan pemerintah daerah provinsi tidak digunakan. Sehingga dari 5706 data laporan keuangan yang telah dikumpulkan dipilih 5336 laporan keuangan. Dari laporan-laporan keuangan yang telah terkumpul dipilih data keuangan yang akan digunakan untuk membentuk variabel-variabel prediktor dalam penelitian ini. Data keuangan yang digunakan berupa total realisasi pendapatan, realisasi Pendapatan Asli Daerah (PAD), realisasi pajak daerah, realisasi pendapatan transfer, total realisasi belanja, realisasi belanja administrasi, realisasi belanja modal, total aset, total aset lancar, total kewajiban, total kewajiban lancar, dan total ekuitas dana.

Dalam membangun model prediksi *financial Distress* ini, penulis menggunakan data keuangan dua tahun sebelum pemerintah daerah mengalami *financial Distress*. Artinya, apabila suatu pemerintah daerah kabupaten/kota mengalami *financial Distress* pada tahun ke- t , maka variabel atribut yang digunakan untuk memprediksi kondisi tersebut dihitung berdasarkan data keuangan pada tahun ke- $(t-2)$. Sehingga, setelah membangun atribut-atribut baru, langkah selanjutnya adalah membuat *dataset* sementara dengan atribut kelas berupa kondisi *financial Distress* tahun ke- t dan atribut rasio keuangan tahun ke- $(t-2)$. *Dataset* tersebut berjumlah empat *dataset*, dengan masing-masing *dataset* terdiri dari atribut-atribut rasio keuangan tahun 2008 hingga 2011, dan atribut kelas yang berisi kondisi *financial Distress* tahun 2010 hingga 2013.

2. Variabel yang digunakan

Variabel yang digunakan dalam penelitian ini terdiri dari variabel target dan variabel prediktor. Variabel target berupa variabel biner yang terdiri dari dua kelas, yaitu kelas *Distress* dan *Non-distress*. Variabel prediktor berupa rasio-rasio keuangan yang digunakan dalam penelitian Trussel dan Patrick (2009) serta Syurmita (2014). Rasio-rasio keuangan yang digunakan adalah *Tax to Total Revenue* (TAXREV), *Intergovernmental Revenue to Total Revenue* (IGR), *Administrative Expenditures to Total Expenditures* (ADMIN), *Debt Level* (DEBTLEVEL), *Debt to Total Revenue* (DEBTREV), *Entity Size* (SIZE), *Revenue Growth* (GROWTH), *Degree of Decentralization* (DoD), *Financial Independence* (FI), *Capital Outlay* (CAPREV), *Solvability* (SOLVABILITY), *Liquidity* (LIQUIDITY), dan *Leverage* (LEVERAGE). Penjelasan masing-masing variabel/atribut yang digunakan dapat dilihat pada Lampiran 1.

3. Langkah penelitian.

Langkah-langkah penelitian ini mengadopsi proses data mining dengan pendekatan CRISP-DM (Cross-Industry Standard Process for Data mining), yaitu meliputi pemahaman data (data understanding), persiapan data (data preparation), pemodelan (modelling), evaluasi, dan penerapan (Shearer, 2000), sebagaimana yang telah dilakukan oleh Geng, et al. (2014).

a. Pemahaman data (data understanding).

Tahapan ini dilakukan dengan memahami jenis-jenis data yang akan digunakan dalam penelitian. Status kondisi keuangan pemerintah daerah kabupaten/kota dibagi menjadi dua kelas yaitu kelas financial distress dan non-financial distress. Kriteria pemberian label kelas ini didasarkan pada penelitian Trussel dan Patrick (2009), di mana pemerintah yang mengalami defisit operasi kumulatif sebesar lebih dari 5% dari total pendapatan selama tiga tahun berturut-turut diberi label kelas financial distress (FD) dan sisanya diberi label kelas non-financial distress (non-FD). Kedua label kelas ini merupakan variabel target dari model prediksi dan merupakan variabel biner.

b. Persiapan data (data preparation).

Persiapan data meliputi pemilihan data, pembersihan data, konstruk data, dan integrasi data. Dari data yang tersisa setelah dibersihkan, kemudian dihitung rasio-rasio yang dibutuhkan untuk dijadikan sebagai variabel input dalam model. Kemudian, data rasio keuangan hasil perhitungan tersebut diintegrasikan sehingga membentuk record atau tabel yang baru.

Hasil akhir dari persiapan data ini adalah *dataset* akhir yang akan dijadikan sebagai data *input* dalam model yang diberi nama *Dataset_final.csv*. *Dataset* ini terdiri dari 13 atribut rasio keuangan dan satu atribut kelas, dengan jumlah *record* sebanyak 1380 *record*. Dari 1380 *record* ini, kelas *Distress* berjumlah 27 *record* dan kelas *Non-Distress* berjumlah 1353

record. Selain itu, *dataset* yang digunakan untuk penerapan model juga disiapkan. *Dataset* ini terdiri dari data keuangan pada tahun 2012, yang digunakan untuk memprediksi kondisi pada akhir tahun anggaran 2014. Hal ini dilakukan karena realisasi tahun 2014 belum diketahui secara keseluruhan, sehingga label kelas belum dapat diketahui. *File dataset* ini diberi nama *dataset_test.csv*.

Selain itu, apabila dilihat dari proporsi jumlah data pada masing-masing kelas, di mana jumlah data pada kelas *Distress* jauh lebih sedikit daripada kelas *Non-Distress*, jenis *dataset* yang digunakan dalam penelitian ini merupakan jenis *dataset* yang tidakimbang (*imbalanced datasets*). Karena algoritma klasifikasi dalam *data mining* selalu berusaha untuk meminimalkan tingkat kesalahan dalam memprediksi label kelas, penerapannya pada *dataset training* yang tidakimbang akan menghasilkan informasi akurasi yang bias, di mana prediksi terhadap kelas mayoritas memiliki akurasi yang lebih tinggi daripada kelas minoritas (Nisa *et al.*, 2013). Sehingga dibutuhkan perlakuan khusus terhadap *dataset* yang tidakimbang agar dapat menghasilkan model prediksi yang baik. Salah satu teknik yang banyak digunakan dalam mengatasi permasalahan *dataset* yang tidakimbang ini, adalah teknik *resampling* data, baik berupa *oversampling* data minoritas, *undersampling* data mayoritas, maupun kombinasi keduanya.

Pada penelitian ini penulis menggunakan teknik *oversampling* data yang diperkenalkan oleh Chawla, *et al.* (2002), yaitu *Synthetic Minority Over-sampling Technique* (SMOTE). Teknik ini dilakukan untuk mereplikasi data pada kelas minoritas dengan membentuk data sintetis. Teknik ini bekerja dengan cara mencari *k nearest neighbor* (ketetanggaan data) pada setiap data kelas minoritas, kemudian membentuk data sintetis sebanyak persentase duplikasi yang diinginkan di antara data kelas minoritas dan *k nearest neighbor* yang dipilih secara acak (Sastrawan, 2010).

Dalam tahap pemodelan, penulis membangun model dengan menggunakan pendekatan *oversampling* terhadap data yang ada pada kelas minoritas dengan SMOTE. Nilai *k* yang digunakan adalah 5, serta persentase *oversampling* yang digunakan adalah 100% hingga 400%. Selain itu, pemodelan dengan *dataset* asli yang tidakimbang juga dilakukan untuk dibandingkan hasilnya dengan *dataset* hasil *oversampling*. Sehingga, terdapat lima jenis *dataset* yang dapat dijadikan sebagai data input dalam pemodelan.

c. Pemodelan (modelling).

Pemodelan dilakukan dengan menggunakan teknik klasifikasi Decision Tree, dengan kombinasi parameter yang disediakan dalam WEKA versi 3.7.12. Rincian parameter yang digunakan dapat dilihat pada Lampiran 2. Pada tahap ini, performa masing-masing model dipaparkan, selanjutnya dipilih model prediksi yang memiliki performa terbaik. Ukuran penilaian performa yang digunakan berupa nilai akurasi, nilai Area Under ROC Curve (AUC), tree size, dan Kappa Statistic.

d. Evaluasi (evaluation).

Setelah model dibangun, langkah selanjutnya adalah mengevaluasi pola yang dihasilkan dari model prediksi yang telah dibentuk dan dipilih pada tahap pemodelan, serta melakukan pembahasan terkait pola yang dihasilkan.

e. Penerapan (deployment).

Pada tahap ini, model prediksi yang telah dihasilkan diterapkan pada dataset baru yang belum memiliki label kelas, untuk memprediksi label kelas dataset yang baru tersebut.

E. Hasil dan Pembahasan

1. Pemodelan dan pemilihan model.

Model yang dihasilkan pada penelitian ini menggunakan lima jenis dataset, yaitu dataset asli yang tak imbang dan dataset yang telah di-*oversampling* sebesar 100% hingga 400%. Masing-masing dataset yang tersedia dilakukan pemodelan dengan tiga proses *pruning decision tree*, yaitu *unpruned*, *on-line pruning*, dan *post-pruning*.

Model *Unpruned* dibuat dengan mengubah parameter *unpruned* pada WEKA. Nilai pada parameter *unpruned* diubah dari FALSE menjadi TRUE. *On-line pruning* dilakukan dengan cara mengubah parameter *MinNumObj* dan *Confidence Factor*. Parameter ini menunjukkan jumlah minimal *instances* pada setiap “daun”. Nilai parameter *MinNumObj* yang digunakan adalah sebesar 20, 30, dan 50, serta nilai *Confidence Factor* ditetapkan sebesar 0,95. Sedangkan *post pruning* dilakukan dengan cara mengubah nilai *Confidence Factor*. Nilai *Confidence Factor* adalah 0,1; 0,5; dan 0,95, serta parameter *MinNumObj* dibiarkan sesuai dengan *default* yaitu 2.

Metode evaluasi yang digunakan untuk menilai kinerja klasifikator ini adalah 10-fold Cross Validation. Metode ini dilakukan dengan memecah dataset menjadi 10 bagian data dengan ukuran yang sama. Pada setiap pemrosesan data berjalan, satu pecahan data berperan sebagai set data uji, sedangkan bagian lainnya berperan sebagai set data latih. Proses ini dilakukan sebanyak 10 kali sehingga setiap data dalam set data berkesempatan untuk menjadi data uji tepat satu kali dan menjadi set data latih sebanyak 9 kali. Jumlah model yang dihasilkan sebanyak 29 model yang diberi nama D1 hingga D29, dengan rincian yang dapat dilihat pada Tabel 3.

Tabel 3. Rincian Masing-masing Model sesuai dengan *Dataset* dan proses *Pruning*.

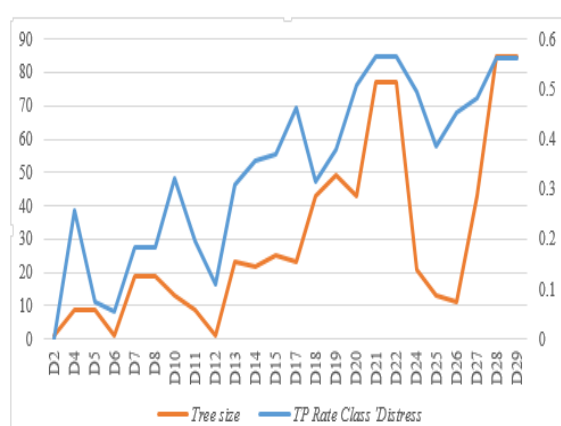
Model	<i>Dataset</i>	<i>Pruning</i>
D1	<i>Imbalanced dataset</i>	<i>Unpruned</i>
D2	<i>Imbalanced dataset</i>	<i>On-line Pruning</i>
D3	SMOTE 100%	<i>Unpruned</i>
D4-D5	SMOTE 100%	<i>On-line Pruning</i>
D6-D8	SMOTE 100%	<i>Post Pruning</i>
D9	SMOTE 200%	<i>Unpruned</i>
D10-D12	SMOTE 200%	<i>On-line Pruning</i>
D13-D15	SMOTE 200%	<i>Post Pruning</i>
D16	SMOTE 300%	<i>Unpruned</i>
D17-D19	SMOTE 300%	<i>On-line Pruning</i>
D20-D22	SMOTE 300%	<i>Post Pruning</i>
D23	SMOTE 400%	<i>Unpruned</i>
D24-D26	SMOTE 400%	<i>On-line Pruning</i>
D27-D29	SMOTE 400%	<i>Post Pruning</i>

Setelah pembangunan model dilakukan, langkah selanjutnya adalah melakukan analisis pada seluruh model yang telah dibangun. Namun hanya model yang telah melalui proses pruning yang dianalisis, baik yang secara on-line pruning maupun post pruning. Setelah analisis dilakukan, selanjutnya dipilih model yang memiliki nilai kinerja terbaik.

Pemilihan model pada penelitian ini tidak didasarkan pada nilai akurasi keseluruhan pada masing-masing model, namun didasarkan pada nilai True Positive (TP) Rate pada kelas Distress. Hal ini dikarenakan bahwa rata-rata akurasi pada semua model yang dibentuk tidak berbeda secara signifikan. Rata-rata nilai akurasi semua model secara keseluruhan sangat tinggi, yakni di atas 90%. Namun, nilai TP rate pada kelas Distress berubah-ubah. Selain itu, tujuan penelitian ini adalah membangun model prediksi kondisi financial Distress, sehingga pemilihan model lebih difokuskan pada kemampuan model dalam memprediksi kelas Distress.

Grafik 1 menunjukkan perbandingan antara True Positive Rate pada kelas Distress dengan ukuran pohon pada masing-masing model. True Positive Rate merupakan nilai yang menunjukkan persentase jumlah instances yang dapat diklasifikasikan dengan benar sesuai dengan kelasnya terhadap total instances yang tersedia pada masing-masing kelas. TP rate pada kelas Distress menunjukkan banyaknya instances pada kelas Distress yang dapat diklasifikasikan secara benar ke dalam kelas Distress oleh model.

Grafik 1. Perbandingan antara Ukuran Pohon dengan True Positive Rate pada kelas Distress



Nilai TP rate tertinggi yang dimiliki oleh model adalah 0,565. Model yang memiliki nilai ini adalah model D21 dan D22. Namun ukuran pohon yang dimiliki masih besar yakni 77. Sedangkan, nilai TP rate terendah adalah 0,000 yang dimiliki oleh model D2 yang menggunakan dataset awal yang belum di-oversample pada kelas minoritasnya.

2. Evaluasi.

a. Evaluasi Model.

Model D24 merupakan salah satu model yang dibangun dengan input berupa dataset yang telah di-oversample kelas minoritasnya sebesar 400%, sehingga dataset yang digunakan terdiri dari 1488 instances. Evaluasi yang lebih mendalam pada Model D24 menunjukkan bahwa akurasi yang dimiliki oleh Model D24 adalah sebesar 92,3387%, yang berarti bahwa secara keseluruhan, Model D24 mampu mengklasifikasikan 1374 instances dengan benar. Kappa Statistic yang dimiliki oleh model masih cukup baik, yaitu 0,4989. Selain itu, nilai AUC yang dimiliki menunjukkan nilai yang tinggi, yaitu 0,855. Nilai AUC ini berada pada interval 0,8-0,9. Berdasarkan Tabel II.2, model dengan nilai AUC yang terletak pada interval 0,8-0,9 dikategorikan sebagai good classification.

Confusion matrix menunjukkan jumlah instances yang diklasifikasikan secara tepat dan salah oleh model. Baris pertama pada Confusion Matrix menunjukkan instances yang pada faktanya berada pada kelas Distress. Angka pertama menunjukkan sebanyak 67 instances diklasifikasikan dengan benar ke dalam kelas Distress. Angka kedua menunjukkan bahwa sebanyak 68 instances yang sebenarnya berada pada kelas Distress diklasifikasikan secara salah ke dalam kelas Non-Distress.

Baris kedua menunjukkan instances yang pada faktanya berada pada kelas Non-Distress. Angka pertama menunjukkan sebanyak 46 instances diklasifikasikan secara salah ke dalam kelas Distress. Angka kedua menunjukkan bahwa sebanyak 1307 instances yang sebenarnya berada pada kelas Non-Distress diklasifikasikan secara benar ke dalam kelas Non-Distress.

Confusion Matrix juga menampilkan rincian akurasi pada masing-masing kelas pada model. Nilai rata-rata TP rate sebesar 0,923 menunjukkan bahwa model mampu mengklasifikasikan instances dengan benar pada masing-masing kelas rata-rata sebesar 92,3%. Tingkat akurasi ini didukung oleh nilai rata-rata False Positive Rate sebesar 0,461, Precision sebesar 0,918, Recall sebesar 0,923, dan F-measure sebesar 0,920.

b. Evaluasi Pola

Model D24 menghasilkan decision tree dengan ukuran pohon sebesar 21 dan jumlah daun sebanyak 11. Apabila dibandingkan dengan pola yang dihasilkan oleh model-model lainnya yang menghasilkan pohon-pohon yang memiliki ukuran yang lebih besar, Model D24 menghilangkan atribut-atribut seperti rasio antara belanja modal dengan total pendapatan (CAPREV), total utang (DEBTLEVEL), rasio antara total utang dengan total pendapatan (DEBTREV), solvabilitas (SOLVABILITY), derajat desentralisasi (DoD), dan kemandirian keuangan (FI). Hal ini menandakan bahwa keenam atribut tersebut tidak mempengaruhi prediksi kondisi financial Distress pada dua tahun yang akan datang.

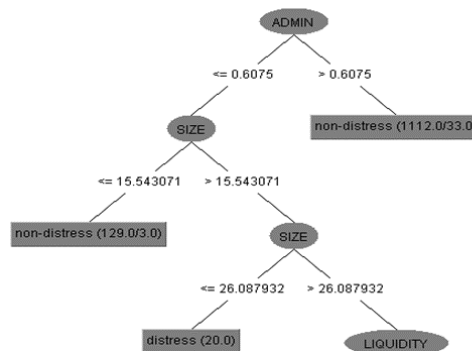
Atribut pertama yang menjadi akar pohon (root node) dan menjadi dasar klasifikasi pada decision tree adalah atribut ADMIN. Atribut ini merupakan rasio antara realisasi belanja administratif dengan total belanja keseluruhan. Belanja administratif dihitung dengan menjumlahkan realisasi belanja pegawai dengan belanja barang dan jasa.

Dalam pohon keputusan, atribut ADMIN dibagi menjadi dua cabang yaitu cabang dengan nilai $ADMIN > 0,6075$ dan cabang dengan nilai $ADMIN \leq 0,6075$. Pada cabang yang berisi pemerintah daerah (pemda) kabupaten/kota dengan nilai $ADMIN > 0,6075$, model langsung mengklasifikasikan pemda tersebut ke dalam kelas Non-Distress. Hal ini berarti bahwa apabila suatu pemda kabupaten/kota memiliki nilai realisasi belanja administratif sebesar lebih dari 60,75% dari total realisasi belanja pada tahun berjalan, maka pemda tersebut dapat diprediksi tidak akan mengalami kondisi financial Distress. Kondisi ini sesuai dengan hasil penelitian Trussel dan Patrick (2009), di mana pada penelitiannya, mereka menemukan bahwa kenaikan belanja administratif mampu mengurangi risiko pemerintah daerah untuk mengalami kondisi financial Distress. Selanjutnya pada cabang kedua, yaitu cabang yang berisi pemerintah daerah (pemda) kabupaten/kota dengan nilai $ADMIN \leq 0,6075$, model akan melihat terlebih dahulu pada atribut SIZE. SIZE merupakan atribut yang menggambarkan

ukuran entitas, yang dalam hal ini adalah ukuran pemerintah daerah kabupaten/kota. Atribut ini dihitung dengan menggunakan logaritma natural dari total realisasi pendapatan. Apabila nilai atribut $SIZE \leq 15,5431$, maka pemda kabupaten/kota akan diprediksi tidak mengalami kondisi financial Distress pada dua tahun mendatang. Apabila nilai atribut $SIZE > 15,5431$ dan $SIZE \leq 26,0879$, maka pemda kabupaten/kota akan diprediksi mengalami kondisi financial Distress pada dua tahun yang akan datang.

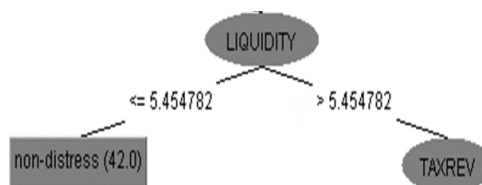
Selanjutnya, apabila nilai atribut $SIZE > 26,8079$, model akan melihat nilai atribut LIQUIDITY. Sampai pada proses ini, dapat disimpulkan bahwa pemerintah daerah kabupaten/kota yang diprediksi mengalami financial Distress adalah pemerintah daerah yang memiliki realisasi belanja administratif kurang dari atau sama dengan 60,75% dari total belanja dan ukuran realisasi pendapatan lebih dari 15,543 sampai dengan 26,088. Potongan pohon sampai pada proses ini ditampilkan pada Gambar 1.

Gambar 1. Potongan 1 *Decision Tree* Model D24



Pada node atribut LIQUIDITY, model akan membentuk dua cabang yaitu cabang dengan nilai $LIQUIDITY > 5,4548$ dan cabang dengan nilai $LIQUIDITY \leq 5,4548$. Atribut LIQUIDITY berisi nilai rasio likuiditas yang dihitung dengan membagi total aset lancar dengan total kewajiban lancar. Apabila nilai atribut LIQUIDITY kurang dari atau sama dengan 5,4548, maka pemerintah daerah kabupaten/kota akan langsung diklasifikasikan pada kelas non-Distress, hal ini berarti pemda tersebut diprediksi tidak mengalami financial Distress dua tahun berikutnya. Apabila nilai LIQUIDITY lebih dari 5,4548, maka model akan melihat nilai atribut TAXREV. Atribut ini merupakan atribut yang berisi nilai rasio antara realisasi pajak daerah dengan total pendapatan. Potongan pohon sampai pada proses ini ditampilkan pada Gambar 2.

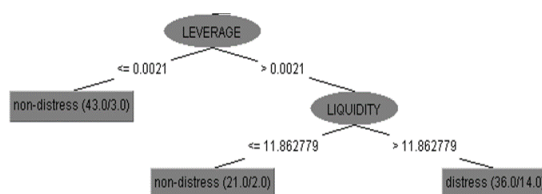
Gambar 2. Potongan 2 *Decision Tree* Model D24



Pada node atribut TAXREV, model membagi node menjadi dua cabang, yaitu cabang dengan nilai TAXREV $\leq 0,0077$ dan cabang dengan nilai TAXREV $> 0,0077$. Pada cabang dengan nilai TAXREV $\leq 0,0077$, model akan melihat pada atribut LEVERAGE. Atribut ini merupakan rasio antara jumlah utang atau kewajiban dengan total ekuitas dana. Apabila nilai atribut LEVERAGE $\leq 0,0021$, maka model akan mengklasifikasikan pemda ke dalam kelas Non-Distress. Hal ini berarti bahwa apabila suatu pemda memiliki nilai rasio utang terhadap ekuitas dana kurang dari 0,0021, maka pemda tersebut diprediksi tidak akan mengalami financial Distress.

Selanjutnya, apabila nilai atribut LEVERAGE $> 0,0021$, model akan melihat kembali pada nilai atribut LIQUIDITY. Apabila nilai atribut LIQUIDITY $\leq 11,8628$, maka model akan mengklasifikasikan pemda ke dalam kelas Non-Distress. Sedangkan, apabila nilai atribut LIQUIDITY $> 11,8628$, maka model akan mengklasifikasikan pemda ke dalam kelas Distress. Potongan pohon sampai pada proses ini ditampilkan pada Gambar 3.

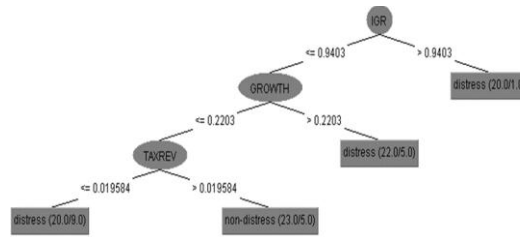
Gambar 3. Potongan 3 *Decision Tree Model D24*



Kemudian, pada cabang dengan nilai TAXREV $> 0,0077$, model akan melihat nilai atribut IGR. Atribut IGR merupakan rasio antara realisasi pendapatan transfer (intergovernmental revenue) dengan total realisasi pendapatan. Apabila nilai IGR lebih dari 0,9403, maka model akan mengklasifikasikan pemda ke dalam kelas Distress. Hal ini berarti bahwa apabila proporsi pendapatan transfer daerah sebesar 94,03%, maka pemda akan berisiko mengalami financial Distress.

Selanjutnya, apabila nilai atribut IGR $\leq 0,2203$, maka model akan melihat pada atribut GROWTH. Atribut ini merupakan rasio antara pertumbuhan pendapatan selama dua tahun berturut-turut dengan total realisasi pendapatan tahun sebelumnya. Apabila nilai pertumbuhan pendapatan lebih dari 0,2203 atau 22,03%, maka model akan mengklasifikasikan pemda ke dalam kelas Distress. Apabila pertumbuhan pendapatan kurang dari atau sama dengan 22,03%, maka model akan melihat kembali pada nilai atribut TAXREV. Apabila nilai atribut TAXREV lebih dari 0,01958, maka model akan mengklasifikasikan pemda ke dalam kelas Non-Distress. Sedangkan, apabila nilai atribut TAXREV kurang dari atau sama dengan 0,01958, maka pemda diprediksi akan mengalami financial Distress. Potongan pohon sampai pada proses ini ditampilkan pada Gambar 4.

Gambar 4. Potongan 4 Decision Tree Model D24



Berdasarkan pola yang dihasilkan pada Model D24, kondisi pemerintah daerah yang diprediksi mengalami kondisi financial Distress, adalah sebagai berikut.

- Rasio realisasi belanja administratif terhadap total realisasi belanja kurang dari atau sama dengan 0,6075 serta ukuran pemda lebih dari 15,543 dan kurang dari atau sama dengan 26,0879.
- Rasio realisasi belanja administratif terhadap total realisasi belanja kurang dari atau sama dengan 0,6075, ukuran pemda lebih dari 26,0879, rasio likuiditas lebih dari 11,8628, rasio pajak daerah terhadap total pendapatan kurang dari 0,0077, dan rasio leverage lebih dari 0,0021.
- Rasio realisasi belanja administratif terhadap total realisasi belanja kurang dari atau sama dengan 0,6075, ukuran pemda lebih dari 26,0879, rasio likuiditas lebih dari 5,4548, rasio pajak daerah terhadap total pendapatan lebih dari 0,0077, dan rasio pendapatan transfer terhadap total belanja lebih dari 0,9403.
- Rasio realisasi belanja administratif terhadap total realisasi belanja kurang dari atau sama dengan 0,6075, ukuran pemda lebih dari 26,0879, rasio likuiditas lebih dari 5,4548, rasio pajak daerah terhadap total pendapatan lebih dari 0,0077, rasio pendapatan transfer terhadap total belanja kurang dari atau sama dengan 0,9403, dan pertumbuhan total pendapatan lebih dari 22,03%.
- Rasio realisasi belanja administratif terhadap total realisasi belanja kurang dari atau sama dengan 0,6075, ukuran pemda lebih dari 26,0879, rasio likuiditas lebih dari 5,4548, rasio pajak daerah terhadap total pendapatan lebih dari 0,0077, rasio pendapatan transfer terhadap total belanja kurang dari atau sama dengan 0,9403, pertumbuhan total pendapatan lebih dari 22,03%, dan rasio pajak daerah terhadap total pendapatan kurang dari 0,01958.

3. Penerapan

Setelah dilakukan pemilihan dan evaluasi terhadap model, langkah selanjutnya adalah menerapkan Model D24 pada dataset yang belum diketahui kelasnya. Ringkasan hasil prediksi yang diperoleh dengan Model D24 disajikan dalam Tabel 4.

Tabel 4. Hasil Prediksi Model D24

<i>Class</i>	<i>Distress</i>	<i>Non-Distress</i>	<i>Total</i>
Jumlah data	1	407	408

Tabel 4 menunjukkan bahwa dari total data test yang berjumlah 408 instances dengan kelas yang belum diketahui, sebanyak 407 diprediksi tidak mengalami financial Distress, dan hanya 1 instance yang diprediksi akan mengalami financial Distress dua tahun berikutnya. Jumlah instance yang diprediksi mengalami financial Distress hanya satu instance karena kondisi financial Distress ini merupakan kondisi yang jarang terjadi (rare case). Hal ini dapat dibuktikan dari jumlah data atau instances yang berhasil dikumpulkan pada dataset training hanya berjumlah 27 data dari 1380 data.

F. Kesimpulan dan Saran

1. Kesimpulan

- a. Model prediksi kondisi financial distress yang dibangun dengan teknik Decision Tree terdiri dari 29 model. Model-model ini dibangun dengan lima dataset yang berbeda. Hal ini dilakukan karena dataset asli merupakan dataset yang tidakimbang, di mana kelas Distress memiliki jumlah instance yang jauh lebih sedikit dari pada jumlah instance pada kelas Non-distress. Oleh karena itu, jumlah instance pada kelas Distress perlu dilakukan oversampling. Teknik oversampling yang digunakan adalah Synthetic Minority Oversampling Technique (SMOTE). Dataset yang digunakan dalam membangun model prediksi terdiri dari dataset asli dan dataset yang telah di-oversample sebanyak 100% hingga 400%. Sehingga, dataset input yang digunakan untuk membentuk model berjumlah 5 jenis dataset.
- b. Dari 29 model yang dibangun dengan Decision Tree, dipilih satu model yang digunakan untuk memprediksi kondisi financial distress pada pemerintah kabupaten/kota di Indonesia, yaitu Model D24. Model ini memiliki ukuran pohon sebesar 21 dengan 11 daun dan nilai akurasi keseluruhan sebesar 92,3387%, serta nilai AUC sebesar 0,855.
- c. Variabel atau atribut yang mempengaruhi proses prediksi financial distress berdasarkan Model D24 adalah rasio belanja administratif terhadap total belanja (ADMIN), ukuran entitas yang dihitung dari logaritma natural total pendapatan (SIZE), rasio likuiditas (LIQUIDITY), rasio pajak daerah terhadap total pendapatan (TAXREV), rasio leverage (LEVERAGE), rasio pendapatan transfer terhadap total pendapatan (IGR), dan pertumbuhan total pendapatan (GROWTH).
- d. Berdasarkan pola yang dihasilkan pada Model D24, kondisi pemerintah daerah yang diprediksi mengalami kondisi financial distress, adalah sebagai berikut:
 - 1) Rasio realisasi belanja administratif terhadap total realisasi belanja kurang dari atau sama dengan 0,6075 serta ukuran pemda lebih dari 15,543 dan kurang dari atau sama dengan 26,0879.

- 2) Rasio realisasi belanja administratif terhadap total realisasi belanja kurang dari atau sama dengan 0,6075, ukuran pemda lebih dari 26,0879, rasio likuiditas lebih dari 11,8628, rasio pajak daerah terhadap total pendapatan kurang dari 0,0077, dan rasio leverage lebih dari 0,0021.
- 3) Rasio realisasi belanja administratif terhadap total realisasi belanja kurang dari atau sama dengan 0,6075, ukuran pemda lebih dari 26,0879, rasio likuiditas lebih dari 5,4548, rasio pajak daerah terhadap total pendapatan lebih dari 0,0077, dan rasio pendapatan transfer terhadap total belanja lebih dari 0,9403.
- 4) Rasio realisasi belanja administratif terhadap total realisasi belanja kurang dari atau sama dengan 0,6075, ukuran pemda lebih dari 26,0879, rasio likuiditas lebih dari 5,4548, rasio pajak daerah terhadap total pendapatan lebih dari 0,0077, rasio pendapatan transfer terhadap total belanja kurang dari atau sama dengan 0,9403, dan pertumbuhan total pendapatan lebih dari 22,03%.
- 5) Rasio realisasi belanja administratif terhadap total realisasi belanja kurang dari atau sama dengan 0,6075, ukuran pemda lebih dari 26,0879, rasio likuiditas lebih dari 5,4548, rasio pajak daerah terhadap total pendapatan lebih dari 0,0077, rasio pendapatan transfer terhadap total belanja kurang dari atau sama dengan 0,9403, pertumbuhan total pendapatan lebih dari 22,03%, dan rasio pajak daerah terhadap total pendapatan kurang dari 0,01958.

2. Saran

Model D24 dapat diterapkan untuk membantu pemerintah, baik pemerintah pusat maupun pemerintah daerah sebagai tool untuk memprediksi kondisi financial distress pada pemerintah daerah kabupaten/kota di Indonesia. Selanjutnya hasil prediksi tersebut dapat digunakan sebagai dasar untuk melakukan tindakan-tindakan preventif guna mencegah terjadinya kondisi financial distress pada pemerintah daerah kabupaten/kota.

Model ini hanya dibangun dengan menggunakan atribut-atribut keuangan. Oleh karena itu, pada penelitian selanjutnya, atribut-atribut non-keuangan sebaiknya ditambahkan guna mendapatkan model prediksi yang memiliki kinerja yang lebih baik.

G. Daftar Pustaka

- Bisri, Achmad dan Romi Satria Wahono. 2015. Penerapan *Adaboost* untuk Penyelesaian Ketidakseimbangan Kelas pada Penentuan Kelulusan Mahasiswa dengan Metode *Decision Tree*. *Journal of Intelligent Systems* 1, no. 1: 27-32
- Chawla, Nitesh V., Kevin W. Bowyer, Lawrence O. Hall, dan W. Phillip Kegelmeyer. 2002. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research* 16: 321-357.

- Chen, Wei-Sen dan Yin Kuan Du. 2008. Using Neural Networks and Data Mining Techniques for The Financial Distress Prediction Model. *Expert System With Application* 36, no. 2: 4075-4086.
- Chien, Chen-Fu dan Li-Fei Chen. 2008. Data Mining to Improve Personnel Selection and Enhance Human Capital: A Case Study in High-Technology Industry. *Expert System With Application* 34: 280-290.
- Ding, Yongsheng, Xinping Song, dan Yueming Zen. 2008. Forecasting Financial Condition of Chinese Listed Companies Based on Support Vector Machines. *Expert System With Application* 34, no. 4: 3081-3089.
- Drazin, Sam dan Matt Montag. 2012. Decision Tree Analysis Using Weka. *Machine Learning – Project II*. Miami: University of Miami.
- Fu, Yongjian. 1997. Data Mining: Tasks, Techniques and Applications. *IEEE Potentials* 16, no. 4: 18-20.
- Geng, Ruibin, Indranil Bose dan Xi Chen. 2015. Prediction of Financial Distress: An Empirical Study of Listed Chinese Companies Using Data Mining. *European Journal of Operating Research* 241: 236-247.
- Gorunescu, Florin. 2011. *Data Mining: Concept, Models, and Techniques*. New York: Springer-Verlag.
- Han, Jiawei, Micheline Kamber dan Jian Pei. 2012. *Data Mining: Concept and Techniques*. Edisi ke-3. San Fransisco: Morgan Kauffman Publishers.
- Hand, David, Heikki Mannila, dan Padhraic Smyth. 2001. *Principles of Data Mining*. Massachusetts: The MIT Press.
- Jones, Stewart dan Walker, R. G. 2007. Explanators of Local Government Distress. *ABACUS* 43, no. 3: 396-418.
- Kloha, Philip, Carol S. Weissert, dan Robert Kleine. 2005. Developing and Testing a Composite Model to Predict Local Fiscal Distress. *Public Administration Review* 65, no. 3: 313-323.
- Moradi, M., M. Salehi, Mohammad Ebrahim Ghorgani, dan Hadi Sadoghi Yazdi. 2013. Financial Distress Prediction of Iranian Companies Using Data Mining Techniques. *Organizacija* 46: 20-27.
- Nisa, Umi Zhahratun, Budi Santosa, dan Stefanus Eko Wiratno. 2013. *Model Prediksi Financial Distress pada Perusahaan Manufaktur Go Public di Indonesia*. Prosiding Seminar Nasional Manajemen Teknologi XVIII. Surabaya.
- Nurcahyo, Alan. 2015. Deteksi dan Karakterisasi Penunggak Pajak Menggunakan Teknik *Data Mining*. Skripsi. Tangerang Selatan: Sekolah Tinggi Akuntansi Negara.
- Nursetyautami, Annisa Fisakinah. 2013. *Examining the Association between Local Government Financial Indicators and Public Service*. Proceedings of 8th Annual London Business Research Conference. London.

- Prasetyo, Eko. 2014. *Data Mining: Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Penerbit Andi.
- Ritonga, Irwan Taufiq, C. Clark dan G. Wickremasinghe. 2012. Assessing Financial Condition of Local Government in Indonesia: An Exploration. *Public Municipal Finance* 1, no. 2: 37-50.
- Shearer, Colin. 2000. The CRISP-DM Model: The New Blueprint for Data Mining. *Journal of Warehousing* 5, no. 4: 13-22.
- Sutaryo. 2013. Nilai Relevan Informasi Laporan Keuangan *Cash Modified Basis*: Kemampuan Rasio Keuangan dalam Memprediksi Status *Financial Distress* Pemerintah Daerah di Indonesia. <http://ssrn.com/abstract=2339632> (diakses pada 20 Agustus 2015).
- Syurmita. 2014. *Prediksi Financial Distress Pemerintah Daerah Kabupaten/Kota di Indonesia*. Simposium Nasional Akuntansi XVII. Mataram.
- Witten, Ian H., Eibe Frank dan Mark A. Hall. 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. Edisi ke-3. San Fransisco: Morgan Kauffman Publishers.
- Trussel, John M., dan Patricia A. Patrick. 2009. A Predictive Model of Fiscal Distress in Local Government. *Journal of Public Budgeting, Accounting and Financial Management* 21, no. 4: 578-616.

Lampiran 1. Penjelasan Atribut Rasio Keuangan yang Digunakan

No.	Nama Atribut	Pengukuran
1	<i>Tax to Total Revenue (TAXREV)</i>	$\frac{\text{Pendapatan Pajak Daerah}}{\text{Total Pendapatan Daerah}}$
2	<i>Intergovernmental Revenue to Total Revenue (IGR)</i>	$\frac{\text{Pendapatan Transfer}}{\text{Total Pendapatan Daerah}}$
3	<i>Administrative Expenditures to Total Expenditures (ADMIN)</i>	$\frac{(\text{Belanja Pegawai} + \text{Belanja Barang dan Jasa})}{\text{Total Belanja Daerah}}$
4	<i>Debt Level (DEBTLEVEL)</i>	ln (Total Kewajiban)
5	<i>Debt to Total Revenue (DEBTREV)</i>	$\frac{\text{Total Kewajiban}}{\text{Total Pendapatan Daerah}}$
6	<i>Entity Size (SIZE)</i>	ln (Total Pendapatan)
7	<i>Revenue Growth (GROWTH)</i>	$\frac{\text{Total Pendapatan Daerah}_t - \text{Total Pendapatan Daerah}_{(t-1)}}{\text{Total Pendapatan Daerah}_{(t-1)}}$
8	<i>Degree of Decentralization (DoD)</i>	$\frac{\text{Pendapatan Asli Daerah}}{\text{Total Pendapatan Daerah}}$
9	<i>Financial Independence (FI)</i>	$\frac{\text{Pendapatan Asli Daerah}}{\text{Total Belanja Daerah}}$
10	<i>Capital Outlay (CAPREV)</i>	$\frac{\text{Belanja Modal}}{\text{Total Pendapatan Daerah}}$
11	<i>Solvability (SOLVABILITY)</i>	$\frac{\text{Total Aset}}{\text{Total Kewajiban}}$
12	<i>Liquidity (LIQUIDITY)</i>	$\frac{\text{Total Aset Lancar}}{\text{Total Kewajiban Lancar}}$
13	<i>Leverage (LEVERAGE)</i>	$\frac{\text{Total Kewajiban}}{\text{Total Ekuitas Dana}}$

Lampiran 2. Parameter yang Digunakan pada Masing-masing Model.

No.	binarysplit	confidenceFactor	debug	minNumobj	numFolds	reduceErrorPruning	saveInstanceData	seed	subtreeraising	unpruned	uselaplace	testmode
1 ModelD1	FALSE	0.25	FALSE	2	3	FALSE	FALSE	1	TRUE	TRUE	FALSE	10-fold cross validation
2 ModelD2	FALSE	0.95	FALSE	20	3	FALSE	FALSE	1	TRUE	FALSE	FALSE	10-fold cross validation
3 ModelD3	FALSE	0.25	FALSE	2	3	FALSE	FALSE	1	TRUE	TRUE	FALSE	10-fold cross validation
4 ModelD4	FALSE	0.95	FALSE	20	3	FALSE	FALSE	1	TRUE	FALSE	FALSE	10-fold cross validation
5 ModelD5	FALSE	0.95	FALSE	30	3	FALSE	FALSE	1	TRUE	FALSE	FALSE	10-fold cross validation
6 ModelD6	FALSE	0.1	FALSE	2	3	FALSE	FALSE	1	TRUE	FALSE	FALSE	10-fold cross validation
7 ModelD7	FALSE	0.5	FALSE	2	3	FALSE	FALSE	1	TRUE	FALSE	FALSE	10-fold cross validation
8 ModelD8	FALSE	0.95	FALSE	2	3	FALSE	FALSE	1	TRUE	FALSE	FALSE	10-fold cross validation
9 ModelD9	FALSE	0.25	FALSE	2	3	FALSE	FALSE	1	TRUE	TRUE	FALSE	10-fold cross validation
10 ModelD10	FALSE	0.95	FALSE	20	3	FALSE	FALSE	1	TRUE	FALSE	FALSE	10-fold cross validation
11 ModelD11	FALSE	0.95	FALSE	30	3	FALSE	FALSE	1	TRUE	FALSE	FALSE	10-fold cross validation
12 ModelD12	FALSE	0.95	FALSE	50	3	FALSE	FALSE	1	TRUE	FALSE	FALSE	10-fold cross validation
13 ModelD13	FALSE	0.1	FALSE	2	3	FALSE	FALSE	1	TRUE	FALSE	FALSE	10-fold cross validation
14 ModelD14	FALSE	0.5	FALSE	2	3	FALSE	FALSE	1	TRUE	FALSE	FALSE	10-fold cross validation
15 ModelD15	FALSE	0.95	FALSE	2	3	FALSE	FALSE	1	TRUE	FALSE	FALSE	10-fold cross validation
16 ModelD16	FALSE	0.25	FALSE	2	3	FALSE	FALSE	1	TRUE	TRUE	FALSE	10-fold cross validation
17 ModelD17	FALSE	0.95	FALSE	20	3	FALSE	FALSE	1	TRUE	FALSE	FALSE	10-fold cross validation
18 ModelD18	FALSE	0.95	FALSE	30	3	FALSE	FALSE	1	TRUE	FALSE	FALSE	10-fold cross validation
19 ModelD19	FALSE	0.95	FALSE	50	3	FALSE	FALSE	1	TRUE	FALSE	FALSE	10-fold cross validation
20 ModelD20	FALSE	0.1	FALSE	2	3	FALSE	FALSE	1	TRUE	FALSE	FALSE	10-fold cross validation
21 ModelD21	FALSE	0.5	FALSE	2	3	FALSE	FALSE	1	TRUE	FALSE	FALSE	10-fold cross validation
22 ModelD22	FALSE	0.95	FALSE	2	3	FALSE	FALSE	1	TRUE	FALSE	FALSE	10-fold cross validation
23 ModelD23	FALSE	0.25	FALSE	2	3	FALSE	FALSE	1	TRUE	TRUE	FALSE	10-fold cross validation
24 ModelD24	FALSE	0.95	FALSE	20	3	FALSE	FALSE	1	TRUE	FALSE	FALSE	10-fold cross validation
25 ModelD25	FALSE	0.95	FALSE	30	3	FALSE	FALSE	1	TRUE	FALSE	FALSE	10-fold cross validation
26 ModelD26	FALSE	0.95	FALSE	50	3	FALSE	FALSE	1	TRUE	FALSE	FALSE	10-fold cross validation
27 ModelD27	FALSE	0.1	FALSE	2	3	FALSE	FALSE	1	TRUE	FALSE	FALSE	10-fold cross validation
28 ModelD28	FALSE	0.5	FALSE	2	3	FALSE	FALSE	1	TRUE	FALSE	FALSE	10-fold cross validation
29 ModelD29	FALSE	0.95	FALSE	2	3	FALSE	FALSE	1	TRUE	FALSE	FALSE	10-fold cross validation

Lampiran 3. Rincian Hasil Evaluasi *10-fold Cross Validation* Masing-masing Model

Model	Accuracy	TP Rate Class Distress	Kappa Statistic	AUC	Tree Size	Num of Leaves
D1	97,7536%	0,000	-0,0051	0,538	9	5
D2	98,0435%	0,000	0	0,460	1	1
D3	95,5935%	0,185	0,2236	0,732	19	10
D4	96,0199%	0,259	0,3145	0,733	9	5
D5	95,7356%	0,074	0,1035	0,703	9	5
D6	96,091%	0,056	0,0903	0,521	1	1
D7	95,5224%	0,185	0,22	0,741	19	10
D8	95,5224%	0,185	0,22	0,741	19	10
D9	94,8396%	0,370	0,4219	0,814	51	26
D10	95,1185%	0,321	0,4036	0,757	13	7
D11	94,0028%	0,198	0,2444	0,733	9	5
D12	93,7268%	0,111	0,1424	0,633	1	1
D13	94,9093%	0,309	0,3827	0,769	23	12
D14	94,8396%	0,358	0,4138	0,796	43	22
D15	93,8396%	0,370	0,4219	0,801	49	25
D16	93,8398%	0,556	0,5383	0,791	79	40
D17	93,9767%	0,463	0,5005	0,807	23	12
D18	93,8398%	0,315	0,4024	0,805	43	22
D19	93,4292%	0,380	0,4274	0,787	49	25
D20	93,5661%	0,509	0,5048	0,779	43	22
D21	93,9083%	0,565	0,5454	0,800	77	39
D22	93,9083%	0,565	0,5454	0,800	77	39
D23	92,8091%	0,563	0,5476	0,791	85	43
D24	92,3387%	0,496	0,4989	0,855	21	11
D25	91,0618%	0,385	0,3913	0,794	13	7
D26	91,8683%	0,452	0,4584	0,786	11	6
D27	93,3468%	0,481	0,5329	0,702	43	22
D28	92,9435%	0,563	0,5529	0,814	85	43
D29	92,8091%	0,563	0,5476	0,810	85	43