

Klasifikasi Kebangkrutan menggunakan *Multilayer Perceptron* dan *k-Nearest Neighbors*

I Putu Indra Aristya

School of Computing

Telkom University

Bandung, Indonesia

indraaristya@student.telkomuniversity.ac.id

Siti Saadah

School of Computing

Telkom University

Bandung, Indonesia

sitisaadah@telkomuniversity.ac.id

Ringkasan—Klasifikasi kebangkrutan yang dilakukan dari ringkasan laporan keuangan setiap perusahaan sangat berguna untuk dapat mempersiapkan dan membantu setiap perusahaan untuk menentukan keputusan secara cepat dan tepat. Dengan menggunakan bantuan metode - metode *machine learning* dapat dibuat model yang dapat menentukan apakah ada kemungkinan perusahaan bangkrut ataupun tidak. Model ini dibuat berdasarkan data terdahulu yang sudah berisi kelas data tersebut (bangkrut atau tidak). Metode yang digunakan adalah *k-Nearest Neighbors* dan *Multilayer Perceptron* dan didapatkan hasil terbaik menggunakan *k-Nearest Neighbors* dengan hasil pengujian sebesar 95,22%.

Index Terms—bangkrut, *multilayer perceptron*, *k-nearest neighbors*

I. PENDAHULUAN

Banyak perusahaan - perusahaan yang sedang berkembang ataupun perusahaan baru yang mulai berjalan setiap tahunnya. Menurut sensus ekonomi oleh Badan Pusat Statistik pada tahun 2016 tercatat 26,7 juta perusahaan yang ada di Indonesia. Angka tersebut mengalami kenaikan sebesar 3,98 juta yang sebelumnya adalah 22,7 juta pada tahun 2006. Lalu, menurut Konfederasi Serikat Pekerja Indonesia (KSPI) akan ada perusahaan yang mulai bangkrut pada tahun 2018 di Indonesia. Contohnya minimarket 7-Eleven yang sudah menutup semua gerainya di Indonesia. Kemudian, dengan mulainya pengurangan tenaga kerja mulai dari tahun 2015 hingga 2016 pada perusahaan - perusahaan besar

seperti PT Tosiha, PT Panasonic, PT Philips, PT DMC dan PT Osung.

Penyebab dari kebangkrutan sebuah perusahaan bisa disebabkan oleh banyak faktor dari dalam ataupun dari luar perusahaan tersebut. Salah satunya adalah kondisi keuangan yang tidak sehat dapat menyebabkan perusahaan bangkrut. Dari permasalahan dan faktor tersebut, dibuatlah model yang dapat mengklasifikasikan kemungkinan perusahaan tersebut akan bangkrut atau tidak dari kondisi keuangannya. Model ini dibuat dengan bantuan *machine learning* menggunakan metode *multilayer perceptron* dan *k-nearest neighbors* untuk dapat membandingkan dan melihat model yang cocok digunakan.

Pada paper ini akan dibahas hal - hal yang digunakan untuk membangun model dengan urutannya adalah pada Bagian 2 membahas penelitian yang sudah pernah dilakukan untuk kasus ini. Bagian 3 membahas metode - metode yang digunakan. Bagian 4 membahas data dan cara pengumpulan data yang digunakan. Bagian 5 adalah bagian dari skenario atau uji eksperimen yang dilakukan dan Bagian 6 akan memuat analisa dan kesimpulan pada permasalahan dan solusi yang dikerjakan.

II. TINJAUAN PUSTAKA

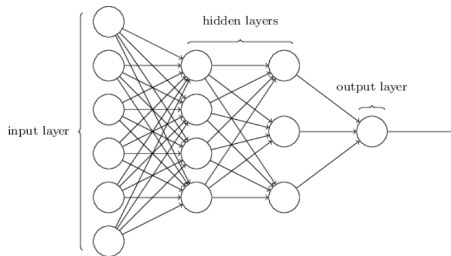
Penelitian yang berhubungan dengan klasifikasi kebangkrutan perusahaan sudah pernah dilakukan sebelumnya. Pada tahun 2011, David L. Olson [4] menggunakan beberapa model seperti *decision tree*,

neural network, *support vector machine* dalam penelitiannya. Dari penelitian tersebut didapatkan nilai akurasi tertinggi menggunakan metode *decision trees C5* sebesar 93.7%. Manil Wagle dkk. [1] berfokus pada metode untuk memilih fitur yang digunakan pada tahap klasifikasi. Digunakan metode *filter* dan *wrapper* serta semua fitur digunakan untuk klasifikasi. Dari hasil percobaannya, didapatkan hasil tertinggi adalah dengan menggunakan fitur yang didapatkan dari proses *filtering* dan klasifikasi menggunakan metode *neural network* dengan nilai akurasi sebesar 85.33%. Penelitian lainnya juga dilakukan oleh G.Pranav Naidu dkk. [2] yang menggunakan metode *neural network* dan *random forest*. Pada pemilihan fitur, Pranav menghitung keterkaitan antar fitur dengan menghitung nilai *co-occurrence*-nya. Dari penelitian tersebut didapatkan hasil tertinggi menggunakan metode *neural network* dengan *error rate* sebesar 4.4349%.

III. METODE PENELITIAN

A. Multilayer Perceptron

Multilayer Perceptron sering disebut juga sebagai MLP. MLP adalah salah satu jenis dari jaringan saraf tiruan yang terdiri dari *layer input*, *1 hidden layer* dan *layer output* [7].



Gambar 1. Arsitektur *Multilayer Perceptron*

Pada proses perhitungan MLP terlebih dahulu dihitung nilai *input* dengan nilai *weight* yang masuk ke dalam fungsi yang sering disebut *sum function*(1).

$$f(x) = w.x + b \quad (1)$$

Setelah itu akan dilanjutkan dengan hasil yang didapat dari proses *sum function* digunakan pada *hidden layer* tergantung dari fungsi aktivasi yang digunakan. Pada penelitian ini digunakan ReLU sebagai

fungsi aktivasi dimana nilai dari keluaran ReLU(2) ini adalah 0 atau nilai masukan.

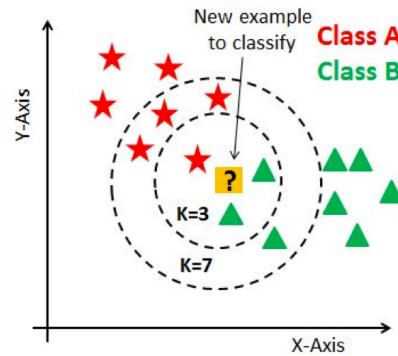
$$f(x) = \begin{cases} x, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases} \quad (2)$$

$$f(x)' = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{if } x < 0 \end{cases} \quad (3)$$

Kemudian akan dihitung *error* dan dilakukan pembelajaran mundur atau *backpropagation* untuk memperbaiki bobot dengan menghitung gradien dari *error*. Pada proses ini memerlukan turunan dari fungsi aktivasi ReLU(3).

B. k-Nearest Neighbors

k-Nearest Neighbors atau sering disingkat menjadi kNN adalah salah satu algoritma *machine learning* yang digunakan untuk proses klasifikasi. Metode kNN ini tidak melalui proses pelatihan atau *training* melainkan secara langsung memasuki proses pengujian. Dengan menggunakan kNN, akan dihitung jarak antar data yang kemudian dilihat sebanyak *k* data terdekat [6].



Gambar 2. Penggambaran *k-Nearest Neighbors*

Penghitungan jarak setiap data dapat dihitung menggunakan *euclidean distance*(4).

$$distance(a, b) = \sqrt{\sum (a - b)^2} \quad (4)$$

Algorithm 1 Algoritma kNN

Input: datatrain, datatest**Output:** kelas setiap datatestinisialisasi nilai k for i in datatest for j in datatrain hitung jarak datatest ke- i ke setiap datatrain

urutkan dari terkecil ke terbesar

 ambil k hasil urutan teratas kelas terbanyak adalah hasil kelas datatest ke- i

Kelas dari data yang dicari tersebut akan ditentukan dari hasil *voting* kelas k data terdekat. Algoritma dari metode kNN ditunjukkan pada Algorithm 1.

C. Principal Component Analysis

Principal Component Analysis (PCA) adalah salah satu metode yang dapat digunakan untuk mereduksi atribut pada suatu data. Cara mereduksi dari PCA membutuhkan perhitungan *covariance*(5)

$$cov(x, y) = \frac{\sum (xi - \bar{x})(yi - \bar{y})}{N} \quad (5)$$

Algorithm 2 Algoritma PCA

Input: data**Output:** data dengan atribut yang sesuai *principal component*inisialisasi nilai *principal component*hitung *covariance* dari datahitung *eigen vectors*urutkan *eigen vectors* berdasarkan *eigen values* dari besar ke kecilpilih *principal component* pertama *eigen vectors**transform* data asli sesuai data hasil reduksi

Dengan menggunakan PCA(2), atribut akan direduksi jumlahnya. PCA melihat atribut - atribut yang tidak memberikan informasi yang berbeda, misalnya ada atribut yang memiliki nilai 1 pada semua data sehingga atribut tersebut tidak berpengaruh untuk proses klasifikasi.

IV. DATA DAN PRE-PROCESSING

Data yang digunakan pada penelitian ini adalah data kebangkrutan dari perusahaan asuransi yang

diambil dari *Emerging Markets Information Service*(EMIS) [5]. Data ini terdiri dari *record* selama 5 tahun dengan 64 atribut setiap datanya dan jumlah yang berbeda setiap tahunnya. Berikut pada Tabel I adalah detail dari data yang digunakan.

Tabel I
JUMLAH DATA YANG DIGUNAKAN

Tahun ke-	Jumlah	Bankruptcy	Non-Bankruptcy
1	7027	271	6756
2	10173	400	9773
3	10503	495	10008
4	9792	515	9277
5	5910	410	5500

Pada penelitian ini, digunakan data pada tahun ke-4 karena perbedaan dari kelas bangkrut dan tidak bangkrut paling seimbang. Lalu, pada pengujian digunakan data tahun ke-1, 2, 3, dan 5.

Sebelum mulai untuk diolah, data - data yang digunakan masuk ke tahap *pre-processing*. *Pre-processing* yang dilakukan adalah menangani kelas data yang tidak seimbang. Data dengan kelas bangkrut pada setiap tahunnya selalu lebih kecil daripada kelas tidak bangkrut. Perbedaan jumlah tersebut dapat mempengaruhi model yang dibuat sehingga dilakukan *oversampling*. Pada proses *oversampling* ini akan dibuat data baru dengan kelas bangkrut. Pembuatan data tersebut menggunakan metode *k-nearest neighbors* dengan mencari 4 data terdekat dari suatu data kemudian dihitung nilai rata - rata setiap atributnya.

V. SKENARIO PENGUJIAN

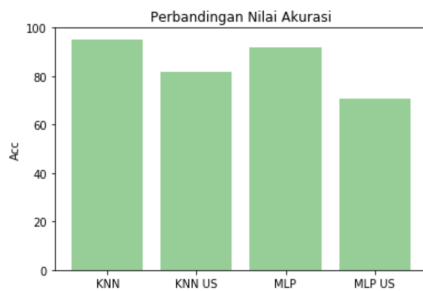
Dilihat dari persebaran data yang tidak seimbang, maka dilakukan proses *oversampling* pada *pre-process*. Dengan itu dicoba untuk melihat perbedaan hasil yang didapatkan saat data dilakukan *oversampling* dan tidak dilakukan *oversampling*. Maka dapat didefinisikan skenario yang dilakukan adalah sebagai berikut.

- melakukan proses pelatihan dan validasi dengan data yang dilakukan *oversampling* kemudian model yang didapatkan digunakan pada proses pengujian.
- melakukan proses pelatihan dan validasi dengan data yang tidak dilakukan *oversampling*

kemudian model yang didapatkan digunakan pada proses pengujian.

VI. HASIL DAN ANALISIS

Setelah mendapatkan model yang baik pada data latih dan data validasi, kemudian dilakukan pengujian menggunakan data uji. Data uji pada penelitian ini adalah data pada tahun ke-1, 2, 3 dan 5 dengan jumlah data sebanyak 33.613 data.



Gambar 3. Akurasi hasil pengujian

Pada pengujian ini didapatkan hasil terbaik menggunakan metode kNN dengan nilai k adalah 10. Kemudian hasil paling kecil didapatkan dengan metode MLP yang menggunakan data dengan *oversampling*. Seperti yang dapat dilihat pada Gambar 3, bahwa diketahui nilai akurasi dari model yang dilatih menggunakan data tanpa *oversampling* memiliki hasil yang lebih baik daripada dengan *oversampling*. Selain itu, dapat diketahui bahwa model terbaik (memiliki akurasi paling tinggi) adalah model dengan metode *k-Nearest Neighbors* dengan akurasi sebesar 95.22%.

Pada percobaan menggunakan MLP tidak ada parameter yang diatur secara langsung, parameter yang digunakan adalah parameter *default* yang diberikan oleh *library* sehingga membuat akurasinya tidak maksimal. Dan dengan menggunakan metode kNN, hanya perlu menentukan nilai k saja, sehingga tidak banyak parameter yang bisa diubah dan penentuannya sudah sesuai.

VII. KESIMPULAN

Pada percobaan ini digunakan 2 metode dan 2 skenario untuk pengujianya. Data uji diujikan pada

model kNN dan MLP yang dilatih dengan data tanpa *oversampling* dan dengan *oversampling*. Seperti hasil yang didapatkan, dapat disimpulkan bahwa dengan dilakukan *oversampling* hasil yang didapatkan lebih kecil daripada tidak dilakukan *oversampling*. Dan metode yang memiliki hasil pengujian tertinggi didapatkan menggunakan kNN dengan $k = 10$. Hal ini karena metode pembandingan MLP yang digunakan tidak mencoba beberapa arsitektur. Arsitektur yang dicoba hanya menggunakan 1 *hidden layer* dengan 100 *neuron*. Pada model kNN, data yang digunakan memiliki jumlah atribut hanya 30 (atribut awal 64) hasil reduksi menggunakan PCA. Atribut pada data berpengaruh jika metode klasifikasi yang digunakan kNN, karena kNN sangat rentan terhadap data dengan dimensi yang tinggi (semakin banyak dimensinya, semakin banyak kemungkinan data terdekat yang bisa saja sebenarnya tidak dekat). Selain itu, dengan PCA dapat dikurangi atribut yang tidak memberikan informasi perbedaan jarak setiap data sehingga penggunaan kNN tidak selalu lebih baik pada semua data.

PUSTAKA

- [1] W. Manil, Y. Zijiang and B. Younes, "Bankruptcy Prediction using Data Mining Techniques," International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES), vol. 8th, 2017.
- [2] G. Pranav Naidu and K. Govinda, "Bankruptcy Prediction using Neural Network," Proceedings of the Second International Conference on Inventive Systems and Control (ICISC), 2018.
- [3] L. Wei-Chao, L. Yu-Hsin and T. Chih-Fong, "Feature selection in single and ensemble learning based bankruptcy prediction models," 2018.
- [4] David L. Olson, D. Dursun and M. Yanyan, "Comparative analysis of data mining methods for bankruptcy prediction," Decision Support Systems, 2011.
- [5] Zieba, M., Tomczak, S. K., Tomczak, J. M. "Ensemble Boosted Trees with Synthetic Features Generation in Application to Bankruptcy Prediction. Expert Systems with Applications," 2016.
- [6] Suyanto. "Data Mining Untuk Klasifikasi dan Klusterisasi Data," 2017.
- [7] Suyanto. "Machine Learning Tingkat Dasar dan Lanjut," 2018.