

Pengaruh *Principal Component Analysis* terhadap Akurasi Model *Machine Learning* dengan Algoritma *Artificial Neural Network* untuk Prediksi Kebangkrutan Perusahaan

Fauzan Akmal Mahdi¹, Cahyo Adi Lukito², Dinda Parwita Aulia Nofri³, Verina Ardiyanti Madjid⁴, Desta Sandya Prasvita⁵

Program Studi Informatika / Fakultas Ilmu Komputer
Universitas Pembangunan Nasional Veteran Jakarta
Jl. R.S Fatmawati No. 1, Jakarta Selatan 12450

¹fauzanam@upnvj.ac.id, ²cahyoadi@upnvj.ac.id, ³dindaparwita@upnvj.ac.id, ⁴verinaam@upnvj.ac.id, ⁵desta.sandya@upnvj.ac.id

Abstrak. Prediksi akan kebangkrutan suatu perusahaan merupakan hal yang diteliti sampai saat ini. Seiring berkembangnya kemajuan ilmu pengetahuan dan teknologi informasi dan komunikasi mengubah cara memprediksi dari cara konvensional atau melakukan analisis dan penarikan kesimpulan secara manual menjadi lebih cepat dan bisa dilakukan dengan menggunakan komputer. Karena itu, kami melakukan prediksi kebangkrutan perusahaan menggunakan model *machine learning Artificial Neural Network*. Dataset yang kami gunakan adalah dataset yang dikumpulkan oleh Taiwan Economic Journal pada tahun 1999-2009 untuk mendefinisikan kebangkrutan perusahaan didasarkan kepada regulasi dari Pasar Saham Taiwan. Kami akan melakukan beberapa tahap untuk pembuatan model seperti pra-proses data, reduksi dimensi, ekstraksi fitur dan fitur seleksi serta klasifikasi. Hasil yang didapatkan dari penelitian ini dengan metode algoritma ANN tanpa tahap reduksi/PCA hasil akurasi sebesar 97.36%, lalu menggunakan algoritma ANN dan PCA yang memiliki akurasi sebesar 96.33%. Sehingga dapat disimpulkan bahwa akurasi yang terbesar adalah dengan menggunakan algoritma *Artificial Neural Network* tanpa tahap reduksi/PCA.

Kata Kunci: *Machine Learning*, Bangkrut, Perusahaan, *Artificial Neural Network*, Model

1. Pendahuluan

Seluruh bagian perusahaan adalah aspek-aspek yang mempengaruhi kehidupan perusahaan. Pada umumnya, aspek keuangan adalah hal-hal yang diperhitungkan sebagai aspek yang mempengaruhi kehidupan perusahaan seperti jumlah pendapatan, jumlah pengeluaran, jumlah keuntungan, jumlah pinjaman, aset-aset perusahaan, jumlah utang, jumlah piutang dan indikator-indikator privat dari perusahaan. Secara spesifik aspek tersebut akan dihitung pada tiap-tiap departemen. Setiap departemen akan diperhitungkan hasil pencapaiannya, perhitungan uang yang masuk dan yang keluar, dan pencapaian khusus sesuai departemen masing-masing. Hal yang sama dilakukan dari setiap divisi dalam suatu departemen. Akan muncul banyak fitur-fitur lainnya saat menghitung aspek keuangan.

Dengan diperhitungkannya aspek-aspek tersebut dapat dilakukan analisis dan penarikan kesimpulan untuk memprediksi kebangkrutan suatu perusahaan. Beberapa penelitian terkait prediksi kebangkrutan suatu perusahaan telah dilakukan sebelumnya. Pada paper *Review Of Bankruptcy Prediction Using Machine Learning And Deep Learning Techniques* [1] memiliki akurasi yang tinggi yaitu 87% ketika melakukan prediksi memakai teknik *Neural Network* (NN). Selanjutnya, pada paper *A neural network model for bankruptcy prediction* [2], prediksi kebangkrutan dengan metode ANN yang menggunakan sampel dari *Moody's Industrial Manuals* menyatakan bahwa *neural network* memiliki performa lebih baik dan lebih konsisten dengan tiga kali pengujian yang menghasilkan akurasi 81.48%, 77.78%, dan 77.78%. Lalu ada prediksi kebangkrutan dengan membandingkan lima metode *binary classification* pada paper *A binary classification method for bankruptcy prediction* [3] yang menghasilkan rasio hit sebesar 76.4% yang berarti terbaik ketiga untuk metode *neural network*.

Metodologi dalam *Machine Learning* yang telah dilakukan bertujuan agar mesin dapat belajar untuk membuat model yang efektif pada suatu kasus tertentu dengan mempelajari data-data yang ada untuk memprediksi keputusan di masa depan. Salah satu model yang akan diterapkan pada penelitian ini adalah model klasifikasi *Artificial Neural Network*. Selain metode yang tepat dibutuhkan juga fitur yang tepat untuk digunakan sebagai variabel prediksi.

Hal ini dikarenakan dimensi fitur yang terlalu luas dapat mengurangi tingkat akurasi karena adanya fitur-fitur yang sebenarnya tidak berpengaruh.

Untuk mendapatkan fitur yang tepat dapat dilakukan reduksi dimensi. Reduksi dimensi membantu mempersempit dimensi fitur agar ditemukan fitur yang tepat. Selain itu, reduksi dimensi juga membantu mengurangi dampak yang akan terjadi akibat dimensi yang terlalu luas. Namun, dengan banyaknya metode reduksi dimensi yang ada, sulit ditemukan metode yang paling tepat untuk sekumpulan data yang ada. Oleh karena itu pada penelitian ini akan dibandingkan beberapa metode reduksi dimensi untuk melakukan prediksi kebangkrutan perusahaan menggunakan metode ANN. Penelitian ini dilakukan dengan tujuan mendapatkan metode reduksi dimensi yang tepat untuk digunakan dengan metode ANN dalam memprediksi kebangkrutan perusahaan.

2. Penelitian Terdahulu

Penelitian yang menggunakan dataset *Company Bankruptcy Prediction* ini sudah beberapa kali digunakan untuk penelitian. Beberapa penelitiannya adalah sebagai berikut.

1. Jaime Becerra Guerrero. Melakukan penelitian model prediksi dengan dataset ini. Menggunakan *Random Forest Classifier* untuk menentukan reduksi dimensi fitur, mendapatkan 15 fitur yang menjadi titik shoulder pada hasil akurasi. Kemudian dengan melakukan analisis deskriptif untuk mencari pola dari data, didapatkan 7 fitur utama yang digunakan dalam proses prediksi. Prediksi dilakukan dengan metode KNN, yang menghasilkan akurasi sebesar 97%, dan dengan metode *Gradient Boosting Classifier*, yang menghasilkan akurasi sebesar 97%. [8]
2. Kabir. Melakukan penelitian model prediksi dengan menggunakan algoritma Naive Bayes. Penelitian ini melakukan dua kali percobaan prediksi, yaitu tanpa reduksi fitur dan dengan reduksi fitur. Pada hasil prediksi percobaan tanpa reduksi fitur, didapatkan hasil akurasi sebesar 83.4%. Sedangkan hasil akurasi pada percobaan dengan reduksi fitur didapatkan akurasi sebesar 83.5%, di mana fitur yang dipakai sebanyak 36 fitur. [9]

3. Landasan Teori

3.1. Normalisasi Data

Normalisasi data dapat diartikan sebagai pendekatan sistematis untuk meminimalkan redundansi data pada suatu database agar database tersebut dapat bekerja dengan optimal [4]. Normalisasi data bertujuan untuk menghilangkan dan mengurangi duplikasi data serta memastikan data berada pada tabel yang tepat. Pada penelitian ini, teknik normalisasi data yang akan digunakan adalah metode Min Max. Metode min-max adalah metode normalisasi data yang dilakukan dengan mentransformasi data asli secara linier. Masukan dari normalisasi data ini ialah dataset yang belum berada pada interval yang sama sehingga dapat mempengaruhi hasil penelitian. Keluaran dan normalisasi ini adalah dataset dengan data yang telah berada pada interval baru yang telah ditentukan. Secara matematis, rumus Normalisasi Min-Max dapat dituliskan sebagai berikut :

$$X = \frac{(new_{max} - new_{min})}{(max - min) + new_{min}}$$

Dimana *max* adalah Nilai maksimal data per kolom, *min* adalah Nilai minimum data per kolom, *new max* adalah Batas maksimum yang kita berikan, dan *new min* adalah Batas minimum yang kita berikan.

3.2. Artificial Neural Network (ANN)

ANN adalah sebuah alat pemodelan data statistik non-linier dan dapat digunakan untuk memodelkan hubungan yang kompleks antara input dan output untuk menemukan pola data. Neuron adalah bagian dasar dari pemrosesan suatu Neural Network (C, Baskoro, Ambarwati, & Wicaksana, 2013). Cara kerja Neural Network dapat dianalogikan sebagaimana halnya manusia belajar dengan menggunakan contoh atau yang disebut sebagai supervised learning.

3.3. Reduksi Dimensi

Reduksi dimensi adalah transformasi informasi digital numerik atau alfabet yang diperoleh secara empiris atau eksperimental menjadi bentuk yang dikoreksi, dipesan, dan disederhanakan. Tujuannya yaitu untuk menghindari kekurangan dari dimensi, membantu menghilangkan fitur yang tidak relevan dan mengurangi noise, serta mengatasi lamanya waktu yang dibutuhkan untuk menganalisis data yang kompleks dalam tiap dataset yang lengkap. Reduksi dimensi terbagi menjadi dua teknik, yaitu ekstraksi fitur dan seleksi fitur. Pada penelitian ini akan digunakan teknik ekstraksi fitur dengan metode PCA (Principal Component Analysis).

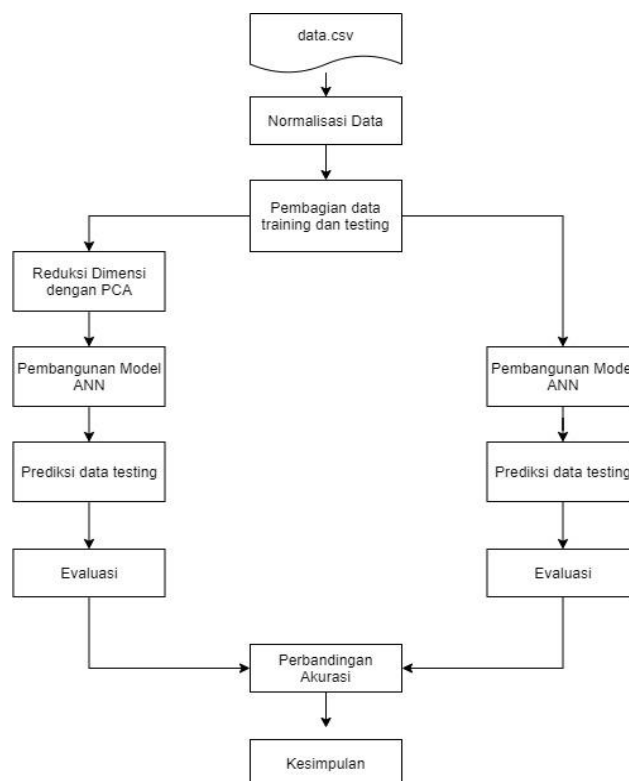
3.4. Principal Component Analysis (PCA)

Secara umum, PCA adalah suatu transformasi ruang vektor yang digunakan untuk mereduksi dimensi dari data yang sangat besar [5]. Dengan menggunakan pemetaan, suatu data asli yang mempunyai banyak variabel bisa diinterpretasikan menjadi beberapa variabel saja. Tujuan dari metode ini adalah untuk menemukan komponen-komponen utama yang akan digunakan pada tahap selanjutnya dari penelitian ini. PCA ini biasanya digunakan untuk mereduksi atau seleksi fitur pada data tanpa mengubah karakteristik data secara signifikan [6].

4. Metodologi

4.1. Flowchart Metodologi

Pendekatan yang dilakukan dalam paper ini adalah dengan membuat model klasifikasi Artificial Neural Network dengan menggunakan dataset yang tersedia pada situs kaggle. Dataset yang digunakan adalah dataset yang dikumpulkan oleh Taiwan Economic Journal pada tahun 1999 sampai 2009 untuk mendefinisikan kebangkrutan perusahaan didasarkan kepada regulasi dari Pasar Saham Taiwan (Taiwan Stock Exchange). Dataset berisi 95 fitur dan 1 kelas atau label. Tahapan pengolahan data pada penelitian ini dapat dilihat pada gambar 1.



Gambar. 1. Metodologi Penelitian

4.2. Deskripsi Dataset

Pencarian dataset *Company Bankruptcy Prediction* [7] didapatkan dari situs *Kaggle*. Pemilihan dataset ini didasarkan pada asumsi untuk mencari sebuah solusi modern untuk memprediksi kebangkrutan perusahaan dengan menggunakan teknologi komputer dan pemrograman. Dataset ini merupakan data yang dikumpulkan oleh *Taiwan Economic Journal* pada tahun 1999 sampai 2009. Kebangkrutan perusahaan merupakan bagian dari regulasi dalam pasar modal Taiwan (*Taiwan Stock Exchange*). Jumlah kolom dalam dataset *Company Bankruptcy Prediction* ada 95 kolom. Terdiri dari 1 kolom kelas *Bankrupt?* dan 95 kolom fitur. *Sample* dataset dapat dilihat pada tabel 1.

Tabel 1. Sample dataset yang berisi 6819 baris data dan 95 kolom fitur dan 1 kolom kelas.

| No | <i>Bankrupt?</i> | <i>ROA(C) before interest and depreciation before interest</i> | <i>ROA(A) before interest and % after tax</i> | <i>ROA(B) before interest and depreciation after tax</i> | <i>Operating Gross Margin</i> | <i>Realized Sales Gross Margin</i> | <i>Operating Profit Rate</i> | ... | <i>Equity to Liability</i> |
|----|------------------|--|---|--|-------------------------------|------------------------------------|------------------------------|-----|----------------------------|
| 1 | 1 | 0.370594 | 0.424389 | 0.40575 | 0.601457 | 0.601457 | 0.998969 | ... | 0.016469 |
| 2 | 1 | 0.464291 | 0.538214 | 0.51673 | 0.610235 | 0.610235 | 0.998946 | ... | 0.020794 |
| 3 | 1 | 0.426071 | 0.499019 | 0.472295 | 0.60145 | 0.601364 | 0.998857 | ... | 0.016474 |
| 4 | 1 | 0.399844 | 0.451265 | 0.457733 | 0.583541 | 0.583541 | 0.9987 | ... | 0.023982 |
| 5 | 1 | 0.465022 | 0.538432 | 0.522298 | 0.598783 | 0.598783 | 0.998973 | ... | 0.03549 |
| 6 | 1 | 0.38868 | 0.415177 | 0.419134 | 0.590171 | 0.590251 | 0.998758 | ... | 0.019534 |
| 7 | 0 | 0.390923 | 0.445704 | 0.436158 | 0.61995 | 0.61995 | 0.998993 | ... | 0.015663 |
| 8 | 0 | 0.508361 | 0.570922 | 0.559077 | 0.601738 | 0.601717 | 0.999009 | ... | 0.034889 |
| 9 | 0 | 0.488519 | 0.545137 | 0.543284 | 0.603612 | 0.603612 | 0.998961 | ... | 0.065826 |

| | | | | | | | | | |
|------|-----|----------|----------|----------|----------|----------|----------|-----|----------|
| 10 | 0 | 0.495686 | 0.550916 | 0.542963 | 0.599209 | 0.599209 | 0.999001 | ... | 0.030801 |
| | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 6819 | 0 | 0.493053 | 0.570105 | 0.549548 | 0.627409 | 0.627409 | 0.99808 | ... | 0.025953 |

Dapat dilihat pada tabel 1 bahwa dataset berisi 6819 baris data yang terbagi menjadi kelas bangkrut yang diberi label 1 pada kolom *Bankrupt?* dan kelas tidak bangkrut yang diberi label 0 pada kolom *Bankrupt?*. Masing masing data pada dataset ini memiliki 95 kolom fitur. Alat yang digunakan dalam penelitian ini adalah Google Collab dan Spyder. Spyder digunakan untuk membantu menampilkan visualisasi dari data yang telah diuji.

4.3. Normalisasi Data

Pada proses penelitian, tahap *preprocessing* data yang dilakukan adalah tahap normalisasi dataset. Tahap *preprocessing* lainnya seperti *cleaning* tidak dilakukan, hal ini disebabkan dataset yang dimiliki telah memenuhi persyaratan dalam tahap *preprocessing* yang lain. Sehingga tahapan *preprocessing* yang perlu dilakukan adalah normalisasi data agar pengolahan data lebih mudah untuk dibaca. Normalisasi dataset dilakukan untuk menyamaratakan interval pada fitur-fitur yang ada dalam dataset. Dalam tahap ini, normalisasi diterapkan pada seluruh kolom fitur dengan interval 0 sampai 1. Teknik yang digunakan untuk melakukan normalisasi dataset adalah menggunakan teknik Min Max. Hasil dari normalisasi data ini ialah keluaran data yang berada pada interval yang telah ditentukan, yaitu pada interval 0 sampai 1. Hasil normalisasi data yang dilakukan dapat dilihat pada tabel 2.

Tabel 2. Sampel data hasil normalisasi data. Dapat dilihat pada tabel bahwa data hasil normalisasi telah berada pada interval 0 sampai 1.

| NO | <i>ROA(C) before interest and depreciation before interest</i> | <i>ROA(A) before interest and % after tax</i> | <i>ROA(B) before interest and depreciation after tax</i> | <i>Operating Gross Margin</i> | <i>Realized Sales Gross Margin</i> | <i>Operating Profit Rate</i> | | <i>Equity to Liability</i> |
|----|--|---|--|-------------------------------|------------------------------------|------------------------------|-------|----------------------------|
| 1 | 0.370594 | 0.424389 | 0.405750 | 0.601457 | 0.601457 | 0.998969 | | 0.016469 |
| 2 | 0.464291 | 0.538214 | 0.51673 | 0.610235 | 0.610235 | 0.998946 | | 0.020794 |
| 3 | 0.426071 | 0.499019 | 0.472295 | 0.60145 | 0.601364 | 0.998857 | | 0.016474 |

| | | | | | | | | |
|-------|----------|----------|----------|----------|----------|----------|-------|----------|
| 4 | 0.399844 | 0.451265 | 0.457733 | 0.583541 | 0.583541 | 0.9987 | | 0.023982 |
| 5 | 0.465022 | 0.538432 | 0.522298 | 0.598783 | 0.598783 | 0.998973 | | 0.03549 |
| | | | | | | | | |
| 6819 | 0.493053 | 0.570105 | 0.549548 | 0.627409 | 0.627409 | 0.99808 | | 0.025953 |

4.4. Pembagian Data Training dan Testing

Pembagian data dalam penelitian ini dipetakan menjadi 80% data training dan 20% data testing. Data training akan digunakan untuk pembangunan model sedangkan data testing akan digunakan untuk menguji model yang telah kita bangun menggunakan data training. Pembagian data training dan data testing didasarkan pada kebutuhan data training yang lebih banyak, karena semakin banyak data yang digunakan untuk melatih model maka model yang dibangun juga semakin baik.

Berdasarkan pembagian yang telah dilakukan, didapatkan data training sebanyak 5455 baris dan data testing sebanyak 1364 baris. Hasil ini sesuai dengan perbandingan data training dan data testing yang diinginkan, yaitu 80% dan 20%.

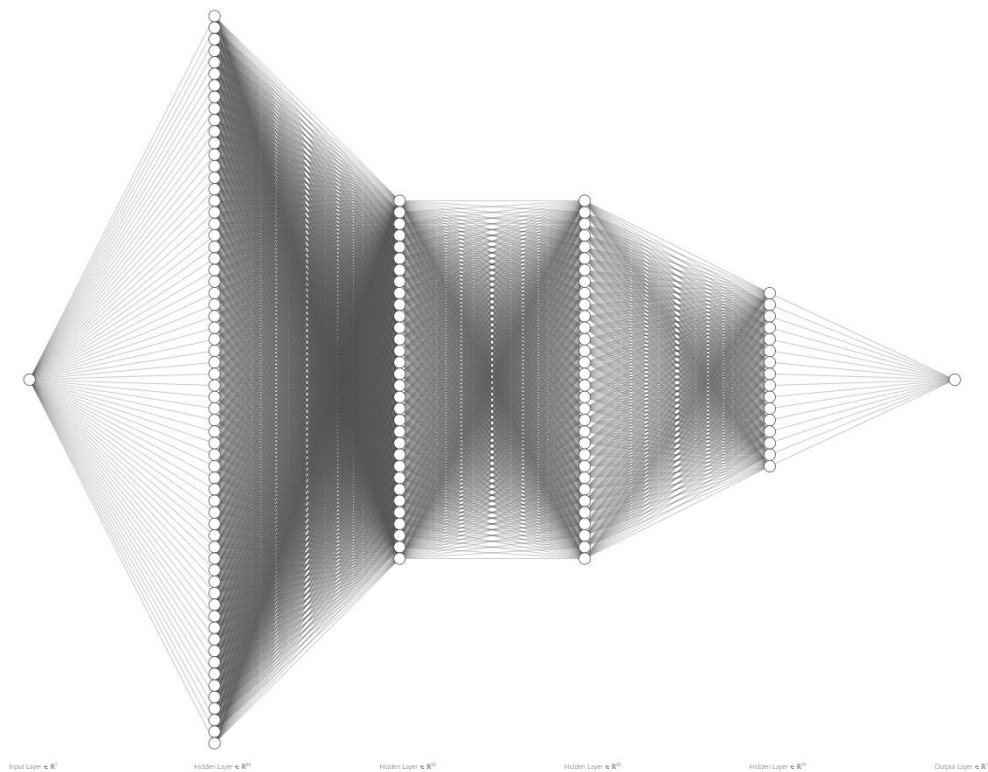
4.5. Reduksi Dimensi

Tahap reduksi dimensi dalam penelitian ini menggunakan teknik PCA (*Principal Component Analysis*). Hal ini dikarenakan PCA merupakan bagian dari praproses data untuk mereduksi dimensi suatu set fitur dari data agar data menjadi lebih ringkas untuk dilakukan pemodelan. Terdapat cara untuk mereduksi dimensi fitur, yaitu dengan cara ekstraksi fitur dan seleksi fitur. Tujuan dari metode ini adalah untuk menemukan komponen-komponen utama yang akan digunakan pada tahap selanjutnya dari penelitian ini.

Pada PCA, ekstraksi fitur dilakukan penghitungan keragaman data yang direpresentasikan oleh nilai eigen / eigenvalue. Proses PCA akan mengurangi dimensi data, pengurangan dimensi data akan mempengaruhi model yang dibuat serta hasil prediksi yang dihitung. Pada tahap ini, ditetapkan jumlah kolom yang akan direduksi menjadi 10 kolom fitur dari 95 fitur yang telah ada.

4.6. Artificial Neural Network (ANN)

Pembangunan model untuk memprediksi data menggunakan library MLP (*Multi Layer Perceptron Classifier*). Pada tahap ini ditentukan jumlah layer input sebanyak 1 layer dengan 1 unit, 4 hidden layer dengan masing-masing jumlah unit sebanyak 64 unit neuron, 32 neuron, 32 neuron, 16 neuron, dan 1 layer output. Arsitektur dari ANN yang dilakukan pada penelitian ini dapat dilihat pada gambar 2.



Gambar. 2. Arsitektur ANN dengan 1 input layer, 4 hidden layer dengan masing-masing jumlah unit sebanyak 64 unit neuron, 32 neuron, 32 neuron, 16 neuron, dan 1 layer output.

Pembangunan model dilakukan dengan menggunakan fungsi aktivasi *logistic*. Pemodelan ini dilakukan dua kali, yaitu pada data yang telah dilakukan reduksi dimensi dengan PCA dan pada data asli yang belum dilakukan reduksi dimensi.

4.7. Evaluasi Model

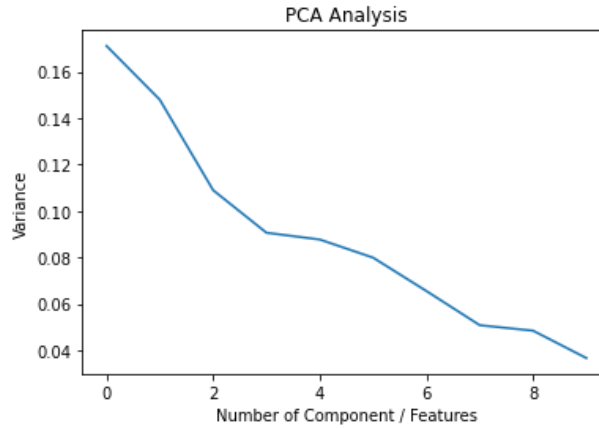
Model yang sudah siap akan dilakukan uji coba untuk dilakukan optimasi agar mendapatkan keakuratan yang tinggi untuk berbagai data latih. Prosedur ini dilakukan agar mendapatkan model yang terbaik dari hasil latih menggunakan data yang sudah ada untuk memprediksi data-data selanjutnya di waktu yang akan datang. Terakhir adalah tahapan perhitungan akurasi dari kedua model. Nantinya kedua hasil klasifikasi tersebutlah yang akan dibandingkan akurasinya.

5. Hasil dan Pembahasan

Pada bab ini akan membahas hasil penelitian mengenai pengaruh reduksi dimensi pada algoritma ANN. Seperti yang telah dijelaskan pada bab sebelumnya, metode reduksi dimensi yang digunakan pada penelitian ini adalah *Principal Component Analysis* dan menggunakan library MLP (*Multi Layer Perceptron*) *Classifier* untuk membangun model ANN.

Karena tujuan penelitian ini adalah membandingkan akurasi dari model ANN dengan menggunakan reduksi dimensi dan tidak menggunakan reduksi dimensi, maka akan dilakukan dua tahap pemodelan. Pada pemodelan pertama, data yang telah dilakukan normalisasi dan dibagi menjadi data training dan testing akan langsung digunakan untuk pembangunan dan pengujian model. Hasil akurasi dari pemodelan tanpa dilakukan reduksi dimensi terlebih dahulu yaitu sebesar 0,9736070381231672 atau 97,36%.

Setelah didapatkan akurasi dari pemodelan tanpa reduksi dimensi, selanjutnya dilakukan pemodelan dengan menerapkan reduksi dimensi dengan teknik PCA terlebih dahulu. Pemeringkatan fitur dari PCA ini dapat dilihat pada gambar 3.



Gambar 3. Analisis pemeringkatan fitur PCA.

Berdasarkan hasil pemeringkatan tersebut, ditetapkan jumlah kolom yang akan direduksi menjadi 10 kolom fitur dari 95 fitur yang telah ada. Hal ini dikarenakan berdasarkan hasil pemeringkatan diatas, penggunaan 10 fitur merupakan titik shoulder dimana hasil percobaan dengan jumlah fitur yang mendekati 10 hasilnya tidak memiliki perbedaan yang signifikan. Hasil reduksi dimensi dari data yang akan digunakan pada pemodelan kedua dapat dilihat pada Gambar 4.

| | pc1 | pc2 | pc3 | pc4 | pc5 | pc6 | pc7 | pc8 | pc9 | pc10 |
|------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 0 | 0.389698 | -0.196544 | -0.061993 | -0.341714 | -0.205487 | -0.155309 | -0.181974 | -0.250419 | -0.271058 | -0.290549 |
| 1 | 0.481681 | -0.773313 | 0.618697 | 0.084185 | -0.366930 | -0.109342 | -0.083648 | 0.010470 | -0.413458 | 0.082511 |
| 2 | -0.071018 | -0.412986 | -0.337838 | -0.074583 | 0.048953 | -0.203020 | -0.210611 | 0.009771 | -0.129327 | 0.166156 |
| 3 | 0.685777 | -0.602494 | 0.343644 | -0.228268 | -0.484178 | -0.132216 | -0.056839 | 0.082115 | 0.763171 | 0.180516 |
| 4 | 0.620666 | 0.494069 | -0.207719 | 0.189122 | -0.247582 | -0.154526 | -0.074669 | 0.060977 | 0.257795 | -0.372660 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 6814 | -0.274150 | -0.064005 | -0.202143 | -0.149684 | 0.168463 | -0.199465 | 0.203945 | -0.021964 | -0.096028 | -0.183629 |
| 6815 | -0.087149 | -0.136130 | 0.601881 | 0.650484 | 0.074065 | 0.418468 | -0.137672 | 0.192984 | 0.119567 | -0.122493 |
| 6816 | -0.252306 | -0.133870 | -0.331603 | -0.129220 | -0.041115 | -0.231741 | -0.149012 | 0.154774 | -0.277667 | 0.187276 |
| 6817 | -0.394147 | -0.092358 | 0.160140 | 0.040519 | -0.016016 | -0.261888 | 0.068668 | 0.032757 | 0.020108 | -0.188401 |
| 6818 | -0.408147 | 0.121739 | -0.147847 | -0.077720 | 0.178023 | -0.474708 | -0.209888 | -0.091617 | 0.045110 | -0.169923 |

Gambar 4. Hasil reduksi dimensi dengan PCA menghasilkan 10 fitur yang akan digunakan pada pemodelan selanjutnya.

Setelah reduksi dimensi dilakukan, selanjutnya data hasil reduksi dimensi akan melewati tahap pemodelan seperti pemodelan pertama yaitu dengan library MLP. Hasil akurasi dari pemodelan dengan reduksi dimensi adalah sebesar 0,9633431085043989 atau 96,33%.Detail perhitungan akurasi dari pemodelan ini dapat dilihat pada tabel 3.

Tabel 3. Perhitungan akurasi pemodelan ANN dengan reduksi dimensi PCA.

| | <i>precision</i> | <i>recall</i> | <i>f1-score</i> | <i>support</i> |
|---------------------|------------------|---------------|-----------------|----------------|
| 0 | 0.97 | 1.00 | 0.98 | 1317 |
| 1 | 0.00 | 0.00 | 0.00 | 47 |
| <i>accuracy</i> | | | 0.97 | 1364 |
| <i>macro avg</i> | 0.48 | 0.50 | 0.49 | 1364 |
| <i>weighted avg</i> | 0.93 | 0.97 | 0.95 | 1364 |

Setelah kedua pemodelan dilakukan, selanjutnya akan dibandingkan hasil evaluasi dari kedua model, yaitu dengan membandingkan akurasi dari masing masing model. Model pertama menggunakan 95 fitur dan model kedua menggunakan 10 fitur hasil dari reduksi dimensi dengan PCA. Perbandingan dari kedua model dapat dilihat pada tabel 4.

Tabel 4. Perbandingan akurasi berdasarkan penggunaan PCA dan tidak menggunakan PCA.

| Pemodelan | Akurasi |
|------------------|--------------------|
| ANN tanpa PCA | 0,9736070381231672 |
| ANN dengan PCA | 0.9655425219941349 |

6. Simpulan dan Saran

Berdasarkan hasil dari eksperimen penelitian Pengaruh *Principal Component Analysis* terhadap Akurasi Model Machine Learning dengan Algoritma Artificial Neural Network untuk Prediksi Kebangkrutan Perusahaan didapatkan hasil bahwa akurasi machine learning dengan algoritma ANN tanpa tahap reduksi/PCA lebih baik dengan hasil akurasi sebesar 97,36% dibandingkan dengan algoritma ANN ditambah PCA yang memiliki akurasi sebesar 96,33 %. Oleh karena itu dapat disimpulkan bahwa dari kedua model yang dilakukan maka tahap reduksi dimensi menggunakan PCA pada algoritma ANN mempengaruhi hasil akurasi model untuk memprediksi kebangkrutan perusahaan. Pengurangan atau reduksi ukuran dimensi dari fitur data yang akan diuji akan mengurangi akurasi dari model klasifikasi yang telah dibangun.

Penelitian selanjutnya diharapkan dapat memperbaiki hasil yang didapat pada penelitian ini dengan menggunakan teknik reduksi dimensi lainnya untuk memperbaiki akurasi algoritma ANN.

Referensi

- [1] Qu, Y., Quan, P., Lei, M. and Shi, Y., 2019. Review of bankruptcy prediction using machine learning and deep learning techniques. *Procedia Computer Science*, 162, pp.895-899.
- [2] Odom, M.D. and Sharda, R., 1990, June. A neural network model for bankruptcy prediction. In *1990 IJCNN International Joint Conference on neural networks* (pp. 163-168). IEEE.
- [3] Min, J.H. and Jeong, C., 2009. A binary classification method for bankruptcy prediction. *Expert Systems with Applications*, 36(3), pp.5256-5263.
- [4] Ihsan, M.A. (2018). REDUKSI ATRIBUT PADA ALGORITMA K-NEAREST NEIGHBOR (KNN) DENGAN MENGGUNAKAN ALGORITMA GENETIKA.
- [5] Kosasih, R. (2014). PENGGUNAAN METODE PCA UNTUK REDUKSI DATA IMAGE PEMBULUH DARAH VENA.
- [6] Adnyana, I.G.N.D., Arjuna, R.M., Indraini, A.N., Prasvita, D.S. (2021) . Pengaruh Seleksi Fitur Pada Algoritma Machine Learning Untuk Memprediksi Pembatalan Pesanan Hotel. *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*.
- [7] <https://www.kaggle.com/fedesoriano/company-bankruptcy-prediction> (diakses pada 20 Juni 2021).
- [8] <https://www.kaggle.com/jaimebecerraguerrero/simple-yet-powerful-bankrupt-prediction-model> (diakses pada 27 Agustus 2021).
- [9] <https://www.kaggle.com/kabirnapal/bankruptcy-prediction-naive-bayes-auc-96> (diakses pada 27 Agustus 2021).