

Assalamualaikum wr wb.perkenalkan nama saya Nabila Ayu Talita. Pada kesempatan kali ini saya akan mempresentasikan hasil review paper mengenai optimasi yang ada di machine learning.

Untuk jurnal yang akan saya review berjumlah dua jurnal, untuk jurnal yang pertama adalah:

JURNAL 1:

- Untuk jurnal pertama yang saya review berjudul 'Optimasi Klasifikasi Data Stunting Melalui Ensemble Learning pada Label Multiclass dengan Imbalance Data', yang diterbitkan tahun 2024. Jurnal ini ditulis oleh Eko Prasetyo dan Kristiawan Nugroho.

- Permasalahan utama yang dibahas pada jurnal ini adalah mengenai stunting, yang masih menjadi isu kesehatan serius di Indonesia. Penulis menjelaskan bahwa data stunting sangat kompleks dan tidak seimbang, sehingga cukup sulit diolah. Selain itu, penelitian sebelumnya terkait machine learning untuk stunting masih terbatas dan performa model yang ada juga belum optimal. Jadi, jurnal ini mencoba mengatasi masalah data yang besar, variabel yang banyak, dan distribusi kelas yang tidak merata.

- Pada bagian metode, penulis menggunakan tiga model klasifikasi tunggal yaitu Decision Tree, Naïve Bayes, dan k-NN. Setelah itu, model-model ini dioptimalkan menggunakan beberapa teknik ensemble learning, seperti Bagging, Boosting (AdaBoost, Gradient Boosting, XGBoost), dan Stacking. Untuk menangani ketidakseimbangan data, penulis menggunakan teknik RUSBoostClassifier, yang membantu menyeimbangkan sampel pada setiap kelas.

- Dataset yang digunakan terdiri dari 2.578 data pertumbuhan balita dari Puskesmas Tlogowungu. Data ini sangat tidak seimbang, misalnya kelas Normal berjumlah lebih dari dua ribu, sedangkan kelas Tinggi hanya ada satu data. Atribut yang digunakan cukup lengkap, seperti jenis kelamin, umur, berat badan lahir, tinggi lahir, berat badan, tinggi badan, Z-Score, dan status stunting sebagai label multiclass.

- Untuk hasilnya, pada model tunggal, Naïve Bayes memberikan akurasi terbaik yaitu 67,20%. Setelah menggunakan metode ensemble, performanya meningkat cukup signifikan. Metode terbaik adalah Bagging Decision Tree dengan akurasi 78,93%, precision 78%, dan recall 77,99%.

Beberapa ensemble lain juga meningkat, tetapi tidak semuanya lebih baik dari model tunggal—seperti AdaBoost yang hasilnya justru turun.

- Kontribusi utama jurnal ini adalah menunjukkan bahwa teknik ensemble learning, khususnya Bagging Decision Tree, bisa memberikan performa yang jauh lebih baik untuk klasifikasi data stunting yang bersifat multiclass dan tidak seimbang. Penelitian ini juga memberikan analisis mendalam menggunakan berbagai metrik, serta menunjukkan bahwa tidak semua metode ensemble cocok untuk kondisi data tertentu.
- kelebihan/kelemahan dari jurnal ini adalah
- Secara keseluruhan, jurnal ini memberikan kontribusi yang cukup besar terutama dalam penerapan ensemble learning pada kasus kesehatan anak, khususnya stunting. Hasil penelitian ini dapat menjadi acuan untuk pengembangan model klasifikasi yang lebih robust di masa depan.

JURNAL 2:

- Untuk jurnal kedua yang saya review berjudul ‘Optimasi Model Artificial Neural Network Untuk Klasifikasi Paket Jaringan’. Jurnal ini diterbitkan pada tahun 2020, dan ditulis oleh Indra Gunawan.
- Permasalahan utama dalam penelitian ini adalah bagaimana melakukan klasifikasi paket jaringan secara akurat sekaligus cepat, tanpa memperlambat lalu lintas internet. Semakin banyak variabel yang dipakai untuk klasifikasi, prosesnya menjadi lebih akurat tetapi lebih lambat. Sebaliknya, semakin sedikit variabel, proses makin cepat tapi akurasinya menurun. Penulis ingin menguji sejauh mana Artificial Neural Network mampu melakukan klasifikasi berbasis per-frame dan sekaligus mencari model ANN yang paling optimal untuk kondisi data jaringan yang kompleks.
- Metode yang digunakan pada penelitian ini adalah algoritma Artificial Neural Network. Dataset diperoleh melalui penyadapan jaringan kampus selama 15 menit pada jam sibuk menggunakan Wireshark. Setelah itu datanya diekspor, dibersihkan dari paket noise, dan diseimbangkan menggunakan dua teknik yaitu SMOTE untuk oversampling dan NEARMISS untuk undersampling.

Penulis melakukan proses optimasi model secara trial and error dengan mengubah berbagai parameter, seperti jumlah neuron pada setiap layer,

jumlah hidden layer, fungsi aktivasi, metode optimasi, nilai dropout, ukuran batch, sampai jumlah epoch.

- Dataset yang digunakan sangat besar, yaitu 990.558 baris data. Variabel input yang digunakan ada empat, yaitu protocol, port, time before previous packet, dan packet length. Outputnya terdiri dari lima label: App, Social Media, Game, Browsing, dan Streaming.

Penulis juga membagi dataset menjadi training, test, dan validation, serta melakukan penyeimbangan label supaya setiap kelas memiliki jumlah data yang lebih seragam.

- Hasil penelitian menunjukkan bahwa model ANN yang paling optimal memiliki struktur: satu input layer dengan 28 neuron, satu hidden layer dengan 10 neuron, dan satu output layer. Parameter terbaiknya menggunakan dropout 0.33, fungsi aktivasi tanh-tanh-softmax, dan optimizer adamax.

Model terbaik menghasilkan nilai loss terendah 0.32 pada data test dan validation. Stabilitas akurasi tercapai di iterasi ke-600, sedangkan stabilitas loss baru tercapai di epoch ke-1000.

Penulis juga menemukan bahwa semakin banyak kolom input, semakin baik hasilnya—empat kolom input menghasilkan loss 0.38 dibandingkan satu sampai tiga kolom yang loss-nya lebih tinggi.

- Kontribusi penting dari penelitian ini adalah menunjukkan bahwa ANN bisa digunakan untuk klasifikasi paket jaringan per-frame, meskipun akurasinya cukup menantang. Penelitian ini juga memberikan panduan yang sangat praktis, yaitu rumus spesifik untuk menentukan jumlah neuron yang optimal pada setiap layer. Selain itu, paper ini memberi baseline yang dapat digunakan untuk penelitian lanjutan di bidang klasifikasi jaringan.

- Secara keseluruhan, jurnal ini memberikan kontribusi yang cukup besar terutama dalam optimasi model ANN untuk klasifikasi jaringan. Penelitian ini bisa menjadi pijakan awal untuk penelitian serupa, terutama terkait arsitektur neuron dan parameter ANN untuk data jaringan per-frame.