

Early Results of Comparison between K-Nearest Neighbor and Artificial Neural Network Method for Facies Estimation

Hasil Awal Perbandingan antara *K-Nearest Neighbor* dan *Artificial Neural Network Method* untuk *Facies Estimation*

Hadyan Pratama^{1,*}, Loris Alif Syahputra¹, M. Fauzan Albany¹, Agus Abdullah¹, Epo Kusuma², Weny Astuti³, Bambang Mujihardi⁴, Sandy Kurniawan Suhardja¹

¹Department of Geophysical Engineering University of Pertamina, Jl. Teuku Nyak Arief, Simprug, Kebayoran Lama, Jakarta 12220

²Department of Geological Engineering University of Pertamina, Jl. Teuku Nyak Arief, Simprug, Kebayoran Lama, Jakarta 12220

³Department of Petroleum Engineering University of Pertamina, Jl. Teuku Nyak Arief, Simprug, Kebayoran Lama, Jakarta 12220

⁴Upstream Research and Technology – RTC PT. Pertamina (Persero), Jakarta 10110

*Email: hadyan.pratama@rocketmail.com

Submit: 2020-05-12; Revised: 2020-07-10; Accepted: 2020-08-18

Abstract: *Artificial Intelligence method has been widely used recently in many aspects to understand big data. Fundamentally, the purpose of Artificial Intelligence is to solve nonlinear problem. Most methods are trying to optimize an output from one or many inputs parameter by identifying any potential patterns that fit or using a statistical data. In Oil & Gas industry, one of the main challenges that can be solved by Artificial Intelligence is estimating facies from well log or seismic data. The main scope of this study is estimating lithofacies by analyzing well logs input using two different methods, K-Nearest Neighbor (KNN) and Artificial Neural Network (ANN). We employed various well log data such as gamma-ray, resistivity, neutron density porosity, and photoelectric effect from well log data at Panoma Council Grove Field, South West Kansas, United States. This study shows that using optimized parameters, KNN method faster than ANN method but, ANN give result better than KNN. Nevertheless, despite the fact this research could estimate lithologies, many aspect should be considered in order to reach optimum result such as insights from geological regional models.*

Keywords: *Artificial Intelligence, K-Nearest Neighbors, Artificial Neural Network, Facies Estimation*

Abstrak: *Artificial Intelligence* atau kecerdasan buatan, saat ini telah digunakan secara luas dibanyak bidang untuk membantu memahami data dengan ukuran yang besar. Secara umum, *Artificial Intelligence* memiliki prinsip untuk mencari solusi dari suatu permasalahan *nonlinear*. Sebagian besar metode ini mencoba mengoptimalkan output dari satu atau beberapa parameter input dengan cara menemukan kemungkinan pola yang cocok ataupun dengan menggunakan data statistik. Dalam industri minyak dan gas, salah satu tantangan yang dapat diselesaikan oleh *Artificial Intelligence* ini adalah mengestimasi *fasies* dari data *well log* atau seismik. Fokus penelitian ini adalah mengestimasi *fasies* dengan menganalisis input data log sumur menggunakan dua metode yang berbeda, *K-Nearest Neighbor (KNN)* dan *Artificial Neural Network (ANN)*. Kami meng-

gunakan input data berupa beberapa parameter log sumur seperti *gamma ray*, *resistivity*, *neutron density porosity*, dan *photoelectric effect log* dari sumur yang berada pada Lapangan Panoma Council Grove, Barat Daya Kansas, Amerika Serikat. Hasil studi ini menunjukkan bahwa dengan parameter yang telah dioptimalkan, metode ANN memerlukan waktu *running process* yang lebih lama dan menunjukkan hasil yang cukup baik dalam mengestimasi *fasies*. Sedangkan metode KNN memerlukan waktu *running process* yang relatif singkat dan menunjukkan hasil estimasi yang *overestimate*. Namun, meskipun dalam penelitian ini, baik metode KNN maupun ANN dapat melakukan estimasi *fasies* dengan akurasi yang cukup baik, beberapa hal harus dipertimbangkan untuk mendapat hasil yang lebih optimal dan lebih akurat seperti pendapat ahli serta model geologi regional wilayah penelitian tersebut.

Kata kunci: *Artificial Intelligence, K-Nearest Neighbors, Artificial Neural Network, Estimasi Fasies*

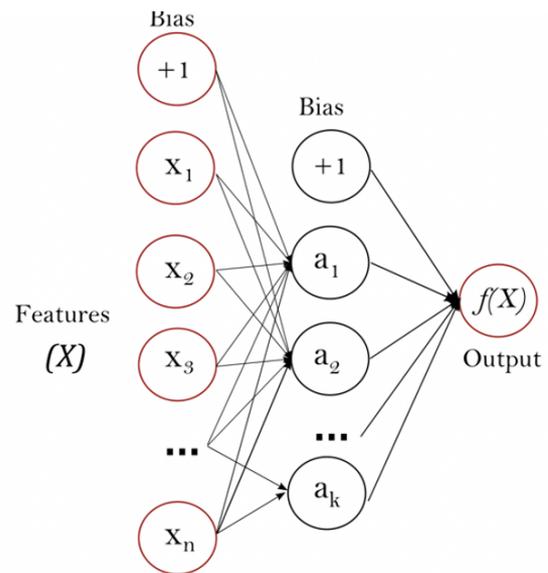
1 PENDAHULUAN

Machine learning merupakan salah satu cabang disiplin ilmu dari *Artificial Intelligence* yang mencakup perancangan dan pengembangan algoritma yang memungkinkan suatu program komputer untuk menganalisa dan mengenali pola atau pun perilaku dengan mengamati data data empiris. Fokus dari *Machine Learning* adalah bagaimana mesin atau dalam hal ini algoritma mengenali secara otomatis pola kompleks dan membuat keputusan berdasarkan data yang telah dilatih sebelumnya. Istilah pembelajaran pertama kali muncul dalam disiplin ilmu kecerdasan buatan dimana kata *learning* berarti menambah pengetahuan, dengan cara memahami, belajar, dan mengikuti perintah. Pada tahun 1959, Arthur Samuel mendefinisikan bahwa *machine learning* adalah bidang studi yang memberikan kemampuan untuk belajar tanpa diprogram secara eksplisit (Koza

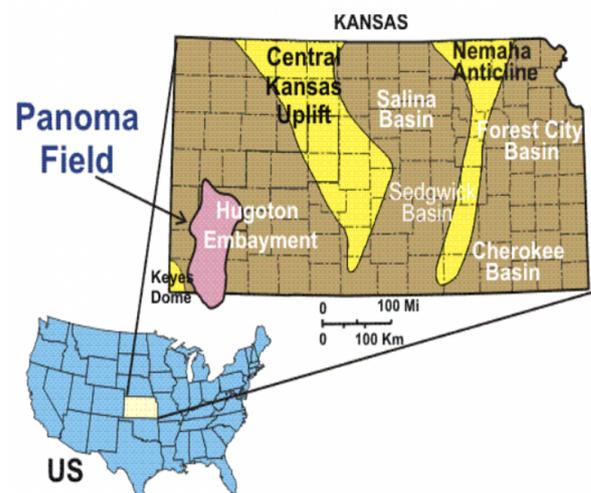
dkk., 1996). Implementasi kemampuan belajar dapat dicapai dengan berbagai teknik, sebagai contoh: dengan pendekatan kaidah, dengan pendekatan statistika, atau pun dengan menggunakan pendekatan fisiologi seperti sistem saraf manusia atau disebut dengan *Artificial Neural Network (ANN)*. Kelebihan dalam menggunakan metode *Machine Learning* dalam dataset yang teroptimisasi adalah waktu pemrosesan yang cepat dengan akurasi cukup baik meski memiliki data yang kompleks sehingga biaya untuk processing lebih murah. Namun, teknik *Machine Learning* juga mempunyai kelemahan. Diantaranya adalah dataset yang digunakan untuk melatih model kadang cukup kecil sehingga ketika model diaplikasikan untuk memprediksi data real, ada kemungkinan memberikan hasil yang bias sehingga memungkinkan untuk mengakibatkan keputusan yang salah. Untuk meminimisasi kesalahan, perlu dilakukan banyak backtesting sehingga model cukup terlatih untuk melakukan prediksi yang akurat.

Machine learning dan *big data* dapat dimanfaatkan dalam ilmu perminyakan dan geosains. Sebagai contoh, *Machine Learning* telah digunakan dalam interpretasi seismik (Waldeland dkk., 2018) untuk *fault detection* pada data seismik 3D. *Machine Learning* juga dapat diterapkan pada *reflection seismology analysis* untuk membuat citra lapisan sedimen and deteksi *water bottom*. *Machine Learning* juga digunakan di tahap produksi oil dan gas dimana *Machine Learning* digunakan untuk meningkatkan produksi minyak dengan mengoptimasi *flow simulation* pada *production forecasting*. Contoh lain yang berhubungan dengan studi ini adalah penentuan facies dengan menggunakan algoritma *Machine Learning* dengan input data *well log* (Hall, 2016). Pada prinsipnya penentuan litologi batuan shale dapat ditentukan dengan melihat nilai *Gamma Ray (GR)* yang relatif lebih tinggi sedangkan nilai GR yang rendah dapat mencerminkan beberapa kemungkinan litologi seperti *sandstone, coal, carbonate, salt* dan lainnya. Untuk memperbaiki hasil interpretasi, diperlukan data log lain seperti densitas, porositas dan *sonic log*. Penambahan data *wellbore* seperti *cutting* dan *core* juga dapat menjadi nilai tambahan untuk mengkonfirmasi hasil akhir interpretasi litologi. Studi ini berusaha membantu mengestimasi facies dengan menggunakan data-data sumur tersebut sebagai input. Pratama (2019) mengaplikasikan algoritma optimisasi *K-Nearest Neighbors (KNN)* untuk estimasi facies dengan hasil yang cukup baik di mana pengklasifikasian terlihat lebih efektif untuk facies yang tebal (Pratama, 2019), namun kurang baik untuk lapisan tipis. Metode KNN sendiri merupakan suatu model yang dapat mengklasifikasikan data berdasarkan jarak terdekat (Coomans & Massart, 1982). Dalam studi ini, kita memfokuskan kepada algoritma *Artificial Neural Network (ANN)*. Secara fundamental, method ANN terdiri dari multilayer perceptron yang membentuk suatu jaringan saraf tiruan. Jaringan saraf tiruan ini terdiri dari sistem neuron sederhana saling berhubungan, atau node, (Gambar.1), neuron ini saling terhubung membentuk suatu hubungan nonlinear dari vektor input ke vektor output.

Tujuan dari penelitian ini adalah untuk membandingkan algoritma ANN dengan algoritma KNN untuk estimasi facies yang lebih baik. Facies sendiri berfungsi untuk mengestimasi litologi secara tepat, sehingga memungkinkan untuk dapat digunakan dalam *real time drilling*.



Gambar 1. Multilayer perceptron dengan satu hidden layer.



Gambar 2. Lokasi Panoma Field terletak di Cekungan Anadarko, sebelah barat daya Kansas (Bohling & Dubois, 2003)

2 DATA DAN METODE

Data dalam studi ini menggunakan data dari lapangan Panama Council Grove dimana sebagian besar merupakan reservoir gas karbonat yang mencakup 2.700 mil persegi di Kansas Barat Daya, USA.

Kami menggunakan data dari delapan sumur. Enam sumur sebagai data training atau data pelatihan, dan dua sumur digunakan sebagai data testing atau data pengujian. Setiap sumur memiliki 7 parameter well log yang dijadikan sebagai input dan 1 label (kelas) facies batuan sebagai output yang dijadikan sebagai dasar dalam mengevaluasi model algoritma. Pengukuran parameter well log ini diambil secara vertikal dengan interval 0.15 m. Parameter well log terdiri dari 5 log yang diambil dari pengukuran wireline logging

Tabel 1. Data parameter well log yang dijadikan sebagai input. Indeks 1 - 5 data hasil pengukuran wireline logging, sedangkan indeks 6 dan 7 berdasarkan geological knowledge. Data diambil secara vertikal dengan interval 0.15 m (Bohling & Dubois, 2003).

No	Parameter	Deskripsi
1	GR	Log Gamma Ray
2	ILD_Log10	Log Resistivity
3	PHIND	Log Neutron Density Average
4	Delta PHI	Log Neutron Density Differences
5	PE	Log Photoelectric Effect
6	NM.M	Log Indikator Lingkungan Pengendapan
7	RELPOS	Log Relative Position (Berkaitan dengan lingkungan pengendapan)

Tabel 2. Kelas fasies yang dijadikan sebagai output. Kelas fasies dilabelkan dengan nomor dan warna tertentu, ke-9 fasies ini saling berdekatan dan bahkan tercampur satu sama lain pada kondisi sebenarnya di alam. Kolom ketiga menunjukkan kedekatan fasies sebagai tetangga terdekat pada setiap fasies (Hall, 2016).

Label	Fasies	Fasies Berdekatan
1	Sandstone	2
2	Coarse Siltstone	1,3
3	Fine Siltstone	2
4	Marine Silt Shale	5
5	Mud Stone	4,6
6	Wack Stone	5,7
7	Dolomite	6,8
8	Packstone	6,7,9
9	Bafflestone	7,8

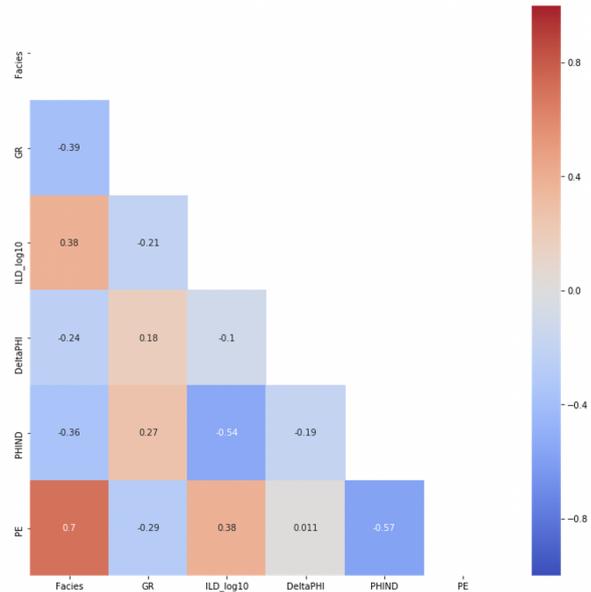
dan dua log dari geological knowledge yang terdiri dari log indikator lingkungan pengendapan dan log relative position (Tabel 1).

Dalam studi ini kita menggunakan label fasies sebagai output yang terdiri dari 9 kelas fasies yang berdekatan posisinya satu sama lain saat berada di alam (Tabel 2). Hal inilah yang membuat algoritma membutuhkan perhitungan yang rumit untuk menghasilkan model yang tepat, dikarenakan fasies yang berdekatan ini cenderung bergabung menjadi satu.

2.1 Korelasi

Karena algoritma Machine Learning yang melihat hubungan input dan output, penting untuk kita mengetahui bagaimana korelasi parameter well log yang dijadikan input dengan label fasies yang dijadikan sebagai output.

Pada Gambar 3. menunjukkan bahwa semakin merah atau semakin biru warna yang tunjukkan maka korelasi akan semakin tinggi. Nilai positif dan nilai negatif disini, menunjukkan adalah hubungan dari kedua parameter, jika kedua parameter memiliki hubungan yang sejajar maka nilainya positif, jika memiliki hubungan yang bertolak belakang, maka nilainya negatif. Dalam dataset yang kita miliki, parameter well log Photoelectric Effect dan Resistivity menunjukkan korelasi yang cukup besar dengan fasies, yaitu masing-masing senilai 0.7 dan 0.38. Ini artinya, parameter Photoelectric Effect dan Resistivity memegang

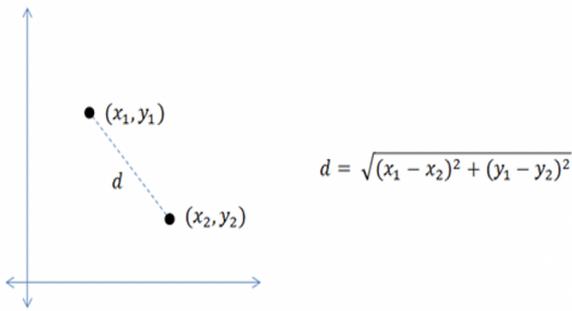


Gambar 3. Hasil perhitungan korelasi antar well log parameter dengan fasies. Hasil korelasi dihitung dengan metode Pearson (Freedman dkk., 2007).

peranan yang penting dalam memprediksi fasies. Sebaliknya, log DeltaPHI yang nilai korelasinya cukup rendah, kurang berpengaruh terhadap penentuan fasies. Selain parameter well log wireline logging, parameter well log yang berasal dari geological knowledge juga memainkan peran penting dalam menentukan fasies. Parameter well log Nonmarine.Marine Indicator akan memisahkan kelas fasies berdasarkan lingkungan pengendapannya. Variabel pembatas geologi ini sangat membantu dalam mengestimasi kelas fasies. Sebagai pengembangan studi ini di masa mendatang dapat mencoba untuk memainkan weighting atau pembobotan pada data data yang memiliki korelasi tinggi antara input dan output data.

2.2 K-Nearest Neighbors

Algoritma K-Nearest Neighbor adalah algoritma supervised learning dimana hasil dari perhitungan pengelompokan diklasifikasikan berdasarkan nilai mayoritas dari kategori k-tetangga terdekat. Tujuan dari algoritma ini adalah untuk mengklasifikasikan obyek baru berdasarkan atribut dan sample dari training data. Untuk mengetahui tentang detail algoritma KNN dapat melihat studi (Zhang dkk., 2018) Dalam pembuatan model, penentuan nilai k sangat penting dimana k merupakan jumlah data tetangga terdekat yang menentukan klasifikasi suatu data tersebut. Lalu setelah menentukan nilai k, dilakukan perhitungan jarak Euclidean yang berfungsi untuk menghitung jarak antar data testing dengan data hasil latihan (training). Secara prinsip setiap poin data yang baru akan dimasukkan ke dalam kategori dengan nilai k yang memiliki tetangga terbanyak berdasarkan nilai jarak euclidean yang optimal.



Gambar 4. Pengukuran jarak dengan Euclidean Distance.

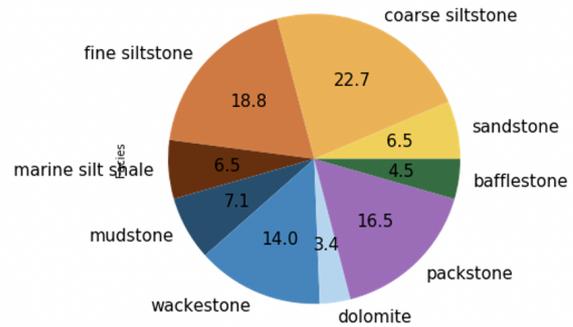
2.3 Artificial Neural Network

Sedangkan Artificial Neural Network (ANN) atau Jaringan Syaraf Tiruan merupakan sebuah teknik atau pendekatan pengolahan informasi yang terinspirasi oleh cara kerja sistem saraf biologis, khususnya pada sel otak manusia dalam memproses informasi. Elemen kunci dari teknik ini adalah struktur sistem pengolahan informasi yang bersifat unik dan beragam untuk tiap aplikasi. Neural Network terdiri dari sejumlah besar elemen pemrosesan informasi (neuron) yang saling terhubung dan bekerja bersama-sama untuk menyelesaikan sebuah masalah tertentu, yang umumnya masalah klasifikasi ataupun prediksi. Cara kerja Artificial Neural Network dapat dianalogikan sebagaimana halnya manusia belajar dengan menggunakan contoh atau yang disebut sebagai supervised learning. Sebuah Artificial Neural Network dapat dikonfigurasi untuk aplikasi tertentu, seperti pengenalan pola atau klasifikasi data, dan kemudian disempurnakan melalui proses pembelajaran. Proses belajar yang terjadi dalam sistem biologis melibatkan penyesuaian koneksi sinaptik yang ada antara neuron. Pada sistem Artificial Neural Network, penyesuaian koneksi sinaptik antar neuron dilakukan dengan menyesuaikan nilai bobot yang ada pada tiap konektivitas baik dari input, neuron maupun output. Neuron menerima masukan dari input layer dimana tiap input dikalikan sebuah bobot yang berbeda - beda (weight) ditambah oleh bias Persamaan 1. Untuk penjelasan yang lebih detail tentang algoritma metode ini dapat dilihat pada studi (Müller & Guido, 2016).

$$Y = \Sigma(weight \times input) + bias \tag{1}$$

2.4 Evaluasi

Berdasarkan pengamatan pada berbagai studi, metode KNN memiliki kelebihan untuk merespon perubahan dengan cepat pada data input. Sedangkan metode ANN, memiliki kemampuannya mempelajari data yang kompleks. Namun kedua metode juga mempunyai beberapa kekurangan. Metode KNN seringkali mengalami kesulitan untuk menarik kesimpulan yang akurat dan tidak cukup kuat ketika berhadapan dengan data yang penuh dengan noise. Sedangkan metode ANN sangat bergantung terhadap kemampuan prosesor, karena memerlukan logika yang lama dalam pembuatan struktur ANN yang berdasarkan trial and error. Studi ini berusaha membandingkan kedua al-



Gambar 5. Distribusi kelas fasies pada dataset.

goritma dalam menghasilkan model penentuan fasies yang mana yang lebih optimal.

Untuk itu, diperlukan sebuah metode evaluasi untuk mengukur keakuratan model dari kedua data ini. Pada studi ini, kami memakai confusion matrix. Confusion Matrix merupakan sebuah tabel yang terdiri atas banyaknya baris data testing yang diprediksi benar dan tidak benar oleh model. Metode ini sering digunakan dengan kasus multiple classifiers atau kelas yang lebih dari dua (Freitas dkk., 2007). Maka dari itu metode ini cocok digunakan dalam penelitian ini untuk mengukur kelas litologi yang lebih dari dua. Perhitungan akurasi dengan menggunakan tabel confusion matrix adalah sebagaimana dirumuskan oleh Persamaan (2).

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

Dengan penjelasan, TP adalah True Positive yang merupakan data positif yang diprediksi benar. TN adalah True Negative yang merupakan data negative yang diprediksi benar. FP adalah False Positive, merupakan data positif yang diprediksi salah. FN adalah False Negative, merupakan data negatif yang diprediksi salah

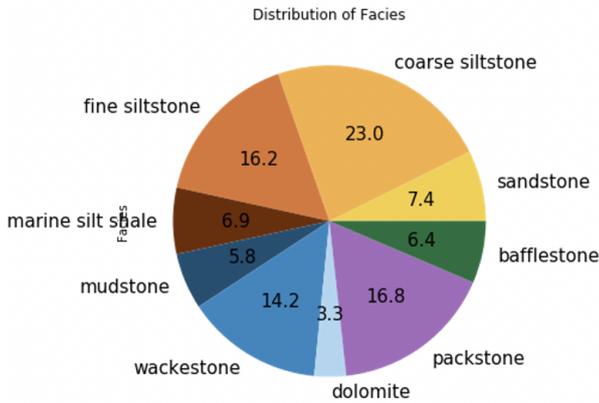
3 HASIL DAN DISKUSI

Sebelum menerapkan algoritma Machine Learning dalam dataset yang kita miliki, terdapat beberapa tahap untuk mengkondisikan dataset yang kita miliki. Hal pertama melihat persebaran output yang kita miliki dalam hal ini kelas fasies dan melakukan data cleaning. Seperti yang dapat dilihat pada Gambar 4. persebaran fasies dalam dataset ini didominasi oleh fasies coarse siltstone sebanyak 22.7% diikuti oleh fine siltstone dan packstone sebanyak 18.8% dan 16.5%. Kelas fasies dengan jumlah data yang paling sedikit disini adalah bafflestone dan dolomite dengan jumlah data 4.5% dan 3.4%.

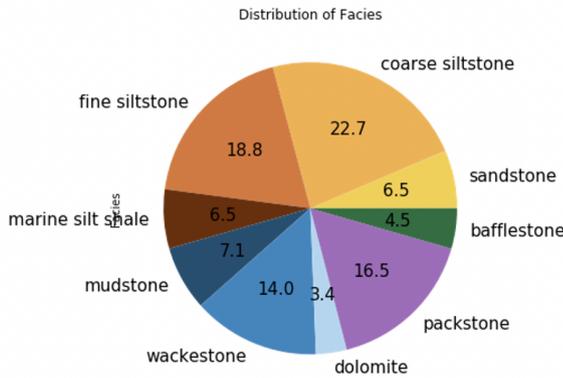
Langkah selanjutnya adalah melakukan normalisasi pada data, hal ini penting dikarenakan besar kecilnya nilai sebuah data mempengaruhi hasil prediksi, untuk itu perlu dilakukan normalisasi agar tidak terdapat kekeliruan dalam memprediksi data. (Alasadi & Bhaya, 2017). Kemudian membagi dataset tersebut menjadi 2 kelompok, enam sumur yang berjumlah 18.496 data akan digunakan sebagai data training, dan dua sumur dengan jumlah data sebe-

Tabel 3. 18 Skema aplikasi ANN dengan pemilihan parameter yang berbeda-beda.

Schema	Parameter			Accuracy		Time (Sec)
	Hidden Layer Size	Max Iteration	Activation Function	Train	Test	
1	(50, 50, 50)	10	Tanh	0.595	0.535	0.27
2	(50, 50, 50)	100	Tanh	0.746	0.535	2.47
3	(50, 50, 50)	1000	Tanh	1	0.468	20.1
4	(50, 100, 50)	10	Tanh	0.605	0.535	0.34
5	(50, 100, 50)	100	Tanh	0.78	0.558	2.88
6	(50, 100, 50)	1000	Tanh	1	0.486	23.7
7	(100, 100, 50)	10	Tanh	0.607	0.531	0.4
8	(100, 100, 50)	100	Tanh	0.808	0.534	3.3
9	(100, 100, 50)	1000	Tanh	1	0.451	20.1
10	(50, 50, 50)	10	ReLU	0.592	0.491	0.3
11	(50, 50, 50)	100	ReLU	0.778	0.505	2.52
12	(50, 50, 50)	1000	ReLU	0.93	0.503	10.5
13	(50, 100, 50)	10	ReLU	0.602	0.536	0.37
14	(50, 100, 50)	100	ReLU	0.803	0.526	3.38
15	(50, 100, 50)	1000	ReLU	0.952	0.507	11.9
16	(100, 100, 50)	10	ReLU	0.622	0.55	0.45
17	(100, 100, 50)	100	ReLU	0.833	0.533	3.84
18	(100, 100, 50)	1000	ReLU	0.983	0.435	17.3



Gambar 6. Distribusi kelas fasies pada data train.



Gambar 7. Distribusi kelas fasies pada data test.

sar 7.360 digunakan sebagai data testing. Hal ini diperlukan untuk mengevaluasi model dalam memprediksi data. Setelah itu, barulah data tersebut diaplikasikan kedalam metode *Machine Learning*.

3.1 Hasil Artificial Neural Network

Pada algoritma ANN diaplikasikan beberapa skema. Dalam setiap skema, terdapat perbedaan dalam pemilihan parameter. Parameter ini dibutuhkan untuk membuat model yang dapat memprediksi fasies menjadi lebih akurat. Parameter yang digunakan sebagai berikut:

Hidden layer sizes

Parameter ini digunakan untuk menentukan jumlah hidden layer yang terdapat dalam arsitektur algoritma ANN. Semakin sedikit jumlah hidden layer dapat menyebabkan algoritma underestimate dalam mengestimasi fasies, sedangkan jika jumlahnya terlalu banyak algoritma akan overestimate dalam mengestimasi fasies. Untuk itu, kita harus bisa menentukan berapa jumlah hidden layer yang sesuai untuk diterapkan pada dataset ini.

Maximum Iteration

Iteration atau pengulang dibutuhkan agar model dapat terlatih dengan baik sehingga dapat memprediksi output dengan tepat. Jika jumlah iterasi terlalu sedikit, maka error atau kesalahan yang dihasilkan cenderung lebih besar, namun jika pemilihan iterasi terlalu besar, maka algoritma tidak dapat memprediksi hasil dengan error yang lebih kecil lagi dan cenderung stagnan pada sebuah nilai error. Untuk itu, kita harus menemukan jumlah iterasi yang baik dan efektif untuk menghasilkan nilai error yang kecil.

Activation Function

Activation Function merupakan sebuah parameter yang menentukan apakah neuron dalam sebuah jaringan ANN harus aktif atau tidak. Hal ini bergantung pada bobot dari setiap input yang dimiliki. Metode Activation Function yang pada umumnya dipakai adalah ReLU (Rectified Linear Unit) dan Tanh.

Seperti yang ditampilkan pada Tabel 3. Skema 9 dengan parameter Hidden Size Layers sebanyak (50, 100, 50), iterasi sebanyak 10 kali dan 'Tanh' sebagai Activation Function memberikan model dengan akurasi proses train 0.78 dan akurasi proses test 0.558 sehingga mendapatkan hasil akurasi keseluruhan yang paling akurat yaitu 0.55. Perhitungan algoritma ANN untuk menghasilkan model dengan akurasi tersebut memakan waktu sekita 2.88 detik.

3.2 Hasil K-Nearest Neighbors

Algoritma KNN dioptimasi dengan mencari nilai k atau tetangga terdekat yang pas. Apabila nilai k tersebut optimal, maka algoritma KNN akan menghasilkan error atau kesalahan yang cenderung kecil.

Terlihat pada Tabel 4. Dengan k bernilai 17 menghasilkan model KNN yang paling akurat dengan akurasi train 0.699 dan akurasi test 0.533 sehingga menghasilkan akurasi keseluruhan sebesar 0.5, walaupun model ini memberikan nilai akurasi dibawah model ANN, waktu yang dibutuhkan KNN relatif lebih cepat dibanding metode ANN yaitu 0.3 detik.

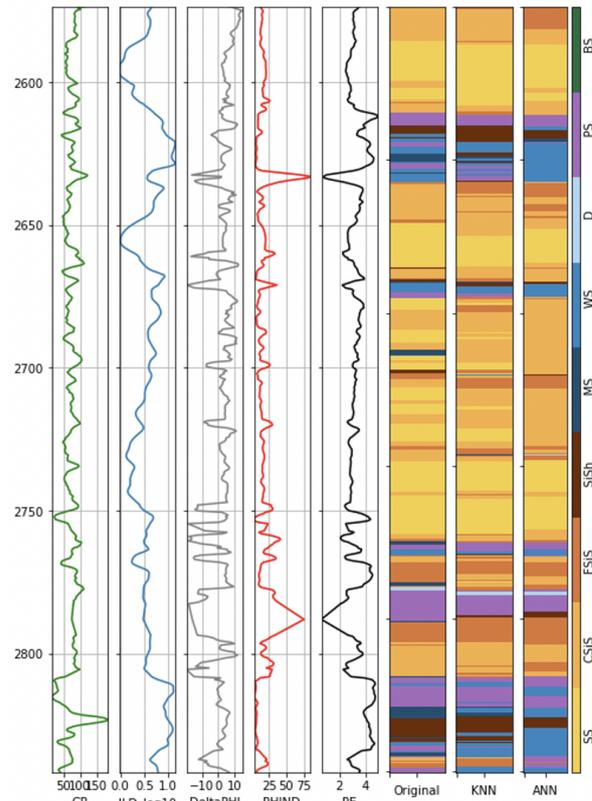
Tabel 4. Jumlah K dan nilai akurasi.

K	Accuracy		Time (Sec)
	Train	Test	
1	0.997	0.428	0.15
2	0.87	0.402	0.16
3	0.872	0.439	0.19
4	0.844	0.44	0.18
5	0.827	0.453	0.2
6	0.796	0.434	0.25
7	0.785	0.477	0.2
8	0.744	0.462	0.22
9	0.756	0.466	0.21
10	0.745	0.46	0.4
11	0.738	0.479	0.4
12	0.73	0.475	0.4
13	0.722	0.48	0.4
14	0.718	0.474	0.4
15	0.712	0.498	0.4
16	0.705	0.492	0.4
17	0.699	0.503	0.3
18	0.692	0.475	0.3

3.3 Perbandingan Hasil Prediksi Fasies

Kedua metode ini diterapkan pada salah satu dari data train dan salah satu sumur dari data test. terdapat 8 kolom pada gambar tersebut. Gambar 6 merupakan visualisasi pengaplikasian kedua metode ini pada Sumur Cross H Cattle, salah satu sumur yang digunakan sebagai data train. Kolom 1 – 5 dari kiri merupakan parameter well log yang dijadikan sebagai input, kolom 6 merupakan kelas fasies yang dijadikan sebagai output atau hasil sebenarnya dari dataset yang diukur berdasarkan hasil coring, kolom 7 merupakan hasil KNN dalam memprediksi kelas fasies, dan kolom terakhir merupakan hasil prediksi ANN dalam menentukan kelas fasies. Gambar 7 merupakan hasil estimasi fasies pada Sumur Shankle, salah satu sumur dari data test. Sama seperti gambar 5, Kolom 1 – 5 pada gambar ini merupakan parameter well log yang dijadikan sebagai input, kolom 6 merupakan kelas fasies hasil pengukuran data coring sebenarnya. Kolom 7 merupakan kelas fasies hasil estimasi metode KNN, dan kolom 8 estimasi kelas fasies oleh model ANN. Terlihat hasil estimasi fasies pada sumur data train lebih baik dibandingkan dengan sumur dari data test. Hasil ini dikarenakan, model mempelajari pola yang terdapat pada data train, kemudian menerapkan hasil pembelajaran itu pada data test. Selain itu, hasil pada data train maupun data test, menunjukkan bahwa metode KNN melakukan terlalu banyak

Well: CROSS H CATTLE

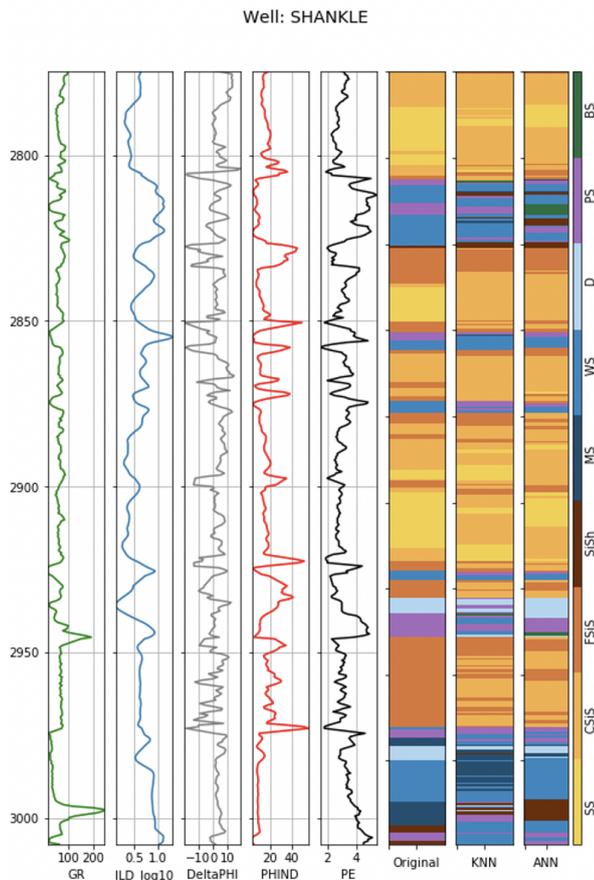


Gambar 8. Hasil prediksi metode KNN dan ANN pada salah satu sumur dari data train. Dari kiri, 5 kolom pertama merupakan parameter well log yang dijadikan sebagai input, kolom ke-6 merupakan kelas fasies sebenarnya hasil coring, kolom ke-7 merupakan kelas fasies hasil prediksi metode KNN, dan kolom terakhir merupakan kelas fasies hasil prediksi metode ANN.

estimasi yang berlebihan dibanding ANN (over-estimate). Metode yang overestimate dapat dilihat dari perbedaan tingkat akurasi yang jauh antara akurasi train dan akurasi test, dan rata-rata skema dari model KNN menghasilkan hal tersebut. Overestimate ini sendiri akan menyebabkan model KNN menampilkan banyak lapisan lapisan baru yang tidak sesuai pada interpretasi fasies yang sebenarnya. Selain disebabkan oleh overestimate, hal ini kemungkinan juga disebabkan oleh adanya efek dari fasies yang berdekatan satu sama lain dan kurang optimalnya dalam memilih parameter. Untuk itu, perlu dilakukan pengembangan selanjutnya, disarankan adanya pengurangan atau pergantian input data dan pembebanan pada input layer awal. Model geologi regional sangat dibutuhkan untuk pengembangan lebih lanjut ketika sudah ada data baru yang didapatkan.

4 KESIMPULAN

Dalam jurnal ini, kita telah berhasil mengaplikasikan algoritma ANN dan membandingkannya dengan algoritma KNN. Dengan menggunakan parameter yang sekarang, menunjukkan algoritma ANN masih memiliki hasil yang lebih



Gambar 9. Hasil prediksi metode KNN dan ANN pada salah satu sumur dari data test. Dari kiri, 5 kolom pertama merupakan parameter well log yang dijadikan sebagai input, kolom ke-6 merupakan kelas fasies sebenarnya hasil coring, kolom ke-7 merupakan kelas fasies hasil prediksi metode KNN, dan kolom terakhir merupakan kelas fasies hasil prediksi metode ANN.

baik dibandingkan algoritma KNN. Namun dari segi kecepatan, algoritma KNN menghasilkan model estimasi yang relative lebih cepat dibandingkan algoritma ANN. Untuk mendapatkan hasil yang lebih optimal dalam mengestimasi fasies, dibutuhkan beberapa hal seperti model geologi regional dan terus memperbarui dataset dengan data baru yang dimiliki.

TERIMA KASIH

Penulis mengucapkan terima kasih kepada Kansas Geological Survey, University of Kansas yang telah menyediakan dataset yang digunakan dalam penelitian ini. Dan Brendon Hall yang telah menyediakan tutorial mengenai facies classification di github.

DAFTAR PUSTAKA

Alasadi, S. A., & Bhaya, W. S. (2017). Review of data preprocessing techniques in data mining. *Journal of Engineering and Applied Sciences*. <https://doi.org/10.3923/jeasci.2017.4102.4107>

- Bohling, G. C., & Dubois, M. K. (2003). An integrated application of neural network and Markov chain techniques to the prediction of lithofacies from well logs: Kansas Geological Survey Open-File Report 2003-50, 6 p. In Group..
- Coomans, D., & Massart, D. L. (1982). Alternative k-nearest neighbour rules in supervised pattern recognition. Part 1. k-Nearest neighbour classification by using alternative voting rules. *Analytica Chimica Acta*. [https://doi.org/10.1016/S0003-2670\(01\)95359-0](https://doi.org/10.1016/S0003-2670(01)95359-0)
- Freedman, D., Pisani, R., & Purves, R. (2007). *Statistics Fourth Edition*. In W.W. Norton & Company.
- Freitas, C. O. A., De Carvalho, J. M., Oliveira, J. J., Aires, S. B. K., & Sabourin, R. (2007). Confusion matrix disagreement for multiple classifiers. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-540-76725-1_41
- Hall, B. (2016). Facies classification using machine learning. *Leading Edge*. <https://doi.org/10.1190/tle35100906.1>
- Koza, J. R., Bennett, F. H., Andre, D., & Keane, M. A. (1996). Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming. In *Artificial Intelligence in Design '96*. https://doi.org/10.1007/978-94-009-0279-4_9
- Müller, A. C., & Guido, S. (2016). Introduction to Machine Learning with Python: a guide for data scientists. In *Journal of Chemical Information and Modeling*. <https://doi.org/10.1017/CBO9781107415324.004>
- Pratama, H. (2019). Machine Learning: Using Optimized KNN (K-Nearest Neighbors) to Predict the Facies Classifications. <https://doi.org/10.1190/segj2018-139.1>
- Waldeland, A. U., Jensen, A. C., Gelius, L.-J., & Solberg, A. H. S. (2018). Convolutional neural networks for automated seismic interpretation. *The Leading Edge*. <https://doi.org/10.1190/tle37070529.1>
- algorithm with data-driven k parameter computation. *Pattern Recognition Letters*. <https://doi.org/10.1016/j.patrec.2017.09.036>