



## Analisa Model *Machine Learning* dalam Memprediksi Laju Produksi Sumur Migas 15/9-F-14H

Devy Ayu Rhamadhani<sup>1)</sup> Eriska Eklezia Dwi Saputri<sup>1\*)</sup>

<sup>1)</sup> Teknik Perminyakan, Universitas Jember

\* corresponding email: [eriska.eklezia@unej.ac.id](mailto:eriska.eklezia@unej.ac.id)

### ABSTRACT

*AI algorithm learns various data streams from various sources sensors and engines to extract the analytics resulting in sound advice smart based on business needs. This deep insight makes it possible for oil and gas companies to have better visibility of the whole process and operations, thereby enabling them to make strategic decisions better. This of course leads to increased operating efficiency, cost reduction, and even reduce the risk of failure. Application of artificial intelligence using machine learning to production of oil and gas wells needs to be done to get predictive results perfect. With the support of existing field data so obtained simulation results that provide an overview of the prediction of production wells can optimizing the implementation of production performance for wells that have same production history. The simulation is carried out using the development of machine learning models, Support Vector Regression (SVR), Elastic Net, dan Linear Regression. The data which contains informations about the well production will be divided into two parts, 70% for training and 30% for testing. Of the three models will be seen which one is the best in predicting the production rate of the well 15/9-F-14H based on the RMSE and  $R^2$  score. SVR is the best model for predicting oil by producing RMSE 5.48 and  $R^2$  0.88 when testing. Elastic-Net is the best model for predicting gas by producing RMSE 966.82 and  $R^2$  0.85 when testing. There is no model that fits to predict the water production.*

**Keywords:** artificial intelligence; machine learning; prediction; SVR; Elastic-Net; Linear Regression

### I. PENDAHULUAN

Penerapan (AI) *Artificial Intelligence* dilakukan dengan pemodelan *machine learning* menggunakan beberapa pendekatan untuk mendapatkan informasi yang akurat. Prediksi kinerja produksi reservoir migas adalah salah satu komponen penting dalam manajemen produksi migas. Di sisi lain, literatur saat ini, berhubungan dengan prediksi kinerja produksi reservoir migas menggunakan AI masih jarang karena merupakan bidang penelitian baru. Seringkali para *engineer* menemukan ketidaklinieran antara produksi sumur dengan parameter fisik minyak yang relevan sehingga menambah kompleksitas tersendiri dalam melakukan pemodelan prediksi produksi reservoir (Ma dan Liu, 2018). Oleh karena itu, mengembangkan model prediksi produksi yang tepat berdasarkan data yang tersedia membantu para *engineer* dalam memiliki pemahaman mendalam tentang kinerja reservoir untuk menyelesaikan masalah terkait manajemen reservoir.

#### 1.1. *Support Vector Regression (SVR)*

Algoritme SVR adalah teori yang diadaptasi dari teori *machine learning* yang sudah digunakan untuk memecahkan masalah klasifikasi, yaitu *Support Vector Machine (SVM)*. SVR menghasilkan keluaran berupa bilangan riil. Konsep algoritme SVR dapat menghasilkan nilai peramalan yang bagus karena SVR mempunyai kemampuan menyelesaikan masalah *overfitting* [2]. *Overfitting* adalah perilaku data saat fase pelatihan atau training menghasilkan akurasi prediksi hampir sempurna [15].

#### 1.2. *Linear Regression*

Regresi linier adalah metode statistika yang digunakan untuk membentuk model atau hubungan antara satu atau lebih variabel bebas X dengan sebuah variabel respon Y. Analisis regresi dengan satu variabel bebas X disebut sebagai regresi linier sederhana, sedangkan jika terdapat lebih dari satu variabel bebas X, disebut sebagai regresi linier berganda [6]. Dalam menganalisis hubungan fungsional antara variabel bebas X dan variabel respon Y, ada kemungkinan terjadi hubungan linier yang berbeda untuk setiap interval X. Apabila regresi X terhadap Y memiliki hubungan linier tertentu pada interval X tertentu, tetapi juga memiliki hubungan linier yang berbeda pada interval X yang lain, maka penggunaan model regresi linier sederhana kurang tepat pada kasus tersebut karena hasil analisis tidak dapat memberikan informasi menyeluruh tentang data [11].



### 1.2. *Elastic-Net*

*Elastic Net* adalah salah satu metode penyusutan seperti *Ridge* dan LASSO yang dapat mengatasi permasalahan multikolinearitas yang diperkenalkan oleh Zou dan Hastie pada tahun 2005. *Elastic-Net* merupakan regularisasi penalti gabungan antara regresi gulud dan LASSO. Regularisasi ini menangani beberapa kekurangan LASSO. *Elastic-Net* berfungsi seperti LASSO, yaitu secara simultan melakukan seleksi variabel dan penyusutan koefisien penduga, serta dapat menangani kekurangan dari LASSO [16].

### 1.3. *Root Mean Square Error (RMSE)*

*Root Mean Square Error (RMSE)* merupakan besarnya tingkat kesalahan hasil prediksi, dimana semakin kecil (mendekati 0) nilai RMSE maka hasil prediksi akan semakin akurat [8]. Nilai RMSE digunakan untuk membedakan kinerja model dalam periode kalibrasi dengan periode validasi serta digunakan untuk perbandingan kinerja antara model prediksi yang satu dengan model prediksi yang lain [4].

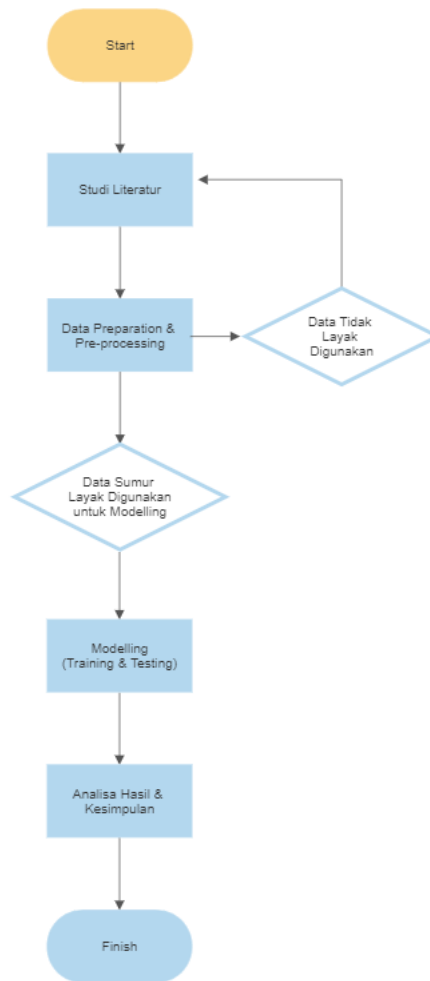
### 1.5. *R-square (R<sup>2</sup>)*

Koefisien determinasi atau  $R^2$  adalah suatu nilai yang menggambarkan seberapa besar perubahan atau variabel dari variabel terikat bisa dijelaskan oleh perubahan atau variasi dari variabel bebas. Dengan mengetahui nilai koefisien determinasi dapat menjelaskan kinerja dari model dalam memprediksi variabel tersebut. Nilai  $R^2$  dikatakan baik jika  $>0,5$  karena nilai  $R^2$  berkisar antara 0 sampai 1 dengan ketentuan semakin mendekati angka 1 artinya model semakin baik. Jika nilai  $R^2$  0,6 berarti 60% sebaran variabel dependen dapat dijelaskan oleh variabel independen. Sisanya 40% tidak dapat dijelaskan oleh variabel independen atau dapat dijelaskan oleh variabel di luar variabel independen (komponen error) [5].

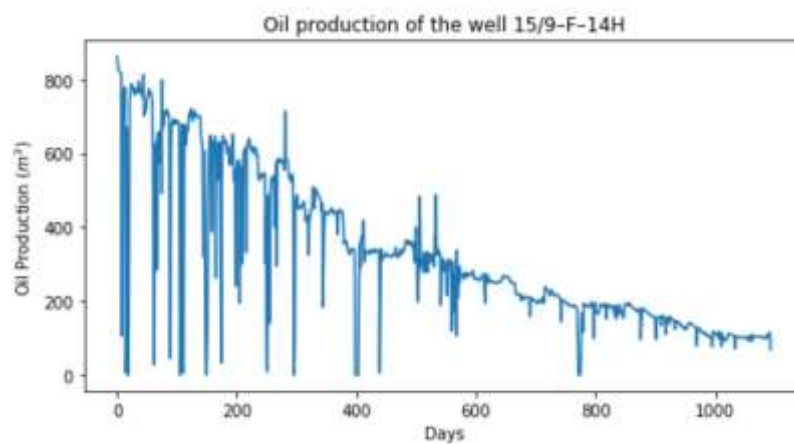
Penelitian ini bertujuan untuk mendapatkan model *machine learning* terbaik dari ketiga model yang digunakan, SVR, *Linear Regression*, *Elastic-Net* dalam melakukan prediksi produksi minyak, gas, dan air di sumur 15/9-F-14H di Lapangan Volve. Lapangan Volve berada di Blok 15/9 bagian selatan Laut Utara Norwegia. Lapangan Volve terletak di kedalaman antara 2750 m dan 3210 m di bawah permukaan laut. Karakteristik batuan pada lapangan ini memiliki nilai permeabilitas sekitar 1000 mD, porositas 0,21, dan rasio *net-to-gross* 0,93, serta saturasi air zona bantalan minyak rata-rata 0,2 [13].

## II. METODOLOGI PENELITIAN

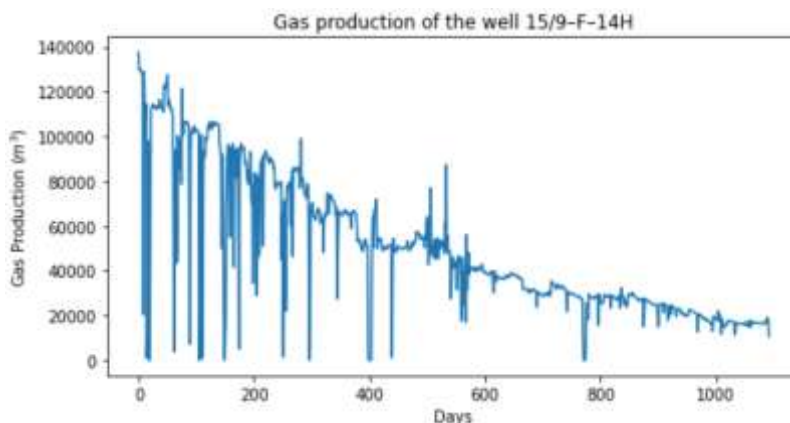
Pertama, dilakukan studi mengenai *machine learning* dan pengembangan model yang akan digunakan, antara lain, *Support Vector Regression (SVR)*, *Linear Regression*, dan *Elastic Net*. Dilanjutkan dengan mencari data lapangan yang dapat diolah. Selanjutnya data yang didapatkan akan dianalisis, apabila tidak memenuhi kriteria maka akan dilakukan pencarian kembali. Data yang akan diolah dibagi menjadi dua bagian, 70% untuk *training* dan 30% untuk *testing*. Sebelumnya dilakukan pemilihan fitur, yang mana adalah parameter atau jenis data dari sumur yang digunakan untuk *modelling* dikarenakan dianggap memiliki pengaruh terhadap *output*. Algoritma dibuat dengan menggunakan bahasa pemrograman Python. Data yang digunakan sebanyak 1093 hari produksi sumur 15/9-F-14H. Pada tahap *training*, ketiga model yang digunakan akan dibuat sebaik mungkin untuk data *training* sampai menghasilkan nilai RMSE dan  $R^2$  terbaik. Kemudian model yang sudah digunakan untuk *training* akan selanjutnya digunakan untuk data *testing*.



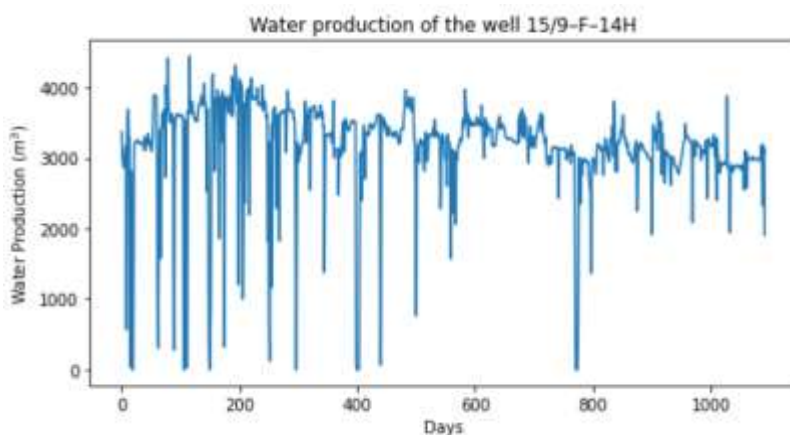
Gambar 1. Flow Chart Penelitian



Gambar 2. Produksi Minyak pada Sumur 15/9-F-14H selama 1093 Hari

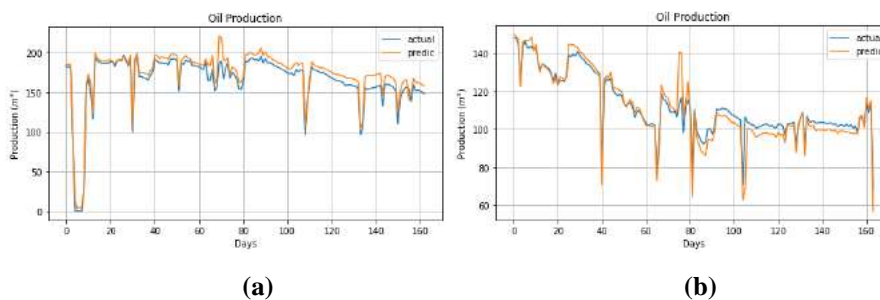


Gambar 3. Produksi Gas pada Sumur 15/9-F-14H selama 1093 Hari

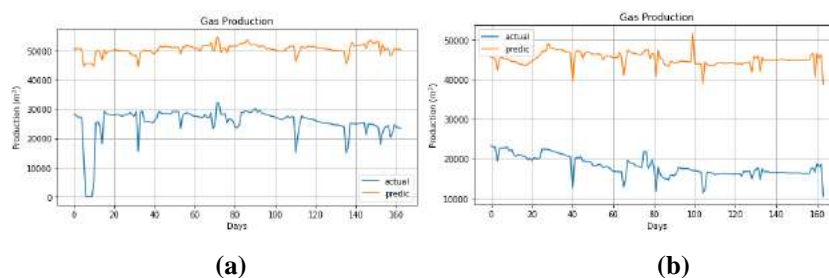


Gambar 4. Produksi Air pada Sumur 15/9-F-14H selama 1093 Hari

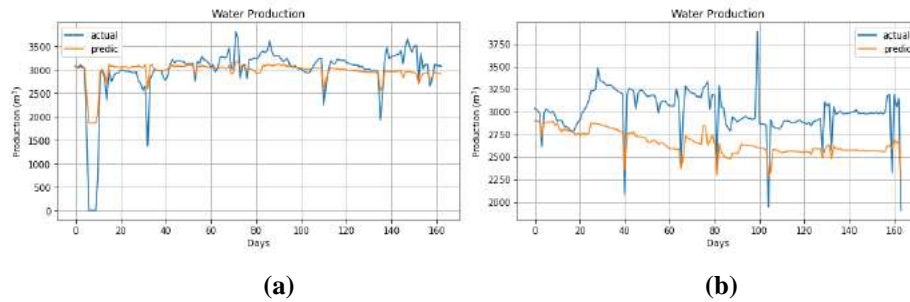
### III. HASIL DAN PEMBAHASAN



Gambar 5. *Training* (a) dan *Testing* (b) Prediksi Produksi Minyak menggunakan SVR



Gambar 6. *Training* (a) dan *Testing* (b) Prediksi Produksi Gas menggunakan SVR

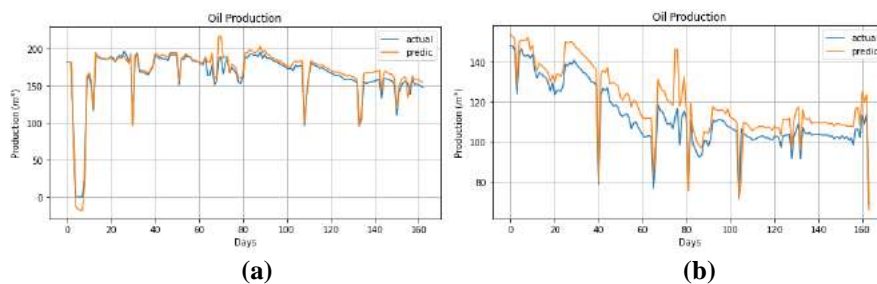


Gambar 7. *Training* (a) dan *Testing* (b) Prediksi Produksi Air menggunakan SVR

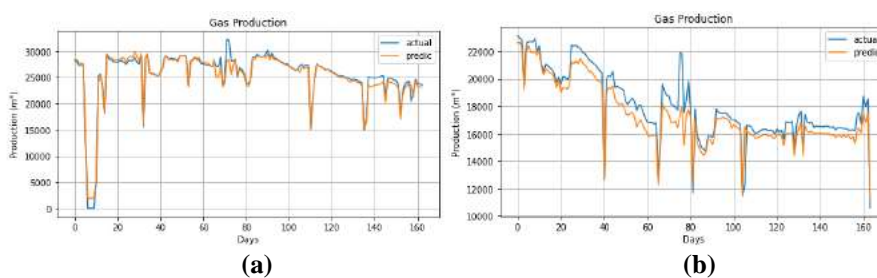
Tabel 1. Hasil *Training* dan *Testing* SVR

Fasa	Training		Testing	
	RMSE	R <sup>2</sup>	RMSE	R <sup>2</sup>
Minyak	9.1570009	0.9319693	5.48076631	0.8795249
Gas	25449.397	-216.764	27141.1856	-116.6198
Air	401.35061	-2.396694	401.305605	-1.821983
(1-R <sup>2</sup> )/3		73.742908		40.187411

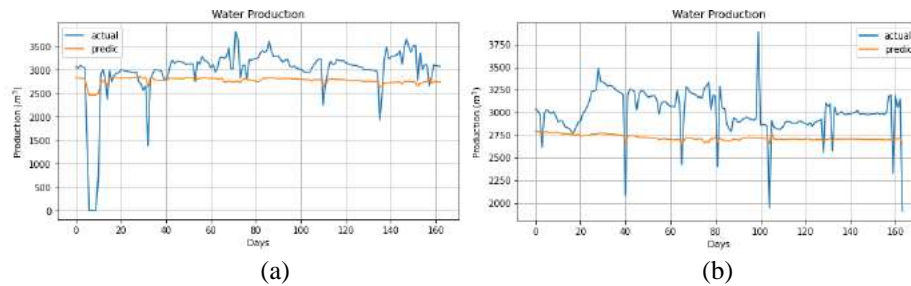
Pada model SVR, dengan nilai RMSE yang lebih rendah daripada ketika training, artinya penyimpangan data yang terjadi tidak begitu besar, namun nilai R<sup>2</sup> menurun artinya komponen error yang ada semakin banyak daripada ketika dilakukan *training* dengan maksud kemampuan model amat terbatas dalam menjelaskan variasi data yang digunakan. Namun dilihat dari grafik yang ada bahwa ketika training maupun testing untuk memprediksi produksi minyak bumi, model ini cukup stabil yang mana artinya model ini bagus untuk digunakan. Nilai RMSE yang begitu tinggi artinya banyak penyimpangan data yang terjadi lalu nilai R<sup>2</sup> yang bernilai negatif artinya model ini dengan parameter yang dipilih seharusnya tidak dapat digunakan untuk melakukan prediksi produksi gas. Model ini tidak dapat digunakan untuk memprediksi air karenanya besarnya penyimpangan data dan komponen error ditandai dengan adanya nilai R<sup>2</sup> yang negatif serta nilai RMSE yang tinggi.



Gambar 8. *Training* (a) dan *Testing* (b) Prediksi Produksi Minyak menggunakan *Elastic-Net*



Gambar 9. *Training* (a) dan *Testing* (b) Prediksi Produksi Gas menggunakan *Elastic-Net*

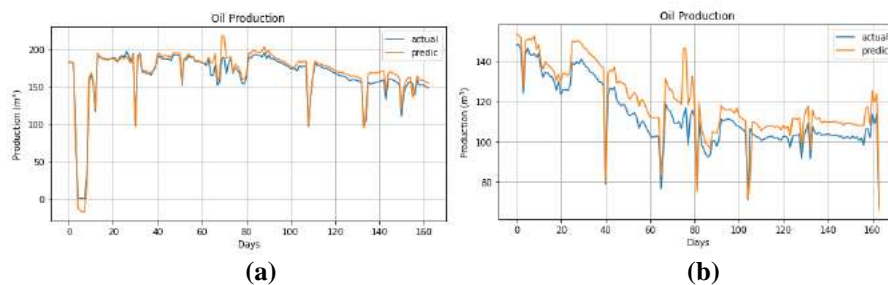


Gambar 10. Training (a) dan Testing (b) Prediksi Produksi Air menggunakan Elastic-Net

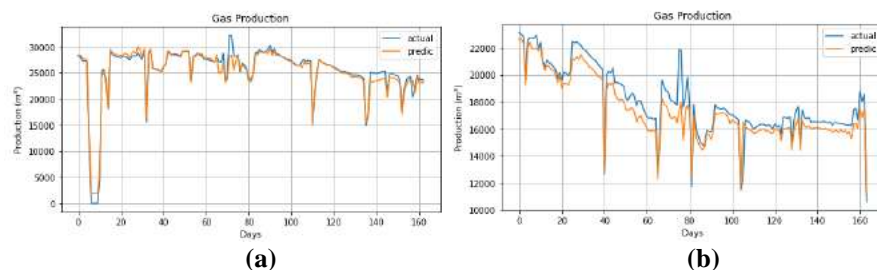
Tabel 2. Hasil *Training* dan *Testing Elastic-Net*

Fasa	Training		Testing	
	RMSE	R <sup>2</sup>	RMSE	R <sup>2</sup>
Minyak	7.4460087	0.9616638	8.77908844	0.6908896
Gas	864.65497	0.9616638	966.821278	0.8507497
Air	573.60905	-64.05381	363.5167	-1.315542
$(1-R^2)/3$		21.710161		0.9246344

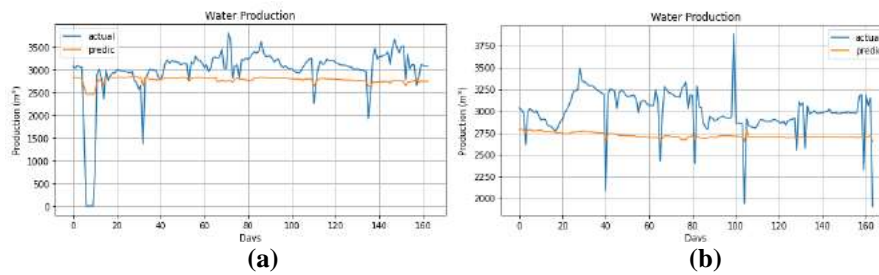
Model *Elastic-Net* dalam memprediksi produksi minyak mengalami penurunan performa dilihat dari nilai RMSE yang dihasilkan ketika testing menjadi lebih tinggi daripada ketika training dan juga nilai R<sup>2</sup> menjadi lebih rendah. Artinya penyimpangan data dan komponen error yang ada lebih banyak daripada yang dihasilkan ketika training. Dari Gambar 8, terlihat seperti model ini terlihat cukup bagus dalam memprediksi ditandai dengan rendahnya nilai R<sup>2</sup>, nilai RMSE terlihat sangat tinggi yang artinya banyak penyimpangan data yang terjadi namun dalam kasus ini RMSE bersifat *non-dimensionless* artinya data input akan memengaruhi besarnya koefisien determinasi RMSE, berbeda dengan R<sup>2</sup>. Model ini tidak dapat digunakan untuk memprediksi air sama halnya dengan model SVR.



Gambar 11. *Training* (a) dan *Testing* (b) Prediksi Produksi Minyak menggunakan *Linear Regression*



Gambar 12. *Training* (a) dan *Testing* (b) Prediksi Produksi Gas menggunakan *Linear Regression*



Gambar 13. *Training* (a) dan *Testing* (b) Prediksi Produksi Air menggunakan *Linear Regression*

Tabel 3. Hasil *Training* dan *Testing Linear Regression*

Fasa	Training		Testing	
	RMSE	R <sup>2</sup>	RMSE	R <sup>2</sup>
Minyak	7.4461801	0.9616627	8.77945996	0.6908635
Gas	864.88498	0.9703307	967.257981	0.8506149
Air	573.62647	-64.04386	363.559587	-1.316089
(1-R <sup>2</sup> )/3		21.703955		0.9248702

Model *Linear Regression* memiliki hasil yang hampir serupa dengan model *Elastic-Net* atau tidak jauh berbeda. Untuk fasa minyak, model yang paling dalam memprediksi produksi minyak adalah SVR karena menghasilkan *output* nilai RMSE saat testing lebih rendah atau yang paling rendah dari kedua model lainnya, sedangkan untuk nilai R<sup>2</sup>nya merupakan yang paling tinggi untuk fasa minyak. Model *Elastic-Net* adalah model yang paling bagus untuk memprediksi gas, namun tidak ada satu pun model yang bagus untuk memprediksi air karena tingginya nilai RMSE dan nilai R<sup>2</sup> yang bernilai negatif. Pada beberapa model terjadi *overfitting* dan *underfitting*, hal ini terjadi ketika model memiliki performa yang terlalu bagus ketika *training* ditandai dengan tingginya nilai RMSE yang dihasilkan dan rendahnya nilai R<sup>2</sup> yang terpaut jauh dengan yang dihasilkan ketika *training*, begitu pun sebaliknya yang terjadi ketika *underfitting*. *Overfitting* terjadi karena model tidak dapat menggeneralisasi dan justru terlalu ‘pas’ dengan data *training*.

Dari analisa yang dilakukan, nilai RMSE yang cukup tinggi dari gas dan air daripada pada minyak dikarenakan produksi gas dan air yang lebih banyak daripada minyak, sehingga hal tersebut yang memengaruhi nilai RMSE dapat mencapai angka ratusan hingga ribuan. Untuk nilai R<sup>2</sup> yang tinggi atau mendekati 1 maka artinya model tersebut bagus dalam memprediksi, contohnya pada gas. Nilai RMSE dari gas tergolong tinggi karena mencapai ratusan namun nilai R<sup>2</sup> dari gas ketika *training* maupun *testing* merupakan yang paling tinggi ketika menggunakan model *Elastic Net* dan *Linear Regression*. Artinya, model-model tersebut bekerja dengan baik dalam memprediksi gas pada sumur tersebut. Sedangkan pada air, nilai RMSE dan R<sup>2</sup> yang dihasilkan tidak sesuai yang diharapkan. Dari ketiga model yang diaplikasikan dalam memprediksi produksi air. Tidak ada satu pun yang menghasilkan *output* yang bagus. Hal ini terjadi karena parameter atau fitur yang digunakan tidak berkorelasi dengan tren produksi pada air itu sendiri. Tren produksi pada minyak dan gas memiliki kemiripan, yaitu menurun. Sedangkan tren produksi pada air dapat dilihat pada grafik ialah lebih acak daripada pada minyak dan gas, sehingga untuk parameter yang digunakan perlu yang berkorelasi seperti tekanan, dll. Pada kasus ini, produksi air sendiri tergantung pada pompa *artificial lift* yang digunakan sedangkan produksi minyak dan gas tergantung pada kondisi reservoir. Sehingga untuk produksi air sendiri lebih dapat dikontrol daripada minyak dan gas. Itu kenapa, untuk parameter atau fitur yang digunakan tidak bisa menyamakan dengan yang digunakan oleh minyak dan gas.

### III. KESIMPULAN

Model *machine learning* yang digunakan cukup akurat dan stabil dalam memprediksi beberapa fasa, seperti minyak. Ketika memprediksi minyak, ketiga model tersebut menunjukkan *output* yang paling baik, ditandai dengan tidaknya adanya nilai R<sup>2</sup> yang bernilai negatif atau hingga menyentuh puluhan hingga ratusan. Namun ketika diaplikasikan kepada fasa yang lain hasil yang didapatkan tidak akurat dan banyak terdapat penyimpangan data serta komponen error berdasarkan nilai RMSE dan R<sup>2</sup> terutama seperti pada air, nilai R<sup>2</sup> ketika memprediksi produksi air pada ketiga model tersebut bernilai negatif sehingga dapat dikatakan model tersebut tidak akurat dalam memprediksi dikarenakan tren produksi yang berbeda antara minyak, air, dan gas. Performa dari model *machine learning* bergantung pada jenis data, fitur serta parameter yang digunakan. Beberapa perlakuan berbeda dilakukan pada data yang diolah *machine learning* dapat membuat penurunan performa pada model.



## DAFTAR PUSTAKA

- [1] Fanny, dkk., 2018. Pendugaan Produktivitas Bagan Perahu dengan Regresi Gulud, LASSO dan Elastic-net, *Xplore: Journal of Statistics*, 2(2), pp. 7–14.
- [2] Furi, R. P., dkk., 2015. *Prediksi Financial Time Series Menggunakan Independent Component Analysis dan Support Vector Regression Studi Kasus : IHSG dan JII*
- [3] Jamaaluddin dan Indah, S., 2021. Buku Ajar Kecerdasan Buatan, p. 121.
- [4] Hosseini, M., et al., 2014. Determination of permeability index using Stoneley slowness analysis, NMR models, and formation evaluations: A case study from a gas reservoir, south of Iran. *Journal of Applied Geophysics*, 109, 80–87.
- [5] Kurniawan, A., 2019. Pengolahan Riset Ekonomi Jadi Mudah Dengan IBM SPSS. Surabaya: Jakad Publishing. h. 31.
- [6] Kurniawan, D., 2008. Regresi Linier. Australia: R Development Core Team.
- [7] Liu, C., and Ma, J., 2018. Development and validation of the Chinese social media addiction scale. *Personality and Individual Differences*, 134, 55–59.
- [8] Mahyudin, M., dkk., 2014. Model Prediksi Liku Kalibrasi Menggunakan Pendekatan Jaringan Saraf Tiruan (ZST) (Studi Kasus: Sub DAS Siak Hulu) (Doctoral dissertation, Riau University)
- [9] Maulana, N.D., dkk., 2019. Implementasi Metode *Support Vector Regression* (SVR) Dalam Peramalan Penjualan Roti (Studi Kasus : Harum Bakery), *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 3(3), pp. 2986–2995.
- [10] Putra, J. W. G., et al., 2020. Pengenalan Konsep Pembelajaran Mesin dan *Deep Learning*, *Computational Linguistics and Natural Language Processing Laboratory*, 4, pp. 1–235.
- [11] Ryan, S. E., and Porth, L. S., 2007. *United States Department of Agriculture A Tutorial on the Piecewise Regression Approach Applied to Bedload Transport Data*.
- [12] Roihan, A., dkk., 2020. Pemanfaatan *Machine Learning* dalam Berbagai Bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 5(1), 75–82.
- [13] Shang Wui, et al., 2021. Well production forecast in Volve field: Application of rigorous machine learning techniques and metaheuristic algorithm. *Journal of Petroleum Science and Engineering*, 208, 109468.
- [14] Sobron, M., dan Lubis, 2021. Implementasi Artificial Intelligence Pada System Manufaktur Terpadu. Seminar Nasional Teknik (SEMNASTEK) UISU, 4(1), 1–7.
- [15] Yasin, H., dkk., 2014. PREDIKSI HARGA SAHAM MENGGUNAKAN SUPPORT VECTOR REGRESSION DENGAN ALGORITMA GRID SEARCH.
- [16] Zou, H., and Hastie, T., 2005. *Regularization and variable selection via the elastic net*. In *J. R. Statist. Soc. B* (Vol. 67, Issue 2).