

## **LAPORAN AKHIR PENELITIAN**



### **IMPLEMENTASI ALGORITMA MACHINE LEARNING DALAM MEMPREDIKSI ACTIVE POWER DI PERUSAHAAN PEMBANGKIT LISTRIK TENAGA BATUBARA**

<b>Ketua</b>	<b>: Ir. Mega Bagus Herlambang, ST, MT, PhD, IPM, ASEAN.Eng</b>	<b>0310038804</b>
<b>Anggota</b>	<b>: Dra. Ir. Gadih Ranti, MT, IPU, ASEAN.Eng Rosyad Burhan</b>	<b>0305056305 1132000066</b>

**PROGRAM STUDI TEKNIK INDUSTRI  
INSTITUT TEKNOLOGI INDONESIA  
Februari, 2024**

## HALAMAN PENGESAHAN

Judul Penelitian : Implementasi Algoritma Machine Learning Dalam Memprediksi Active Power di Perusahaan Pembangkit Listrik Tenaga Batubara

Jenis Penelitian : Terapan

Bidang Penelitian : Data Science

Tujuan Sosial Ekonomi : Peningkatan efisiensi dan efektifitas kinerja PLTU

Peneliti

a. Nama Lengkap : Ir. Mega Bagus Herlambang, ST, MT, PhD, IPM, ASEAN.Eng

b. NIDN : 0310038804

c. Jabatan Fungsional : Lektor

d. Program Studi : Teknik Industri

e. Nomor HP : 081390447744

f. Alamat Surel (*e-mail*) : [mega.herlambang@iti.ac.id](mailto:mega.herlambang@iti.ac.id)

Anggota Peneliti

a. Nama Lengkap : Dra. Ir. Gadih Ranti, MT, IPU, ASEAN.Eng

b. NIDN : 0305056305

c. Institusi : Program Studi Teknik Industri ITI

Anggota Mahasiswa

a. Nama Lengkap : Rosyad Burhan

b. NRP : 1132000066

Institusi Sumber Dana : Mandiri

Biaya Penelitian : Rp 15.000.000,-

Kerjasama Mitra : PT. PLN Indonesia Power

Kota Tangerang Selatan, 24 Februari 2024

Mengetahui,


Program Studi Teknik Industri  
Ketua  
  
(Ir. Mega Bagus Herlambang, ST, MT, PhD, IPM, ASEAN. Eng)  
NIDN : 0310038804

Ketua Tim

  
(Ir. Mega Bagus Herlambang, ST, MT, PhD, IPM, ASEAN. Eng)  
NIDN : 0310038804

Menyetujui,

Pusat Riset dan Pengabdian Masyarakat (PRPM) – ITI

Kepala  
  
(Prof. Dr. Ir. Ratnawati, M.Eng.Sc., IPM)  
NIDN : 0301036303



## **PRAKATA**

Puji Puji Syukur kehadiran Tuhan YME yang telah melimpahkan rahmat, nikmat, dan petunjuk Nya sehingga penulisan laporan akhir penelitian dengan Dana Mandiri dapat diselesaikan. Laporan ini dibuat bertujuan agar hasil penelitian ini bermanfaat bagi pembaca baik kalangan mahasiswa, dosen, maupun masyarakat secara umum.

Kami mengucapkan banyak terima kasih kepada berbagai pihak yang telah membantu proses penyusunan laporan penelitian ini. Semoga hasil penelitian ini bermanfaat bagi pengembangan keilmuan Teknik Industri.

Tangerang Selatan, 24 Februari 2024  
Ketua Tim,

Ir. Mega Bagus Herlambang, ST, MT, PhD, IPM, ASEAN. Eng

## ABSTRAK

Terdapat 6 variabel yang digunakan oleh perusahaan pembangkit listrik tenaga batubara untuk memprediksi nilai *active power*, antara lain *main steam flow*, *coal flow*, *drum pressure*, *feedwater flow*, *main steam pressure*, dan *feedwater pressure*. Dengan menggunakan 4 algoritma *machine learning* yaitu *linear regression*, *decision tree*, *support vector machines*, dan *random forest* didapat bahwa algoritma *linear regression* mampu memberikan prediksi dengan metrik nilai kesalahan terkecil. Proses pembuatan model *machine learning* menggunakan data historis operasional selama 2 bulan yang terdiri dari 56 baris data. Melalui model regresi yang dihasilkan didapat variabel yang paling berpengaruh secara positif terhadap nilai *active power* adalah *feedwater pressure*. Sementara variabel yang paling berpengaruh negatif terhadap nilai *active power* adalah *drum pressure*.

Kata Kunci : machine learning, active power, regresi linear, supervised learning

## ABSTRACT

*There are 6 variables utilized by coal-fired power generation companies to predict the value of active power, namely main steam flow, coal flow, drum pressure, feedwater flow, main steam pressure, and feedwater pressure. By employing 4 machine learning algorithms, namely linear regression, decision tree, support vector machines, and random forest, it was found that the linear regression algorithm is capable of providing predictions with the smallest error metric. The process of creating the machine learning model utilized historical operational data spanning 2 months consisting of 56 rows of data. Through the generated regression model, it was determined that the variable most positively affecting the value of active power is feedwater pressure. Meanwhile, the variable exerting the most negative influence on the value of active power is drum pressure.*

*Keywords: machine learning, active power, linear regression, supervised learning*

## DAFTAR ISI

<b>HALAMAN PENGESAHAN.....</b>	<b>ii</b>
<b>PRAKATA .....</b>	<b>iv</b>
<b>ABSTRAK .....</b>	<b>v</b>
<b>ABSTRACT .....</b>	<b>v</b>
<b>DAFTAR ISI.....</b>	<b>vii</b>
<b>DAFTAR GAMBAR.....</b>	<b>viii</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Rumusan masalah .....	2
1.3 Tujuan .....	2
1.4 Manfaat .....	2
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>3</b>
2.1 Machine learning .....	3
2.2 Linear regression.....	4
2.3 Support vector machines.....	5
2.4 Decision tree .....	6
2.5 Random forest.....	7
<b>BAB III METODE.....</b>	<b>9</b>
3.1 Diagram Alir .....	9
3.2 Keterangan Diagram Alir Metodologi Penelitian .....	11
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>15</b>
4.1 Alur Proses di piranti lunak Orange.....	15
4.2 Business Understanding.....	15
4.3 Data Understanding .....	16
4.4 Data Preparation.....	18
4.5 Modeling .....	18
4.6 Evaluasi.....	19
4.7 Deployment.....	21
4.8 Koefisien regresi .....	21

<b>BAB V KESIMPULAN .....</b>	<b>23</b>
5.1 Kesimpulan .....	233
5.2 Saran .....	23
<b>DAFTAR PUSTAKA .....</b>	<b>24</b>

## DAFTAR GAMBAR

Gambar 1.1 Ilustrasi active power .....	1
Gambar 2.1 Ilustrasi model machine learning .....	4
Gambar 2.2 Ilustrasi hyperplane di support vector machines .....	6
Gambar 2.3 Ilustrasi decision tree.....	7
Gambar 3.1 Gambar diagram alir metodologi penelitian.....	11
Gambar 4.1 Tampilan workflow di Orange .....	15
Gambar 4.2 Tampilan data yang diolah dalam format tabel.....	16
Gambar 4.3 Pengamatan data menggunakan Feature Statistics.....	17
Gambar 4.4 Konfigurasi cross-validation pada widget Test & Score.....	19
Gambar 4.5 Evaluasi model regresi di test set .....	20
Gambar 4.6 Scatter plot prediksi linear regression dengan nilai sesungguhnya....	21

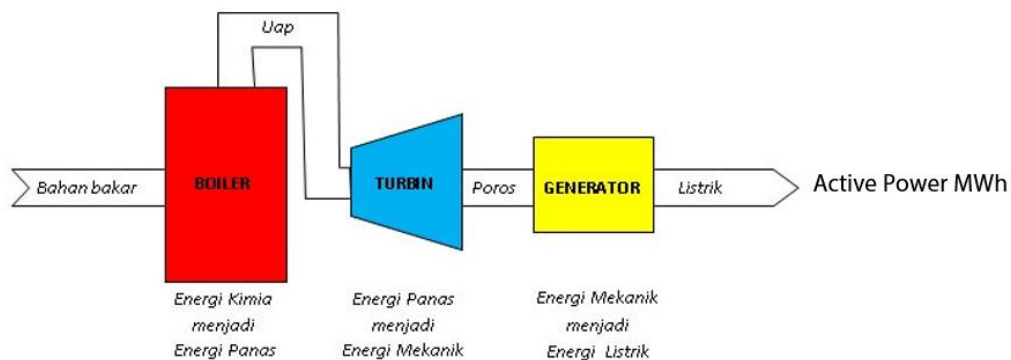


# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Pembangkit listrik tenaga batubara merupakan salah satu sumber utama dalam memenuhi kebutuhan energi listrik di banyak negara. Namun, pengoperasian pembangkit listrik ini sering kali dihadapkan pada tantangan dalam mengoptimalkan produksi energi, terutama terkait dengan prediksi dan pengaturan *active power* (daya aktif) yang dihasilkan.

*Active power* dalam konteks pembangkit listrik tenaga uap merujuk pada daya yang dihasilkan oleh pembangkit tersebut yang benar-benar digunakan untuk melakukan pekerjaan atau menghasilkan energi listrik. *Active power* dinyatakan dalam watt (W) atau megawatt (MW) dan merupakan komponen utama dari daya listrik yang dihasilkan oleh pembangkit. Ilustrasi *active power* dapat dilihat pada gambar 1.



Gambar 1.1 Ilustrasi *active power*

Pada perusahaan pembangkit listrik tenaga batubara, pengaturan daya aktif membutuhkan pemahaman yang mendalam tentang berbagai faktor yang memengaruhinya, seperti kondisi operasional boiler, kondisi cuaca, permintaan listrik, dan karakteristik bahan bakar batubara. Pengaturan yang tidak tepat dapat mengakibatkan penggunaan bahan bakar yang tidak efisien, kerugian energi, dan bahkan berpotensi menyebabkan ketidakstabilan pada jaringan listrik.

Dalam menghadapi kompleksitas tersebut, implementasi algoritma *machine learning* menjanjikan pendekatan yang efektif dalam memprediksi *active power*. *Machine learning* mampu mempelajari pola-pola kompleks dari data historis, memodelkan hubungan antara berbagai variabel masukan (seperti suhu, tekanan uap, dan komposisi batubara) dengan *active power* yang dihasilkan, dan akhirnya memberikan prediksi yang akurat untuk membantu dalam pengaturan operasional.

Meskipun *machine learning* menawarkan peluang yang besar, implementasinya dalam konteks pembangkit listrik tenaga batubara memiliki tantangan tersendiri. Tantangan tersebut antara lain adalah kurangnya ketersediaan data berkualitas tinggi, kompleksitas model yang dibutuhkan, dan integrasi sistem yang sesuai dengan infrastruktur perusahaan. Namun, dengan strategi yang tepat, tantangan-tantangan ini dapat diatasi, dan implementasi *machine learning* dapat memberikan manfaat signifikan dalam meningkatkan efisiensi dan kinerja pembangkit listrik

## **1.2 Rumusan masalah**

Penelitian ini menginvestigasi dan menerapkan berbagai algoritma *machine learning* dalam memprediksi *active power* di perusahaan pembangkit listrik tenaga batubara.

## **1.3 Tujuan**

Tujuan penelitian ini antara lain:

1. Menganalisis data historis dari pembangkit listrik untuk memahami pola-pola dan hubungan antara variabel masukan dengan *active power*.
2. Memilih, mengimplementasikan, dan membandingkan kinerja beberapa algoritma *machine learning* yang tepat untuk memprediksi *active power*.

## **1.4 Manfaat**

Penelitian ini diharapkan dapat meningkatkan pemahaman tentang peran dan potensi aplikasi *machine learning* dalam pengaturan operasional pembangkit listrik tenaga batubara, serta memberikan kontribusi nyata dalam meningkatkan efisiensi dan kinerja industri pembangkit listrik secara keseluruhan.

## BAB II TINJAUAN PUSTAKA

### 2.1 *Machine learning*

*Machine Learning* (ML) adalah cabang dari kecerdasan buatan yang berkaitan dengan pengembangan sistem yang dapat belajar dari data, mengidentifikasi pola, dan membuat keputusan atau prediksi dengan minimal intervensi manusia. Tujuan utama dari machine learning adalah untuk mengembangkan model atau algoritma yang dapat digunakan untuk memahami dan memanfaatkan pola dalam data untuk melakukan tugas tertentu.

Terdapat beberapa penerapan nyata dari *machine learning*. Contoh aplikasi dari *machine learning* di kehidupan sehari-hari seperti aplikasi email yang bisa mendeteksi mana email *spam* dan mana yang bukan. Kemudian *machine learning* juga bisa digunakan untuk memprediksi harga rumah, memprediksi curah hujan, memprediksi proyeksi ekonomi ke depan, dan masih banyak lagi.

Secara umum, *machine learning* dibagi menjadi tiga jenis (Herlambang, 2022), antara lain:

- *Supervised learning*
- *Unsupervised learning*
- *Reinforcement learning*

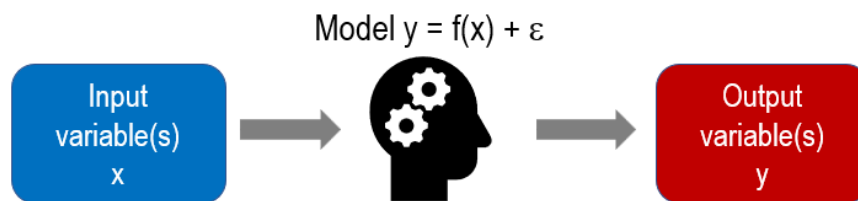
*Supervised learning* adalah cabang dari *machine learning* di mana kita sudah menentukan variabel mana yang akan diprediksi. Jika variabel yang diprediksi bertipe numerik (angka), maka menggunakan algoritma regresi. Kemudian jika variabel yang diprediksi bertipe kategori, maka menggunakan algoritma klasifikasi.

*Unsupervised learning* merupakan bagian dari *machine learning* di mana tidak terdapat variabel yang ingin diprediksi. Tipe *machine learning* ini lebih fokus mengelompokkan semua data yang dimiliki berdasarkan kemiripan fitur (*feature*) yang dimiliki oleh semua data yang ada. Beberapa algoritma *unsupervised learning* yang sering digunakan adalah *clustering*.

*Reinforcement learning* adalah tipe *machine learning* dengan menggunakan konsep *reward/punishment*, di mana model *machine learning* akan berusaha memaksimalkan *reward* (atau dengan mengurangi *penalty/punishment*) jika model berhasil melakukan

tujuan dari model yang dibuat. Penerapan *reinforcement learning* sering ditemui pada robotik, seperti misalnya mengajarkan bagaimana robot berjalan, berlari, dan melakukan aktivitas lainnya.

Implementasi *machine learning* yang dipakai di dunia nyata berupa model *machine learning* (Gambar 2.1). Model ini merupakan algoritma statistik yang bisa merupakan hasil dari implementasi *supervised*, *unsupervised*, atau *reinforcement learning*.



Gambar 2.1 Ilustrasi model machine learning (Herlambang, 2022)

*Machine Learning* adalah bidang yang terus berkembang dalam kecerdasan buatan yang menawarkan berbagai teknik dan algoritma untuk memahami dan memanfaatkan pola dalam data. Dengan kemampuan untuk belajar dari pengalaman dan meningkatkan kinerja seiring waktu, machine learning telah menjadi inti dari banyak aplikasi teknologi modern.

## 2.2 *Linear Regression*

*Linear regression* (regresi linear) adalah metode statistik yang digunakan untuk memodelkan hubungan linier antara satu atau lebih variabel independen (biasanya disebut sebagai "fitur" atau "variabel prediktor") dan variabel dependen (biasanya disebut sebagai "variabel respons" atau "variabel yang diprediksi"). Tujuan utama dari linear regresi adalah untuk menemukan garis lurus terbaik yang mendekati pola hubungan antara variabel-variabel ini.

Pertama, mari kita lihat model linear regresi sederhana yang melibatkan satu variabel independen dan satu variabel dependen. Model ini memiliki bentuk matematis sebagai berikut:

$$Y = \beta_0 + \beta_1 X + \epsilon$$

Di mana:

- $Y$  adalah variabel dependen.
- $X$  adalah variabel independen.
- $\beta_0$  adalah *intercept* (titik potong) dari garis regresi di sumbu  $Y$ .
- $\beta_1$  adalah koefisien regresi yang mengukur perubahan rata-rata dalam  $Y$  untuk setiap satu unit perubahan dalam  $X$ .
- $\epsilon$  adalah *error*.

Tujuan utama dalam analisis regresi adalah untuk menemukan estimasi terbaik untuk koefisien regresi  $\beta_0$  dan  $\beta_1$  sehingga garis regresi yang dihasilkan adalah yang paling cocok dengan data yang diamati. Estimasi ini sering dilakukan dengan menggunakan metode kuadrat terkecil (*least squares method*), di mana kita mencari garis regresi yang memiliki kesalahan kuadrat terkecil antara nilai yang diprediksi oleh garis tersebut dan nilai yang diamati dalam data.

### 2.3 Support Vector Machines

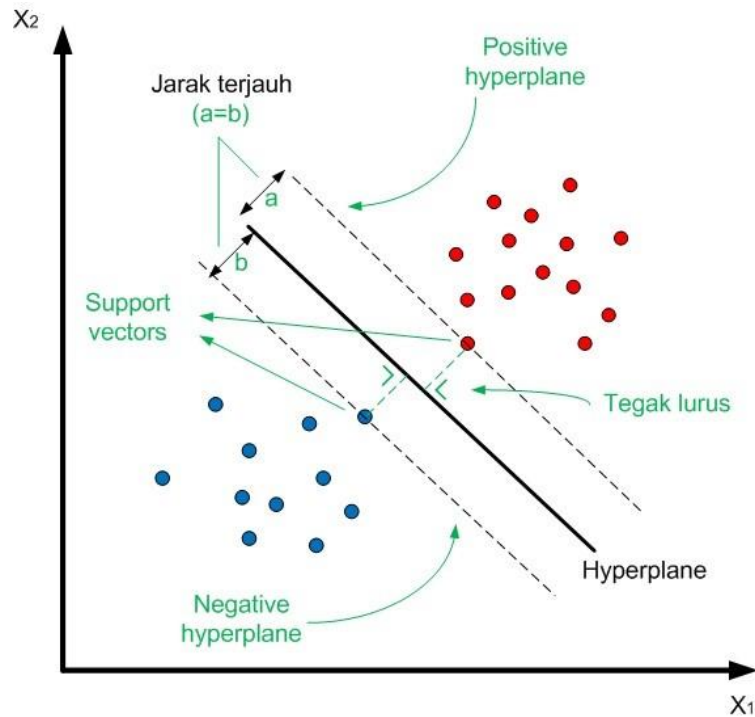
Support Vector Machine (SVM) adalah salah satu algoritma pembelajaran mesin yang digunakan untuk pemisahan kelas dan regresi. Tujuannya adalah untuk menemukan hyperplane optimal yang memisahkan dua kelas data dengan margin maksimum. SVM dapat digunakan baik untuk klasifikasi (memisahkan data menjadi kelas-kelas diskret) maupun regresi (memperkirakan nilai-nilai berkelanjutan).

*Hyperplane* adalah generalisasi dari konsep garis atau bidang dalam dimensi yang lebih tinggi. Dalam SVM, *hyperplane* ini digunakan untuk memisahkan dua kelas data dalam ruang fitur. Ilustrasi *hyperplane* bisa dilihat pada gambar 2.2. Secara matematis, sebuah *hyperplane* dalam dimensi  $n$  didefinisikan oleh persamaan:

$$\mathbf{w} \cdot \mathbf{x} + b = 0$$

Di mana:

- $\mathbf{w}$  adalah vektor normal terhadap hyperplane.
- $\mathbf{x}$  adalah vektor fitur dari titik data.
- $b$  adalah bias (intercept).



Gambar 2.2. Ilustrasi *hyperplane* di *support vector machines*.

SVM memiliki banyak aplikasi dalam klasifikasi, regresi, pemisahan pola, dan deteksi anomali. Beberapa contoh aplikasi termasuk klasifikasi teks, pengenalan wajah, pengenalan tulisan tangan, dan bioinformatika.

## 2.4 Decision Tree

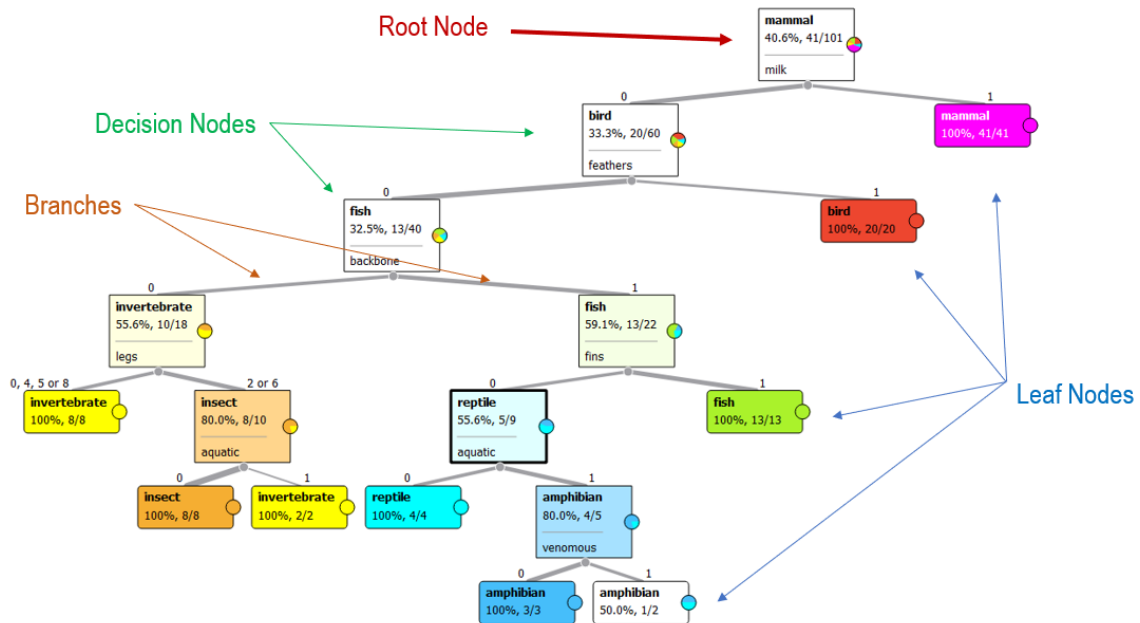
*Decision Tree* (pohon keputusan) adalah model prediktif dalam *machine learning* yang mengambil bentuk struktur pohon, di mana setiap simpul internal mewakili keputusan berdasarkan fitur-fitur input, setiap cabang merepresentasikan hasil dari keputusan tersebut, dan setiap daun mewakili hasil prediksi. *Decision tree* adalah algoritma yang serbaguna dan mudah diinterpretasikan, yang dapat digunakan untuk klasifikasi dan regresi.

Pohon keputusan terdiri dari simpul-simpul (*nodes*) dan cabang-cabang (*branches*) yang menghubungkannya, di mana:

- Simpul (*nodes*): Mewakili keputusan berdasarkan fitur input.

- Cabang (*branches*): Merepresentasikan kemungkinan hasil dari keputusan tersebut.
- Daun (*leaf nodes*): Mewakili hasil prediksi atau kelas yang ditetapkan.

Ilustrasi dari *decision tree* dapat dilihat melalui gambar 2.3 berikut ini.



Gambar 2.3. Ilustrasi *decision tree* (Herlambang, 2022)

## 2.5 Random Forest

*Random Forest* adalah sebuah metode *ensemble learning* yang digunakan dalam *machine learning*, terutama untuk klasifikasi dan regresi. Ini adalah gabungan dari banyak pohon keputusan yang dibuat selama proses pelatihan dan menggabungkan hasil prediksi dari setiap pohon untuk memperbaiki kinerja dan mengurangi *overfitting*.

*Ensemble Learning* adalah pendekatan dalam *machine learning* di mana beberapa model (disebut "*learner*" atau "*base learner*") digabungkan untuk membentuk model yang lebih kuat. Prinsip di balik *ensemble learning* adalah bahwa gabungan beberapa model dapat menghasilkan hasil yang lebih baik daripada model tunggal.

Setiap pohon dalam *Random Forest* adalah pohon keputusan, yang dibuat dengan menggunakan algoritma pohon keputusan seperti ID3, CART, atau C4.5. Pohon keputusan dalam *Random Forest* dibangun secara independen satu sama lain.

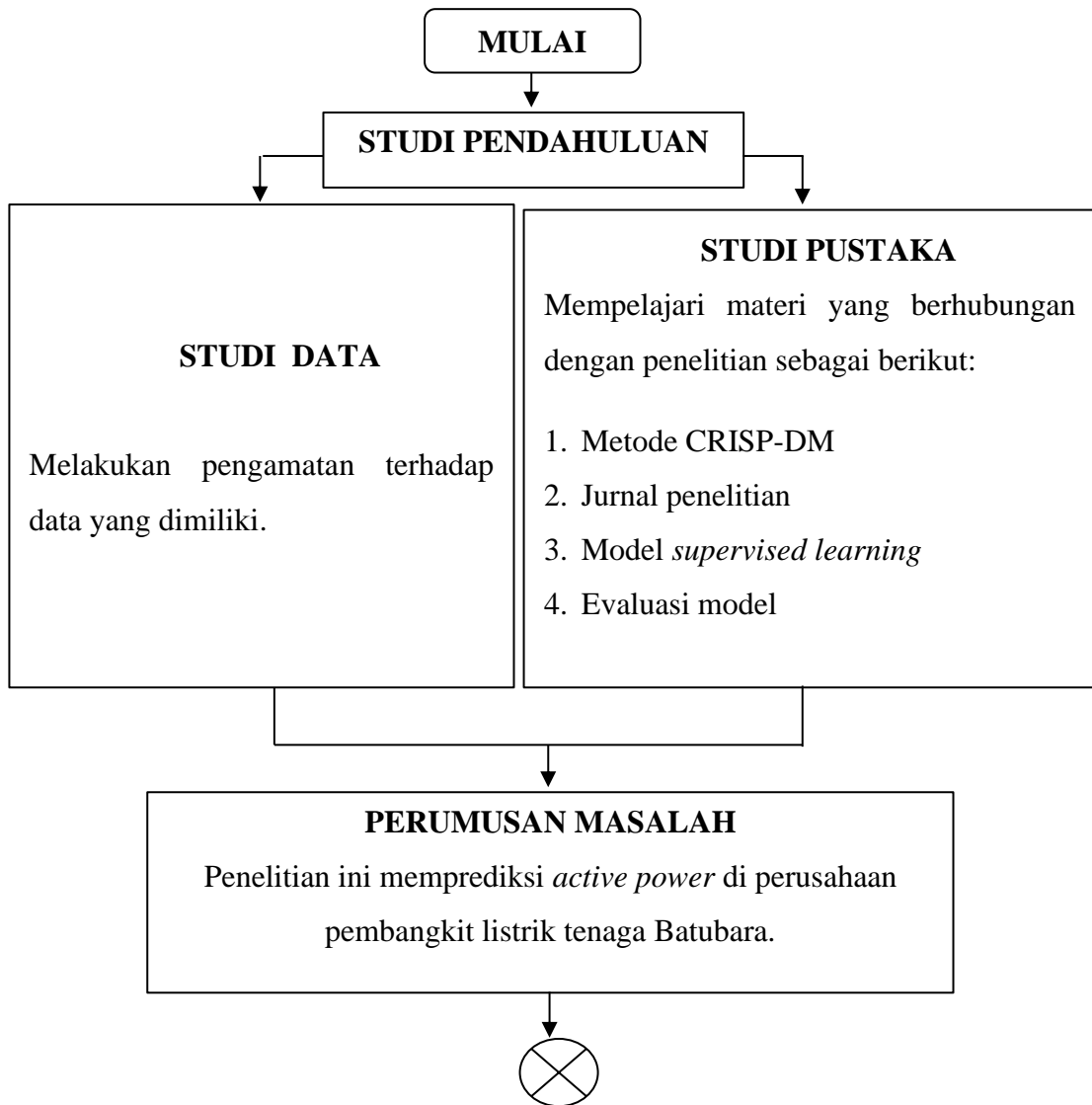
Untuk melakukan prediksi, setiap pohon dalam *Random Forest* memberikan prediksi berdasarkan input, dan hasil prediksi dari semua pohon diensemble secara

bersama-sama untuk menghasilkan prediksi final. Dalam klasifikasi, ini sering dilakukan dengan mengambil mayoritas suara dari semua pohon, sedangkan dalam regresi, ini bisa berupa rata-rata prediksi dari semua pohon.



## BAB III METODE

### 3.1 Diagram Alir Metodologi Penelitian





### **TUJUAN PENELITIAN**

Tujuan dari penelitian ini sebagai berikut:

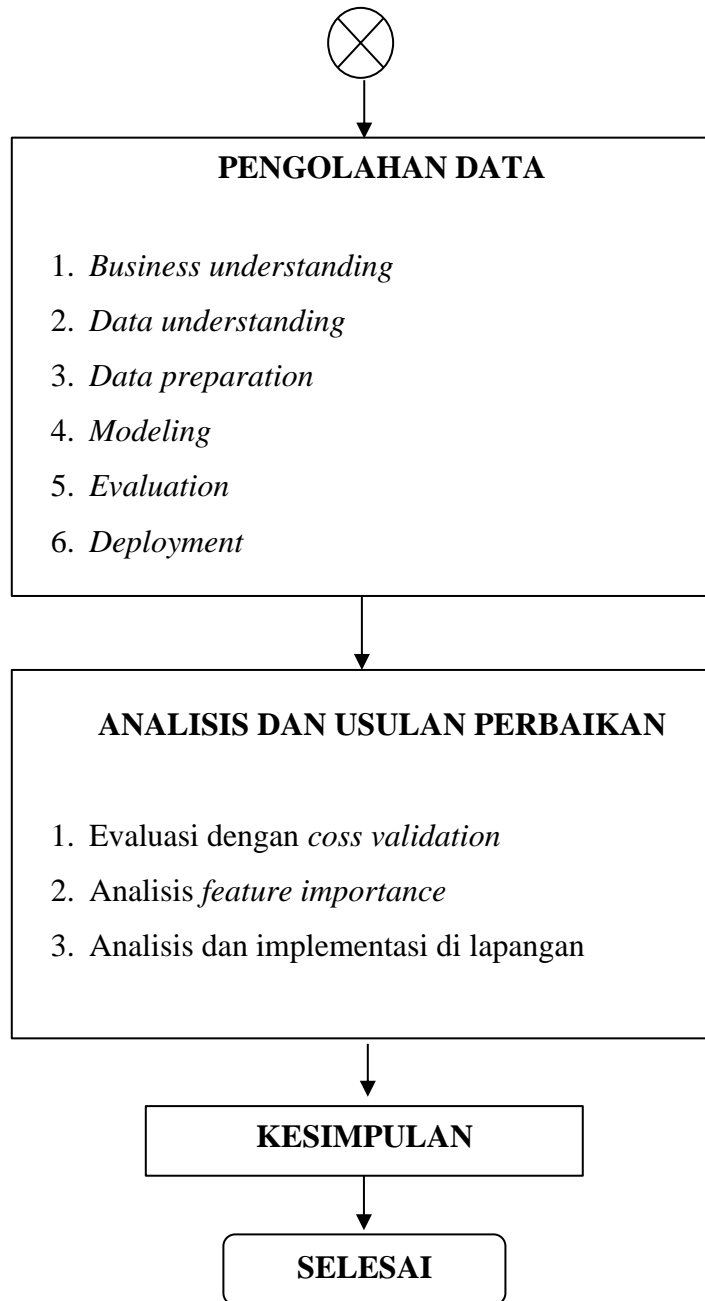
1. Menganalisis data historis dari pembangkit listrik untuk memahami pola-pola dan hubungan antara variabel masukan dengan active power.
2. Memilih, mengimplementasikan, dan membandingkan kinerja beberapa algoritma machine learning yang tepat untuk memprediksi *active power*.



### **PENGUMPULAN DATA**

1. Data historis operasional 2 bulan ke belakang
2. Data *main steam flow*
3. Data *coal flow*
4. Data *drum pressure*
5. Data *feedwater flow*
6. Data *main steam pressure*
7. Data *feedwater pressure*
8. Data *active power*





Gambar 3. 1 Gambar Diagram Alir Metodologi Penelitian

### 3.2 Keterangan Diagram Alir Metodologi Penelitian

1. Mulai  
Merupakan tahap awal dalam melakukan penelitian terkait indeks demam berdarah
2. Studi Pendahuluan

Studi pendahuluan merupakan suatu langkah yang dilakukan untuk melihat variabel, sampel atau populasi yang akan diteliti. Melalui studi ini dapat memperoleh suatu informasi yang akan dibahas dalam permasalahan yang diangkat dan dapat memperjelas topik utama yang akan di bahas di dalam penelitian.

3. Studi Data

Studi data merupakan suatu studi yang dilakukan secara langsung dengan melakukan pengamatan pada objek data yang diteliti. Pada penelitian ini data yang digunakan adalah data-data yang terkait dengan jumlah penyebaran demam berdarah di propinsi di Indonesia.

4. Studi Pustaka

Studi pustaka dapat digunakan untuk mengumpulkan informasi – informasi yang berkaitan pada teori, referensi, dan literatur di dalam kegiatan penelitian. Studi Pustaka yang dilakukan adalah melakukan kajian terkait proses yang akan dilakukan di dalam penelitian ini seperti menggunakan pendekatan metode CRISP-DM dan beberapa algoritma *machine learning* dalam pembuatan modelnya. Beberapa jurnal penelitian yang relevan juga menjadi bahan kajian dalam penelitian ini.

5. Perumusan Masalah

Setelah melakukan kajian Pustaka dan terhadap data yang ada, maka dirumuskan masalah dalam penelitian kali ini. Penelitian ini menginvestigasi dan menerapkan berbagai algoritma *machine learning* dalam memprediksi *active power* di perusahaan pembangkit listrik tenaga batubara.

6. Tujuan Penelitian

Adapun tujuan pada penelitian ini adalah sebagai berikut :

- a. Menganalisis data historis dari pembangkit listrik untuk memahami pola-pola dan hubungan antara variabel masukan dengan active power.
- b. Memilih, mengimplementasikan, dan membandingkan kinerja beberapa algoritma machine learning yang tepat untuk memprediksi *active power*

## 7. Pengumpulan Data dan Pengolahan Data

Pada tahap ini dilakukannya pengumpulan data yang dilanjutkan dengan pengolahan data, dimana data tersebut menyangkut dalam penelitian yang diteliti. Dalam pengolahan data di dasarkan pada landasan teori yang dibahas dengan menggunakan metode yang telah ditentukan berdasarkan teori atau referensi yang ada. Pengumpulan data dan pengolahan data sebagai berikut :

### a. Pengumpulan Data

- 1) Data operasional 2 bulan terakhir.
- 2) Data *main steam flow*  
Jumlah uap yang dihasilkan dari pembakaran batubara.
- 3) Data *coal flow*  
Jumlah/banyaknya batubara.
- 4) Data *drum pressure*  
Tekanan yang ada pada Drum Boiler.
- 5) Data *feedwater flow*  
Jumlah air yang masuk ke dalam pembakaran *boiler*.
- 6) Data *main steam pressure*  
Tekanan uap *main steam*
- 7) Data *feedwater pressure*  
Tekanan air umpan yang akan masuk ke dalam *boiler*
- 8) Data *Active power*  
Daya yang dihasilkan pembangkit. Variabel ini menjadi variabel dependen (target)

### b. Pengolahan Data

Pengolahan data dilakukan dengan metode CRISP-DM, yang terdiri dari beberapa tahapan, antara lain:

- 1) *Business understanding*.
- 2) *Data understanding*
- 3) *Data preparation*
- 4) *Modeling*
- 5) *Evaluation*

## 6) *Deployment*

### 8. Analisis

Setelah melakukan pengumpulan dan pengolahan data, maka pada tahap selanjut nya melakukan analisis dari hasil pengolahan data dan memberikan solusi untuk menyelesaikan permasalahan serta memberikan usulan kepada pihak terkait.

### 9. Kesimpulan dan Saran

Pada tahap ini dapat dilakukan kesimpulan yang berisi dari hasil pengolahan data dan analisis data sesuai dengan tujuan pada penelitian ini.

### 10. Selesai

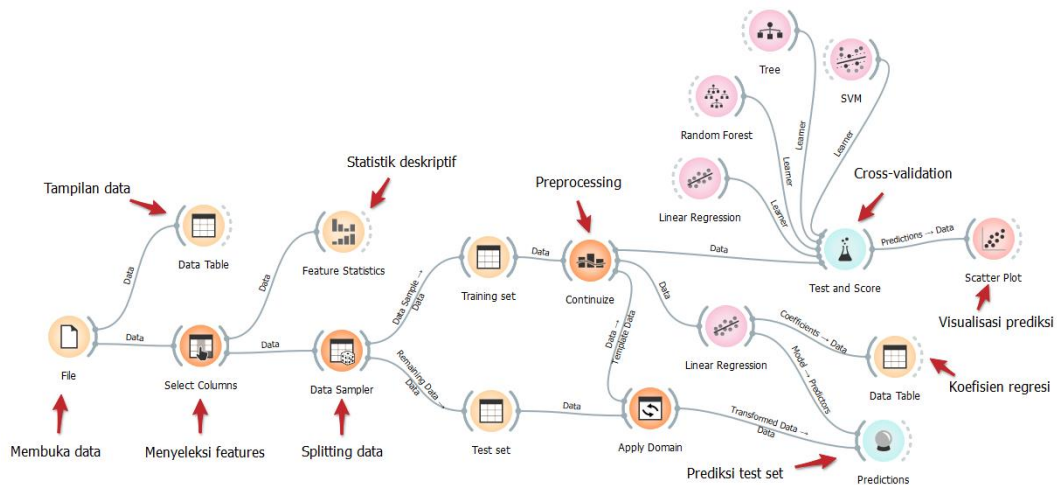
Setelah semua langkah telah diselesaikan, maka dalam penyusunan laporan penelitian ini telah selesai.

## BAB IV HASIL DAN PEMBAHASAN

### 4.1 Alur Proses di piranti lunak Orange

Dengan menggunakan *software* Orange, kita dapat membuat alur proses (*workflow*) mulai dari membuka data hingga didapatkan pembagian kelompok (*clusters*) menggunakan algoritma yang diinginkan.

Tampilan dari *workflow* di *software* Orange secara keseluruhan dapat dilihat pada gambar di bawah ini:



Gambar 4. 1 Tampilan *workflow* di Orange

Melalui gambar di atas dapat dilihat alur proses dari analisis data dari awal membuka data hingga tahapan evaluasi model *clustering* yang sebenarnya merupakan penjabaran dari alur proses CRISP-DM.

### 4.2 Business Understanding

Tahapan awal dari CRISP-DM adalah mendefinisikan masalah. Pada tahapan ini dirumuskan bahwa permasalahan yang ada adalah perusahaan pembangkit ingin memprediksi *active power* berdasarkan beberapa variabel yang dimiliki. Oleh

karena itu, metode yang sesuai adalah dengan menggunakan modle *supervised learning* regresi.

### 4.3 Data Understanding

Tahapan kedua di metode CRISP-DM adalah memahami data yang diperlukan dan yang akan dianalisis. Dalam hal ini data didapatkan dari sensor yang dimiliki oleh perusahaan pembangkit dan diambil dalam operasional 2 bulan terakhir. Kemudian proses analisisnya menggunakan *software* Orange.

Tahapan pertama pada Gambar 4.1 dimulai dengan membuka data menggunakan *widget* 'File'. Kemudian untuk melihat tampilan data dalam format tabel, digunakan *widget* 'Data Table'. Tampilan format tabel dari data yang dianalisis adalah sebagai berikut:

	ACTIVE POWER	MAIN STEAM FLOW	FEEDWATER FLOW	IAL FEEDWATER TE	FEEDWATER PRESS	DRUM PRESS	MAIN STEAM PRESS	MAIN STEAM
1	261.44	787.35	691.81	232.62	15.45	14.98	13.98	
2	292.31	917.27	883.23	273.04	17.97	17.46	16.29	
3	278.54	899.00	860.84	269.95	17.64	17.14	16.03	
4	292.80	909.49	896.74	272.81	18.02	17.50	16.31	
5	286.31	881.83	833.99	271.83	17.23	16.74	15.60	
6	292.24	921.48	913.35	272.87	17.73	17.21	15.97	
7	289.09	906.71	869.59	271.97	17.38	16.87	15.68	
8	262.80	811.37	803.25	266.47	16.61	16.13	15.04	
9	288.05	918.78	902.74	272.16	17.99	17.47	16.27	
10	271.40	878.20	840.05	268.79	17.27	16.77	15.66	
11	217.87	689.58	647.39	256.46	14.21	13.77	12.85	
12	214.40	691.41	652.88	255.61	14.06	13.62	12.69	
13	294.44	924.09	906.75	273.29	17.85	17.33	16.10	
14	291.18	893.10	879.86	272.08	17.35	16.84	15.66	
15	269.19	855.09	826.75	268.31	17.01	16.52	15.41	
16	291.18	917.05	897.58	271.88	18.17	17.65	16.40	
17	216.82	695.39	643.21	255.47	15.14	14.71	13.86	
18	240.98	762.59	729.59	261.50	15.68	15.23	14.22	
19	215.68	701.43	648.54	255.71	14.64	14.20	13.30	
20	275.46	873.26	801.99	270.63	16.68	16.21	15.13	
21	288.18	903.72	837.75	272.11	17.43	16.95	15.81	
22	296.05	918.46	869.40	273.68	18.00	17.49	16.32	
23	260.41	809.21	765.89	266.03	15.67	15.20	14.15	
24	215.41	709.49	639.06	255.68	14.28	13.84	12.92	
25	198.22	643.42	593.98	251.05	13.78	13.36	12.55	
26	195.70	637.66	550.62	250.57	13.50	13.09	12.30	
27	286.74	912.04	886.10	271.41	17.62	17.11	15.90	
28	176.07	588.11	541.18	245.20	13.73	13.32	12.66	

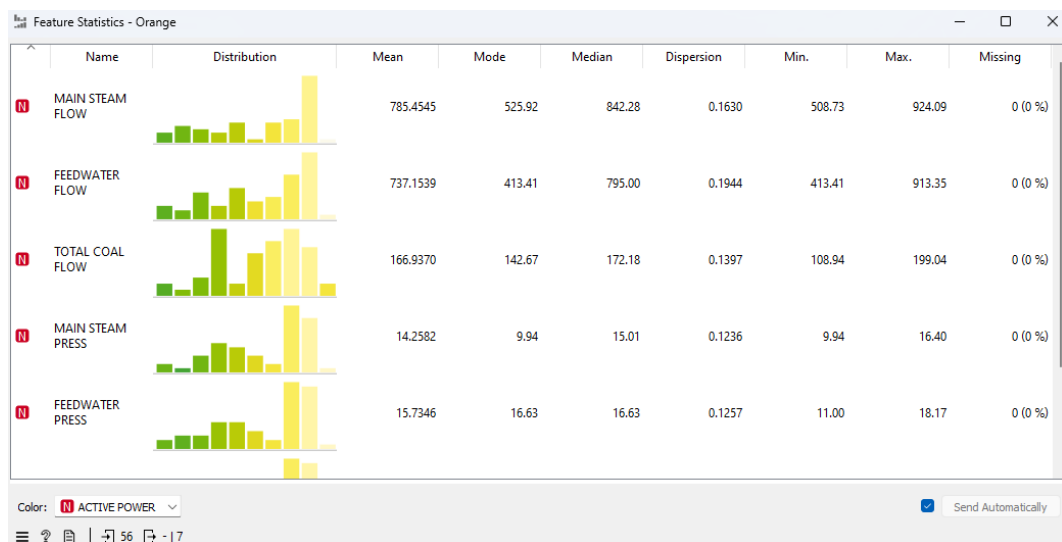
Gambar 4.2 Tampilan data yang diolah dalam format tabel



Dapat diamati pada gambar 4.2 bahwa data yang diolah merupakan data di setiap kota besar di setiap propinsi dengan total adalah sebanyak 56 baris data. Variabel yang digunakan sebagai dasar regresi terdiri dari 8 variabel utama, antara lain:

- *Main steam flow*
- *Coal flow*
- *Drum pressure*
- *Feedwater flow*
- *Main steam pressure*
- *Feedwater pressure*
- *Active power*

Pengamatan terhadap data di Orange dapat dilihat menggunakan *widget* 'Feature Statistics'. Hasil dari tampilan pengamatan data dapat dilihat sebagai berikut:



Gambar 4.3 Pengamatan data menggunakan *Feature Statistics*

Melalui pengamatan di atas maka tidak terdapat nilai yang hilang (*missing values*). Kemudian semua data memiliki tipe data numerik. Dengan demikian, cocok untuk menggunakan model regresi.

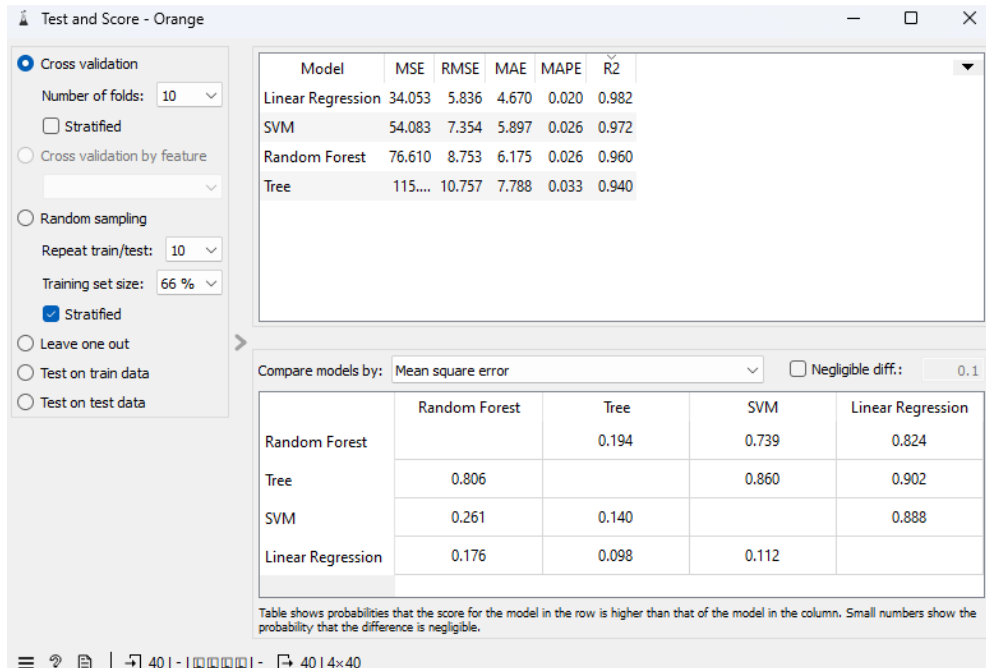
#### 4.4 Data Preparation

Sebelum data siap untuk dibuat modelnya, maka harus mengalami persiapan data. Sebagai contoh jika terdapat nilai kosong (*missing values*) maka harus dibuang nilai kosong atau dilakukan imputasi. Karena tidak terdapat nilai kosong, maka tidak perlu dilakukan imputasi. Walau demikian proses *feature scaling* (menyamakan skala data) perlu dilakukan karena tidak semua skala data di setiap variabel adalah seragam. Oleh karena itu diperlukan *widget* *continuize* di Orange (Gambar 4.1). Proses ini dilakukan baik di *training* maupun *test set*. Setelah itu baru data siap untuk dilakukan *modeling*. Untuk melakukan pembagian *training* dan *test set* dengan proporsi 70:30 digunakan *widget Data Sampler*.

#### 4.5 Modeling

Untuk membuat model regresi maka dapat digunakan beberapa *widget* seperti *linear regression*, *tree*, *SVM*, dan *random forest*. Dalam hal ini kita menggunakan 4 algoritma yaitu regresi linear, *decision tree*, *support vector machines*, dan *random forest*.

Kemudian untuk bisa mendapatkan model yang baik maka kita evaluasi keempat model ini melalui proses *cross-validation*. Di Orange untuk dapat melakukan *cross-validation*, maka kita gunakan *widget Test & Score*. Tampilan konfigurasi untuk *widget Test & Score* dapat dilihat di gambar 4.4.

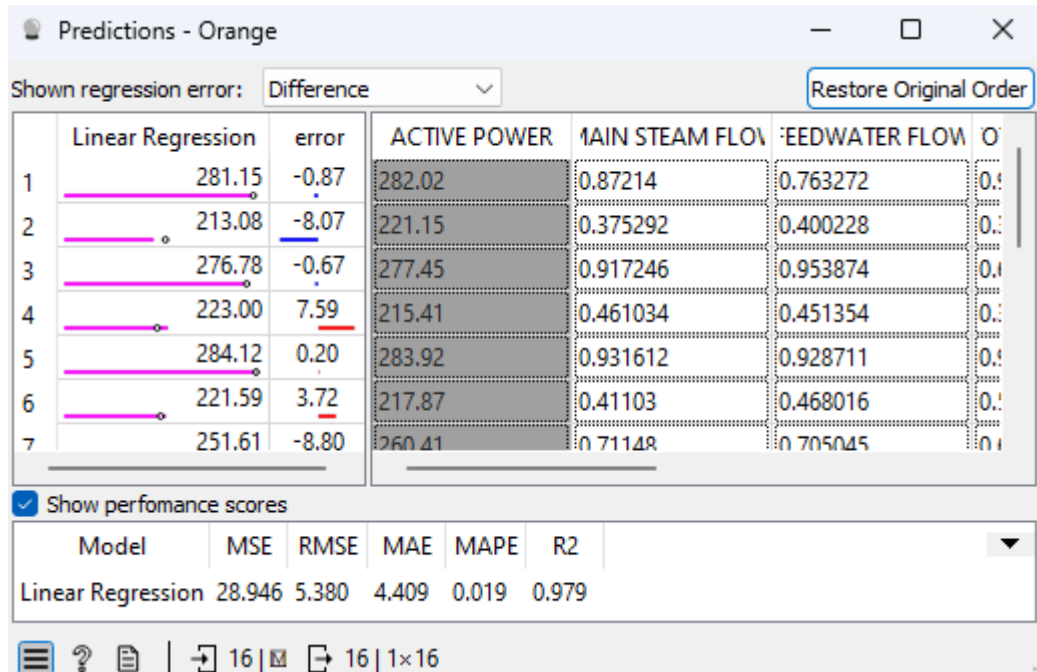


Gambar 4.4. Konfigurasi *Cross-validation* pada widget *Test & Score*

## 4.6 Evaluasi

Proses evaluasi dari model regresi *supervised learning* dilakukan melalui *cross-validation* untuk bisa dilihat mada model yang memberikan performa terbaik di *training set*. Melalui gambar 4.4 dapat dilihat bahwa model *linear regression* menunjukkan nilai error untuk keempat metrik yaitu MSE (*mean squared error*), RMSE (*root mean squared error*), MAE (*mean absolute error*), MAPE (*mean absolute percentage error*). Kemudian *linear regression* juga menunjukkan nilai metrik R<sup>2</sup> (*coefficient of determination*) yang paling tinggi. Dengan demikian bisa dikatakan bahwa untuk prediksi *active power* di *training set*, maka model *linear regression* adalah modle yang terbaik.

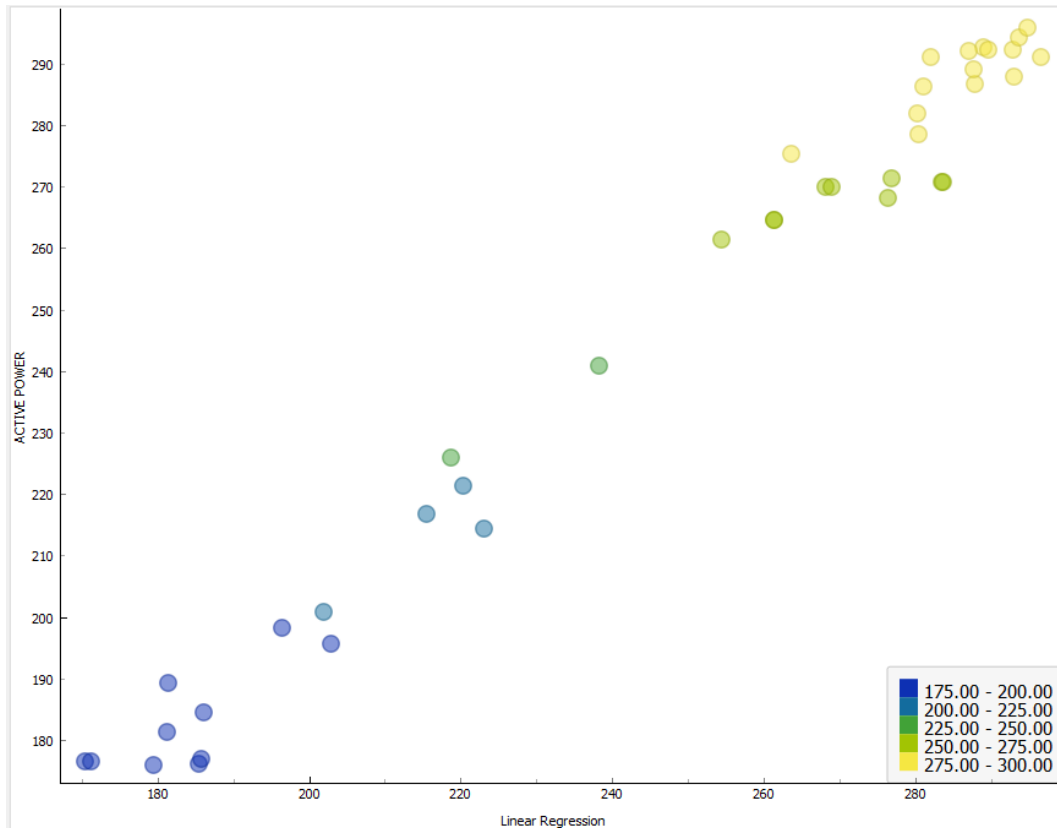
Proses evaluasi selanjutnya adalah menguji performa model terbaik (*linear regression*) untuk data yang ada di *test set*. Proses evaluasi di *test set* juga membandingkan kelima metriks. Hasil evaluasinya dapat dilihat di gambar 4.5.



Gambar 4.5 Evaluasi model regresi di *test set*

Pada gambar 4.5 dapat dilihat bahwa performa *linear regression* di *test set* tidak jauh berbeda dengan *training set*. Sebagai contoh nilai RMSE di *training set* adalah 5,836 sedangkan di *test set* adalah 5,38 artinya model *linear regression* dapat memprediksi dengan cukup akurat untuk data yang tidak pernah dilihat sebelumnya (*test set*).

Evaluasi model juga dapat dilakukan secara visual. Melalui *scatter plot* yang dibuat dengan memetakan nilai *prediksi linear regression* dengan data sesungguhnya, maka dapat diamati bahwa model yang dihasilkan memang linear dengan performa yang sangat baik. Ilustrasi *scatter plot* dapat dilihat di gambar 4.6.



Gambar 4.6. Scatter plot prediksi *linear regression* dengan nilai sesungguhnya

## 4.7 Deployment

Proses *deployment* yang paling sederhana adalah dengan menuliskan laporan (*report*) dari keseluruhan proses modeling dari awal hingga akhir (Larose et al., 2015). Dalam hal ini *deployment* dituangkan dalam bentuk laporan penelitian.

## 4.8 Koefisien Regresi

Dengan menggunakan model *linear regression* kita bisa membentuk sebuah formula matematis hubungan antara variabel independen (*features*) dengan variabel dependen (*target*). Model matematis yang dihasilkan adalah sebagai berikut:

$$\begin{aligned}
 \text{Active Power} = & 163.92 + 1949.41 * \text{Feedwater Pressure} - 1483.96 \\
 & * \text{Drum Pressure} + 354.74 * \text{Main Steam Pressure} - 95.29 \\
 & * \text{Feedwater Flow} + 25.58 * \text{Coal Flow} + 93.85 \\
 & * \text{Main Steam Flow}
 \end{aligned}$$

Melalui persamaan matematis yang dihasilkan dapat dilihat bahwa nilai *intercept* adalah 163.92, artinya nilai *active power* yang dihasilkan jika semua variabel bernilai nol adalah 163.92 MW. Kemudian dari formula di atas dapat diperhatikan juga bahwa variabel yang memiliki nilai koefisien positif artinya semakin besar nilainya maka semakin besar pula nilai *active power* yang dihasilkan. Sementara nilai koefisien yang bernilai negatif justru sebaliknya, di mana semakin besar nilainya justru semakin kecil nilai *active power* yang dihasilkan.

Variabel yang paling berpengaruh secara positif terhadap nilai *active power* adalah *feedwater pressure*. Semakin besar nilai tekanan umpan ke dalam *boiler*, maka semakin besar nilai *active power* yang dihasilkan. Kemudian variabel yang paling berpengaruh negatif terhadap variabel target adalah *drum pressure*. Semakin besar tekanan yang ada di *drum boiler*, maka semakin kecil nilai *active power* yang dihasilkan.

Melalui persamaan matematis yang dihasilkan, dapat disimpulkan bahwa jika perusahaan ingin meningkatkan atau menurunkan nilai *active power*, maka perusahaan bisa prioritaskan kepada variabel-variabel yang memiliki pengaruh terbesar terhadap nilai *active power*.

## **BAB V KESIMPULAN**

### **5.1 Kesimpulan**

Berdasarkan hasil pengolahan data dan analisis pada penelitian ini dapat disimpulkan bahwa :

1. *Machine learning* dapat digunakan di industri pembangkit untuk memprediksi nilai *active power* dengan menggunakan nilai variabel independen lainnya.
2. Algoritma *linear regression* adalah algoritma yang memberikan performa terbaik yang ditandai dengan nilai metrik *error* terkecil dibandingkan dengan algoritma *machine learning* lainnya.
3. Variabel yang paling berpengaruh secara positif terhadap nilai *active power* adalah *feedwater pressure*. Semakin besar nilai tekanan umpan ke dalam *boiler*, maka semakin besar nilai *active power* yang dihasilkan.
4. Variabel yang paling berpengaruh negatif terhadap nilai *active power* adalah *drum pressure*. Semakin besar tekanan yang ada di *drum boiler*, maka semakin kecil nilai *active power* yang dihasilkan.

### **5.2 Saran**

Berdasarkan hasil pengamatan sampai dengan analisis data yang dilakukan dalam penelitian ini, peneliti memberikan usulan kepada perusahaan pembangkit untuk memprioritaskan pemantauan terhadap nilai variabel yang paling berpengaruh terhadap nilai *active power*.

## DAFTAR PUSTAKA

- Conway, D. (2013). <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>
- Herlambang, M. B. (2022). Data Science A-Z Tanpa Coding.
- Larose, D. T., Larose, C. D. (2015). Data Mining and Predictive Analytics. Wiley Publication.