

BAB 1 PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi komputansi dan digitalisasi informasi berdampak pada meningkatnya jumlah penyimpanan data dalam arsip maya dalam dimensi besar untuk dikelola secara manual. Oleh karena itu, teknologi *Machine Learning* menjadi salah satu alternatif yang dapat digunakan.

Teknologi *Machine Learning* mengadopsi proses belajar yang dilakukan oleh manusia, untuk diterapkan pada mesin dengan membuat model matematis yang dapat merefleksikan pola-pola data yang ada. Teknik ini berusaha menciptakan *intelligent agent* untuk mengumpulkan data dari lingkungan kemudian diolah menjadi informasi dan atau pengetahuan baru. Ketersediaan data menjadi salah satu faktor utama yang dibutuhkan pada implementasi teknologi ini. Dengan data, mesin dilatih untuk membaca dan menerjemahkan pola yang ada menggunakan algoritma yang dapat diterapkan.

Di bidang kedokteran, teknologi *Machine Learning* dapat digunakan untuk mendeteksi penyakit seseorang berdasarkan gejala yang ada. Pada bidang *computer vision*, penerapan pengenalan dan pelabelan wajah serta penerjemahan tulisan tangan menjadi teks sudah diaplikasikan.

Selain bertujuan memperoleh ilmu pengetahuan baru (*knowledge discovery*), teknologi *Machine Learning* juga dapat digunakan untuk memprediksi masa depan (*unobserved event*) secara kualitatif maupun kuantitatif, baik pada data diskret maupun kontinyu. Salah satu penerapannya adalah untuk memprediksi jumlah daya listrik yang dihasilkan oleh sistem Pembangkit Listrik Tenaga Surya (PLTS). Daya output pada sistem PLTS yang bersifat fluktuatif dapat diprediksi menggunakan variabel terkait untuk mengetahui produksi energi pada periode berikutnya.

Beberapa penelitian terkait prediksi daya output pada sistem PLTS telah dilakukan, antara lain yang dilakukan oleh K. Anuradha, et al pada tahun 2021. Penelitian dilakukan dengan pendekatan berbasis pembelajaran mesin untuk memprediksi daya yang dihasilkan di seluruh negara bagian India berdasarkan data lingkungan. Metode yang digunakan adalah *Linear Regression* (LR), *Support Vector Machine Regression* (SVMR) dan *Random Forest* (RF). Hasil penelitian menunjukkan bahwa model regresi RF

mempunyai performa lebih baik dari model lainnya, dengan akurasi sebesar 94,01%. Hasil prediksi yang diperoleh, tentunya sangat bergantung terhadap keterkaitan antar variabel pada data set yang tersedia dan metode yang digunakan. Oleh karena itu, dipandang perlu untuk melakukan penelitian prediksi daya output pada sistem PLTS menggunakan metode lain untuk menguji seberapa kuat hubungan antara variabel yang digunakan terhadap hasil prediksi yang didapatkan

1.2 Perumusan Masalah

Berdasarkan latar belakang tersebut di atas, perlu dilakukan analisis prediksi daya output menggunakan metode regresi linier berganda untuk memprediksi jumlah daya listrik yang diproduksi oleh sistem PLTS. Hasil prediksi yang akan diperoleh dapat digunakan untuk menunjukkan berapa besar tingkat akurasi dan korelasi antara variabel yang menggunakan metode regresi linier berganda.

1.3 Tujuan Penelitian

Tujuan dari Tugas Akhir ini adalah mengetahui tingkat akurasi dari metode regresi linier menggunakan Weka 3.8 untuk melakukan prediksi terhadap produksi energi dari sistem PLTS berdasarkan korelasinya dengan data iradiasi, suhu modul, dan suhu lingkungan. Dengan demikian, dapat diperoleh hasil prediksi jumlah daya yang dihasilkan oleh sistem PLTS pada periode berikutnya, salah satunya untuk dapat digunakan sebagai bahan evaluasi performa sistem.

1.4 Batasan Masalah

Analisis yang dilakukan pada Tugas Akhir ini adalah prediksi daya output sistem Pembangkit Listrik Tenaga Surya (PLTS) ditinjau dari rekaman data iradiasi, suhu modul dan suhu lingkungan dengan metode regresi linier menggunakan aplikasi Weka, pada data set `Generation_Data` dan `Weather_Sensor_Data` yang bersumber dari <https://www.kaggle.com/anikannal/solar-power-generation-data>. Tidak dilakukan modifikasi terhadap data yang tersedia, termasuk pengujian terhadap akurasi data akibat kendala koneksi dan kerusakan pada perangkat.

1.5 State of The Art

Penyusunan skripsi ini mengambil beberapa referensi penelitian sebelumnya termasuk jurnal-jurnal yang berhubungan dengan penelitian ini, antara lain sebagai berikut:

JURNAL/PUBLIKASI	KESIMPULAN	PERSAMAAN	PERBEDAAN
<p>Data Mining Menggunakan Regresi Linear untuk Prediksi Harga Saham Perusahaan Pelayaran.</p> <p>(Ekka Pujo Ariesanto Akhmad, 2020)</p> <p>Jurnal Aplikasi Pelayaran dan Kepelabuhanan, Volume 10, Nomer 2</p>	<p>Penelitian ini menggunakan konsep CRISP-DM yang terdiri dari 6 fase, yaitu <i>Bussiness Understanding, Data Understanding, Data Preparation, Modelling, Evaluation</i> dan <i>Deployment</i>.</p> <p>Hasil penelitian menunjukkan masih ada selisih antara harga penutupan saham luaran data testing dengan harga penutupan saham aktual yang ada di bursa saham.</p> <p>Evaluasi nilai <i>Root Mean Square Error (RMSE)</i> menunjukkan angka plus 7,522 dari data aktual harga penutupan saham periode harian PT. BULL.</p>	<ul style="list-style-type: none"> ▪ Prediksi menggunakan regresi ▪ Metode regresi linear berganda ▪ Evaluasi tingkat akurasi menggunakan RMSE 	<ul style="list-style-type: none"> ▪ Objek prediksi adalah harga saham ▪ Menggunakan aplikasi RapidMiner
<p>Data Mining dengan Regresi Linier Berganda untuk Peramalan Tingkat Inflasi.</p> <p>(Amrin, 2016)</p> <p>Jurnal Techno Nusa Mandiri Vol. XIII, No. 1.</p>	<p>Penelitian ini menggunakan metode prediksi regresi linier berganda, untuk memprediksi tingkat inflasi bulanan di Indonesia.</p> <p>Dari hasil analisis data yang dilakukan disimpulkan bahwa model regresi linier berganda yang dihasilkan pada penelitian ini adalah $Y = 0,241X_1 + 0,164X_2 + 0,271X_3 + 0,07X_4 + 0,040X_5 + 0,060X_6 + 0,169X_7 - 0,010$.</p> <p>Adapun nilai koefisien regresi sebesar 0,999 dan koefisien determinasi sebesar 0,997 (99,7%).</p> <p>Performa model regresi linier berganda yang dibentuk dari data training dan divalidasi pada data testing memberikantingkat akurasi prediksi yang cukup baik dengan nilai dengan nilai <i>Mean Absolute Deviation (MAD)</i> sebesar 0.0380, <i>Mean Square Error (MSE)</i> 0.0023, dan nilai <i>Root Mean Square Error (RMSE)</i> sebesar 0.0481.</p>	<ul style="list-style-type: none"> ▪ Prediksi menggunakan regresi ▪ Metode regresi linear berganda ▪ Evaluasi tingkat akurasi menggunakan RMSE 	<ul style="list-style-type: none"> ▪ Objek prediksi adalah tingkat Inflasi ▪ Menggunakan SPSS 17.0
<p><i>Analysis of Solar Power Generation Forecasting using Machine Learning Techniques.</i></p>	<p>Penelitian ini menyajikan pendekatan berbasis pembelajaran mesin untuk analisis pembangkit listrik tenaga surya untuk memprediksi daya yang dihasilkan di seluruh negara bagian India berdasarkan data lingkungan.</p>	<ul style="list-style-type: none"> ▪ Objek prediksi adalah daya output pada sistem PLTS 	<ul style="list-style-type: none"> ▪ Metode yang digunakan adalah LR, SVMR dan RF

JURNAL/PUBLIKASI	KESIMPULAN	PERSAMAAN	PERBEDAAN
(K. Anuradha, Deekshitha Erlapally, G. Karuna, V. Srilakshmi, K. Adilakshmi, 2021) E3S Web of Conferences 309, 01163. ICMED 2021.	Metode yang digunakan adalah <i>Linear Regression</i> (LR), <i>Support Vector Machine Regression</i> (SVMR) dan <i>Random Forest</i> (RF). Perbandingan antara suhu dengan data yang diberikan, pada suhu sekitar 56-55F, daya output meningkat 30% dibandingkan dengan suhu lainnya. Model regresi <i>Random Forest</i> mempunyai performa lebih baik dari model lainnya, dengan akurasi sebesar 94,01%		
Analisis Data Iklim Indonesia Menggunakan Aplikasi Weka dengan Metode Klasifikasi Naive Bayes. (Aswendy, 2016) Jurnal Teknologi Rekayasa Volume 21 No.3	Penelitian ini membahas tentang analisis data iklim Indonesia menggunakan aplikasi Weka dengan metode Naive Bayes. Dari hasil analisis sangat diharapkan dapat membantu memberi peringatan atas iklim ekstrim yang ada di Indonesia. Berdasarkan pengolahan data pada penelitian ini, wilayah Papua memiliki prediksi suhu minimum yang paling tinggi yakni 242 dan suhu maksimum dengan 322. Suhu rata-rata yang tertinggi pada provinsi Papua dengan 282 dan untuk prediksi kelembaban rata-rata yang memiliki intensitas tertinggi pada provinsi Jawa Barat dan Sulawesi Selatan dengan prediksi kelembaban 100. Untuk lama penyinaran yang memiliki prediksi tertinggi yakni pada provinsi DI Yogyakarta dengan prediksi sebesar 17.1.	<ul style="list-style-type: none"> ▪ Tujuannya adalah prediksi ▪ Menggunakan aplikasi Weka 	<ul style="list-style-type: none"> ▪ Prediksi menggunakan klasifikasi ▪ Metode yang digunakan adalah Naive Bayes
<i>Predicting the Power Output of Solar Panels based on Weather and Air Pollution Features using Machine Learning.</i> (Tserenpurev Chuluunsaikhan, Da Bin Choi, Aziz Nasridinov, Woo Seok Choi, Sang Hyun Choi, Young Myoung Kim, 2021) Journal of Korea Multimedia Society Vol. 24, No. 2.	Penelitian ini memprediksi output daya panel surya berdasarkan situasi lingkungan seperti cuaca dan polusi udara. Metode data mining yang digunakan adalah <i>Linear Regression</i> , <i>K-Nearest Neighbors</i> (kNN), <i>Support Vector Regression</i> (SVR), <i>MultiLayer Perceptron</i> (MLP), <i>Random Forest Regressor</i> (RF) dan <i>Gradient Boosting Regressor</i> (GB). Data set yang digunakan berasal dari daerah Seoul, Korea Selatan, antara tahun 2017 dan 2019. Hasil eksperimen menunjukkan	<ul style="list-style-type: none"> ▪ Objek prediksi adalah daya pada sistem PLTS ▪ Prediksi menggunakan regresi ▪ Evaluasi tingkat akurasi menggunakan RMSE 	<ul style="list-style-type: none"> ▪ Data set yang digunakan berasal dari daerah Seoul, Korea Selatan, tahun 2017-2019 ▪ Metode yang digunakan adalah <i>Linear Regression</i>, kNN, SVR, MLP, RF dan GB

JURNAL/PUBLIKASI	KESIMPULAN	PERSAMAAN	PERBEDAAN
	bahwa fitur cuaca dan polusi udara dapat menjadi faktor yang efisien untuk memprediksi output daya solar panel, dengan model terbaik pada RF dengan akurasi 98%.		

1.6 Sistematika Penulisan

Sistematika penulisan merupakan sebuah metode atau urutan dalam menyelesaikan sebuah penelitian. Penulisan skripsi ini dikelompokkan menjadi beberapa subbab dengan sistematika penulisan sebagai berikut:

Bab 1 Pendahuluan

Bab ini memuat pendahuluan penelitian yang terdiri dari latar belakang, perumusan masalah, tujuan penelitian, batasan masalah, *state of the art*, dan sistematika penulisan.

Bab 2 Tinjauan Pustaka

Bab ini memuat konsep, metode, teori dasar dan teori umum yang digunakan berkaitan dengan *data analytic*, *machine learning*, algoritma regresi linier, *data set*, WEKA dan Pembangkit Listrik Tenaga Surya (PLTS).

Bab 3 Metode Penelitian

Bab ini memuat tentang metode penelitian, analisis data set, preprocessing data serta analisis algoritma regresi linier berganda.

Bab 4 Hasil dan Pembahasan

Bab ini memuat dan membahas mengenai implementasi proses prediksi yang dilakukan menggunakan algoritma regresi linier berganda, hasil pengujian, serta analisis dari hasil yang didapatkan.

Bab 5 Kesimpulan dan Saran

Bab terakhir ini berisi kesimpulan dan saran dari proses prediksi menggunakan algoritma regresi linier berganda yang telah dilakukan.

Daftar Referensi

Lampiran