

BAB I

PENDAHULUAN

1.1 Latar Belakang

Seiring dengan berjalannya waktu perkembangan teknologi dalam bidang IT (*Information Technology*) semakin meningkat, dan sudah banyak sekali penerapannya dalam membantu pekerjaan manusia, hal ini tidak lepas dari bidang perawatan manufaktur industri. *Information Technology* (IT) merupakan istilah umum terhadap sebuah teknologi yang dapat membantu manusia dalam menerima, mengolah, menyimpan maupun menyebarkan informasi. IT sendiri teknologi yang menyatukan komputasi dan komunikasi berkecepatan tinggi, agar informasi yang disebarkan dapat mudah diakses. Selain itu IT juga mampu melakukan visualisasi terhadap informasi yang didapatkannya, sehingga tidak ada batasan dalam jenis informasi yang akan digunakan.

Dalam memanfaatkan ilmu ini memerlukan teknologi IT yang sering disebut dengan teknologi *Machine Learning*. *Machine Learning* ini memberikan sebuah keputusan yang optimal berdasarkan data yang diterimanya, oleh karena itu pengumpulan data yang akurat dan sesuai merupakan hal yang sangat penting. *Machine Learning* merupakan mesin yang dikembangkan untuk mampu belajar dengan sendirinya tanpa adanya arahan dari pengguna, biasanya penggunaan *Machine Learning* ini memerlukan ilmu statistika, matematika, dan *data mining* yang kuat agar mesin mampu belajar untuk menganalisa data tanpa perlu diperintah kembali. Secara umum, terdapat 2 teknik pembelajaran *Machine Learning* yaitu *Supervised Learning* dan *Unsupervised Learning*. Hasil yang dapat diberikan sangat bervariasi baik itu kualitatif maupun kuantitatif, dengan data diskrit ataupun data kontinu (Budi Santosa & Ardian Umam, 2018).

Dalam melakukan pengolahan data dan menciptakan model *Machine Learning* memerlukan ilmu *Data Science*. Ilmu *Data Science* merupakan ilmu gabungan dari ilmu *Data Analytics* dan ilmu *Science* dalam mengolah data besar (*Big Data*) untuk menemukan suatu pola atau *insight* yang dapat digunakan untuk pengambilan keputusan atau penyelesaian permasalahan. Seseorang yang menerapkan ilmu ini disebut sebagai *Data Scientist*. Seorang *Data Scientist* memerlukan kemampuan analisis yang kuat, pemahaman akan data yang baik, serta kemampuan dalam mengkomunikasikan hasil temuan. Kemampuan-kemampuan ini bisa dipelajari secara otodidak maupun dengan

pembelajaran secara intensif. Alasan seorang *Data Scientist* memerlukan kemampuan tersebut dikarenakan dalam data yang jumlahnya sangat banyak, seorang *Data Scientist* perlu menentukan data seperti apa yang seharusnya diambil dan diolah untuk menyelesaikan permasalahan yang sedang dihadapi saat itu (Fadhilla, 2019).

Untuk melakukan penelitian ini, penulis melakukan kerja sama dengan pihak PT. INTEGRAL INDUSTRIAL INDONESIA, dalam pengambilan data. PT. INTEGRAL INDUSTRIAL INDONESIA merupakan sebuah perusahaan yang bergerak pada bidang lingkungan industrial yang kompleks, dimana optimalisasi desain, operasi, dan pemeliharaan aset merupakan hal yang sangat penting. PT. INTEGRAL INDUSTRIAL INDONESIA membangun platform perangkat lunak khusus untuk mengotomatiskan pekerjaan pengetahuan dan membangun keunggulan kompetitif yang berkelanjutan dengan memberikan pengambilan yang tinggi selama seluruh siklus hidup aset.

Data yang didapatkan merupakan data masa lalu pada sensor getaran yang dipasangkan pada salah satu *gearbox* mesin kertas yang sedang melakukan produksi pada suatu perusahaan kertas yang sudah melakukan kerja sama dengan PT. INTEGRAL INDUSTRIAL INDONESIA. Adapun permasalahan yang ingin diselesaikan oleh pihak PT. INTEGRAL INDUSTRIAL INDONESIA yaitu melakukan penelitian terhadap variabel-variabel getaran yang diterima sensor dapat melakukan prediksi terhadap temperatur *gearbox* pada mesin kertas. Dengan mengetahui hasil prediksi temperatur ini pihak PT. INTEGRAL INDUSTRIAL INDONESIA mampu melakukan pengembangan yang lebih baik pada sensor yang dimilikinya.

Alasan pihak PT. INTEGRAL INDUSTRIAL INDONESIA ingin menyelesaikan permasalahan ini dikarenakan untuk menghindari terjadi *overheat* pada *gearbox* mesin kertas, sehingga dapat memberikan pihak konsumen untuk memantau secara berkala. Dikarenakan jika *gearbox* mesin kertas terjadi *overheat* akan menyebabkan kerugian yang cukup besar bagi pihak konsumen, seperti kemungkinan terbakarnya mesin kertas sehingga perlu pengeluaran untuk mengganti mesin tersebut, ataupun dikarenakan *gearbox* mengalami *overheat* maka mesin akan mengalami *breakdown* atau kerusakan yang cukup sehingga produksi harus diberhentikan yang akan menyebabkan tidak dapat memenuhi *demand* atau *target* produksi.

Pengambilan pada jangka waktu selama 52 hari dimulai pada tanggal 12 Januari 2023 hingga tanggal 04 Maret 2023. Dalam penelitian tugas akhir ini, memiliki tujuan yaitu melakukan prediksi temperatur dengan getaran yang dihasilkan *gearbox* mesin kertas sebagai *feature*. Terdapat 3 model *Machine Learning* yang digunakan yaitu Regresi Linear Berganda, Regresi *Random Forest*, dan Regresi *Support Vector Machine*. Dari 3 model ini akan dibandingkan dengan melihat nilai, RMSE, MAPE, dan *R-Squared* yang dihasilkan untuk menentukan model yang terbaik. Dilanjutkan dengan penentuan *Feature Importance* untuk menentukan variabel yang paling mempengaruhi dalam model untuk melakukan prediksi.

Dengan adanya penelitian diharapkan dapat memberikan keuntungan bagi penulis, pihak PT. INTEGRAL INDUSTRIAL INDONESIA, dan konsumen yang akan membeli produk sensor dari PT. INTEGRAL INDUSTRIAL INDONESIA. Dikarenakan penulis akan mendapatkn ilmu tentang penerapan *Machine Learning* pada bidang industri dan mengetahui cara untuk melakukannya. Untuk PT. INTEGRAL INDUSTRIAL INDONESIA mampu melakukan pengembangan sensor dan dapat memperluas pasar penjualannya sehingga mampu memberikan kenaikan keuntungan penjualan. Dan untuk pihak konsumen yang membeli produk sensor ini, dapat menjaga aset produksi yang dimiliki saat ini dengan lebih baik, dengan menghindari kemungkinan terjadinya kerusakan seperti *overheat* pada *gearbox* sehingga mampu memperpanjang umur mesin yang dimilikinya saat ini.

1.2 Perumusan Masalah

Adapun perumusan masalah pada tugas akhir ini yaitu:

1. Bagaimana suatu model *Machine Learning* melakukan prediksi temperatur berdasarkan model Regresi Linear Berganda, Regresi *Random Forest*, Regresi *Support Vector Machine*, dan Regresi *Support Vector Machine*.
2. Bagaimana menentukan model *Machine Learning* yang terbaik berdasarkan model yang sudah dibuat dengan membandingkan nilai RMSE, MAPE, dan *R-Squared*.
3. Bagaimana untuk menentukan variabel yang paling mempengaruhi perubahan temperatur dalam melakukan prediksi menggunakan *Machine Learning*.

4. Bagaimana dengan mengetahui hasil penelitian ini mampu memberikan dampak pada finansial perusahaan.

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian tugas akhir ini yaitu sebagai berikut:

1. Mampu membuat model *Machine Learning* dalam melakukan prediksi temperatur *gearbox* menggunakan Regresi Linear Berganda, Regresi *Random Forest*, dan Regresi *Support Vector Machine*.
2. Mampu menentukan atau memilih model *Machine Learning* yang terbaik diantara model yang sudah dibuat dengan membandingkan nilai RMSE, MAPE, dan *R-Squared*.
3. Mampu untuk menentukan variabel yang paling mempengaruhi perubahan temperatur dalam melakukan prediksi menggunakan *Machine Learning*.
4. Mampu memberikan analisis finansial bagi perusahaan berdasarkan hasil penelitian yang didapatkan.

1.4 Batasan Masalah

Adapun batasan masalah yang dibuat agar tugas akhir ini lebih terarah, berikut merupakan batasan masalah dalam tugas akhir, sebagai berikut:

1. *Dataset* yang digunakan dalam membuat model *Machine Learning* didapatkan dari sensor yang dipasangkan pada *Gearbox* mesin kertas.
2. *Dataset* yang didapatkan berisikan data getaran mesin dalam bentuk *JSON file*.
3. Data yang digunakan dalam pembuatan model sebanyak 1000 data yang berisikan UID, Tanggal, Waktu, *XYZ Peak*, *XYZ rms*, Temperatur, dan *Rate*.
4. Perancangan *Machine Learning* akan menggunakan 3 model yaitu Regresi Linear Berganda, Regresi *Random Forest*, dan Regresi *Support Vector Machine*.
5. Dalam melakukan pemilihan model yang terbaik dilakukan dengan membandingkan hasil RMSE, MAPE, dan *R-Squared*.
6. Dalam melakukan analisis menggunakan *Feature Importances*.

1.5 State of The Art

Penyusunan tugas akhir ini menggunakan beberapa referensi penelitian yang sebelumnya telah dilakukan dan dituliskan dalam bentuk jurnal, berikut merupakan beberapa jurnal yang menjadi referensi dalam tugas akhir ini:

Judul Jurnal	Pembahasan
<p>Judul:</p> <p>Comparison of multivariate linear regression and a machine learning algorithm developed for prediction of precision warfarin dosing in a Korean population</p> <p>Peneliti:</p> <p>Van Lam Nguyen, Hoang Dat Nguyen, Yong-Soon Cho, Ho-Sook Kim, Il-Yong Han, Dae-Kyeong Kim, Sangzin Ahn, Jae-Gook Shin</p> <p>Lokasi:</p> <p>Inje University College of Medicine, Busan, Korea Inje University Busan Paik Hospital, Busan, Korea</p> <p>Tahun:</p> <p>2021</p>	<p><u>Hasil penelitian:</u></p> <p>Seiring perkembangan teknologi implementasi <i>machine learning</i> sudah cukup luas, salah satunya dalam mencari hubungan antara dosis warfarin dengan kovariat. Penelitian dilakukan terhadap 650 pasien di korea yang menggunakan obat warfarin dan akan diambil informasi demografis, indikasi, komorbiditas, kebiasaan, dan factor genetic. <i>Dataset</i> akan dibagi menjadi 90% sebagai <i>training set</i> dan 10% sebagai <i>test set</i>. Pengolahan akan menggunakan metode Regresi Linear (LR), dan <i>Gradient Boosting Machine</i> (GBM). Berdasarkan 2 metode yang digunakan mendapatkan hasil akurasi 75,38% untuk LR dan 73,85% untuk GBM dengan nilai korelasi masing-masing sebesar 0,77 dan 0,73. Untuk nilai MAE yang dihasilkan 0,58 mg/hari untuk LR dan 0,64 mg/hari untuk GBM, dengan nilai RMSE sebesar 0,82 mg/hari untuk LR dan 0,9 mg/hari untuk GBM. Dapat disimpulkan bahwa kedua metode tersebut mampu memberikan hasil yang baik, hanya saja LR lebih unggul sedikit dibandingkan GBM.</p> <p><u>Alasan menjadi referensi penelitian:</u></p> <p>Pada penelitian ini menggunakan salah satu metode yang sama yaitu Regresi Linear, dengan tujuan</p>

	<p>mengetahui nilai akurasi, korelasi, dan nilai <i>error</i> yang dihasilkan, dilanjutkan dengan perbandingan hasil dengan metode lainnya untuk mencari metode yang terbaik. Perbedaan penelitian ini yaitu pada bidang dan data yang digunakan.</p>
<p>Judul:</p> <p>A Machine Learning-Based Comparative Approach to Predict the Crop Yield Using Supervised Learning With Regression Models</p> <p>Peneliti:</p> <p>Bharati Panigrahi, Krishna Chaitanya Rao Kathala, M. Sujatha</p> <p>Lokasi:</p> <p>Mukesh Patel School of Technology Management & Engineering (MPSTME), NMIMS University, India</p> <p>University of Massachusetts Amherst, United States of America</p> <p>Jyothishmathi Institute of Technology and Sciences, India</p>	<p><u>Hasil penelitian:</u></p> <p>Penerapan <i>Machine Learning</i> sudah menyebar ke berbagai bidang, salah satunya pada bidang pertanian. <i>Machine Learning</i> akan digunakan untuk melakukan prediksi hasil panen berdasarkan beberapa data yang didapatkan. Tujuan dari penelitian ini yaitu menciptakan model <i>Machine Learning</i> yang mampu memprediksi hasil panen untuk Bengal gram, Groundnut, dan Maize pada daerah Telangana India dari tahun 2016 hingga tahun 2018. Data yang didapatkan berupa penggunaan air, benih, petisida, dan insektisida. Data yang didapatkan akan dilakukan <i>pre-processing</i> data serta akan mengisi data yang hilang dengan nilai median masing-masing kategori, setelah itu data akan dipisahkan menjadi 80% data <i>training</i> dan 20% data <i>test</i>. Penelitian ini menggunakan 6 metode regresi yaitu: <i>Linear Regression</i>, <i>Decision Tree Regression</i>, <i>Gradient Boosting Regression</i>, <i>Random Forest Regression</i>, <i>Xgboost Regression</i>, dan <i>Voting Regression</i>. Dari model-model tersebut akan dicari nilai MAE, MSE, R2 Score, dan <i>Cross-validation Score</i>, sehingga mampu menentukan metode yang akan dijadikan model <i>machine learning</i>. Berdasarkan hasil penelitian maka didapatkan metode <i>Random Forest Regression</i> merupakan</p>

<p>Tahun: 2023</p>	<p>metode yang terbaik untuk dijadikan model dengan nilai MAE 468,16 dan <i>Cross-validation Score</i> sebesar 0,6087.</p> <p><u>Alasan menjadi referensi penelitian:</u></p> <p>Pada penelitian tugas akhir ini menggunakan 2 metode yang sama dengan jurnal, yaitu <i>Linear Regression</i> dan <i>Random Forest Regression</i>. Dengan tujuan melakukan prediksi berdasarkan hasil korelasi yang dihasilkan dan akan dibandingkan berdasarkan nilai MAE, <i>R2 Score</i> untuk menentukan metode yang terbaik. Perbedaan dalam penelitian tugas akhir ini yaitu menggunakan data getaran <i>gear box</i> mesin kertas.</p>
<p>Judul: Modelling and forecasting new cases of Covid-19 in Nigeria: Comparison of regression, ARIMA and machine learning models</p> <p>Peneliti: S.I. Busari, T.K. Samson</p> <p>Lokasi: Statistics Programme, College of Agriculture, Engineering and Science, Bowen University, Iwo, Osun State, Nigeria</p>	<p><u>Hasil penelitian:</u></p> <p>Perkembangan dalam implementasi <i>Machine Learning</i> sudah sangat luas, salah satunya pada bidang kesehatan. Pada penelitian yang dilakukan pada jurnal ini, <i>Machine Learning</i> digunakan untuk meramalkan kasus penyebaran Covid-19 di negara Nigeria. Data yang digunakan yaitu data penyebaran Covid-19 dimulai pada tanggal 27 Februari 2020 hingga 30 November 2021. Metode <i>Machine Learning</i> yang digunakan yaitu cukup bervariasi yaitu menggunakan regresi dengan 4 model berbeda, metode ARIMA dengan 6 model berbeda, dan 10 model <i>Machine Learning</i> yang berbeda. Dalam menentukan model yang terbaik dapat dilakukan perbandingan dengan melihat hasil MSE dan RMSE, pilih yang menghasilkan nilai terkecil. Pada penelitian pada jurnal ini model yang terbaik pada</p>

<p>Tahun:</p> <p>2022</p>	<p><i>Machine Learning</i> model <i>Fine Tree</i> dengan nilai MSE dan RMSE sebesar 0,04913 dan 0,22165.</p> <p><u>Alasan menjadi referensi penelitian:</u></p> <p>Berdasarkan jurnal yang didapatkan ini metode dan tujuan yang ingin dicapai memiliki kesamaan yaitu pada penggunaan regresi untuk melakukan prediksi kedepannya serta melakukan perbandingan terhadap model yang terbaik berdasarkan nilai <i>error</i> terkecil. Perbedaan pada penelitian tugas akhir ini yaitu implementasi <i>Machine Learning</i> dan jenis data yang digunakannya.</p>
<p>Judul:</p> <p>Machine learning-based multi-target regression to effectively predict turning movements at signalized intersections</p> <p>Peneliti:</p> <p>Khaled Shaaban, Ali Hamdi, Mohammad Ghanim, Khaled Bashir Shaban</p> <p>Lokasi:</p> <p>Department of Engineering, Utah Valley University, Orem, United States</p>	<p><u>Hasil penelitian:</u></p> <p>Implementasi <i>Machine Learning</i> pada penelitian dalam jurnal ini yaitu ingin melakukan prediksi pergerakan kendaraan berbelok pada suatu persimpangan. Dalam melakukan prediksi membutuhkan pengumpulan data, kalibrasi, dan upaya pemodelan yang ekstensif. Penelitian ini akan menggunakan 3 model yaitu <i>Multiple Linear Regression</i>, <i>Random Forest Regression</i>, dan <i>Artificial Neural Network</i>. Untuk memilih model yang terbaik maka perlu dilakukan perbandingan nilai MSE, <i>R-Squared</i>, RMSE, dan MAPE. Sebelum memulai pengolahan data, akan dilakukan <i>Cross-validation</i> sebanyak 10 <i>fold</i>, untuk memastikan bahwa data yang digunakan sudah <i>valid</i>. Setelah dilakukan pengolahan data, maka dapat diketahui bahwa model <i>Artificial Neural Networks</i> dengan mendapatkan hasil MSE, <i>R-Squared</i>, RMSE, dan</p>

<p>Faculty of Engineering and IT, RMIT University, Melbourne, Australia</p> <p>Ministry of Transport, Doha, Qatar</p> <p>Computer Science and Engineering Department, Qatar University, Doha, Qatar</p> <p>Tahun: 2022</p>	<p>MAPE secara berturut-turut adalah 1514, 92, 38.9, dan 59.6. Maka dengan ini, penggunaan model <i>Artificial Neural Networks</i> dapat digunakan untuk melakukan prediksi dengan data yang baru.</p> <p><u>Alasan menjadi referensi penelitian:</u></p> <p>Pada jurnal penelitian ini memiliki kesamaan dalam tujuan, model yang digunakan, dan cara untuk melakukan perbandingan model yang terbaik dengan penelitian tugas akhir ini. Model yang digunakan yaitu <i>Multiple Linear Regression</i> dan <i>Random Forest Regression</i>. Cara untuk memilih model yang terbaik dapat dilakukan dengan cara membandingkan nilai <i>error</i> yang dihasilkan masing-masing model. Perbedaannya yaitu pada penelitian tugas akhir ini data yang digunakan merupakan data getaran <i>gearbox</i> pada mesin kertas.</p>
<p>Judul:</p> <p>Application of supervised machine learning to predict the enhanced gas recovery by CO2 injection in shale gas reservoirs</p> <p>Peneliti:</p> <p>Moataz Mansi, Mohamed Almobarak, Jamiu Ekundayo, Christopher Lagat, Quan Xie</p> <p>Lokasi:</p>	<p><u>Hasil penelitian:</u></p> <p>CO2-EGR adalah proses geofisika kompleks yang dikendalikan oleh beberapa parameter <i>Shale Properties</i> dan <i>Engineering Desain</i>. Pada jurnal penelitian ini memiliki tujuan yaitu ingin menilai efisiensi CO2-EGR dengan cara melakukan prediksi peningkatan bertahap CH4 yang dihasilkan CO2 menggunakan <i>Machine Learning</i>. <i>Dataset</i> yang didapatkan pada jurnal ini cukup banyak, tetapi sebelum melakukan pengolahan data, <i>dataset</i> akan dilakukan eliminasi data dengan melihat hasil korelasi <i>Spearman</i>. Setelah eliminasi data dilakukan maka data yang tersisa akan dijadikan <i>feature</i> dalam</p>

<p>Discipline of Petroleum Engineering, WA School of Mines: Minerals, Energy and Chemical Engineering, Curtin University, Western Australia, Australia</p> <p>Tahun: 2023</p>	<p>melakukan prediksi. Model <i>Machine Learning</i> yang digunakan yaitu <i>Linear Regression</i> dan <i>Artificial Neural Network</i> (ANN). Pemilihan model terbaik dilihat dari hasil <i>R-Squared</i> masing-masing model. Untuk ANN memiliki 3 model yang berbeda, ini dilakukan karena jumlah <i>hidden layer</i> yang dibuat berbeda-beda untuk mendapatkan hasil yang terbaik. Berdasarkan hasil pengolahan data yang didapatkan, maka ANN dengan 100 <i>hidden layer</i> memberikan hasil <i>R-Squared</i> terbaik dengan nilai 0.78, jika dibandingkan dengan hasil <i>Linear Regression</i> mendapatkan nilai sebesar 0.68. Dengan demikian perhitungan prediksi peningkatan bertahap CH₄ yang dihasilkan CO₂ dilakukan dengan menggunakan model ANN 100 <i>hidden layer</i>.</p> <p><u>Alasan menjadi referensi penelitian:</u></p> <p>Pemilihan jurnal penelitian ini dipilih sebagai referensi dikarenakan penggunaan model yang sama yaitu <i>Linear Regression</i> dan memiliki tujuan yang mirip yaitu melakukan prediksi dengan membandingkan 2 model <i>Machine Learning</i>. Cara melakukan pemilihan model yang terbaik juga hampir sama yaitu dengan melihat hasil <i>R-Squared</i> dari masing-masing model. Perbedaan jurnal ini dengan penelitian tugas akhir ini yaitu pada jenis <i>dataset</i> yang digunakannya.</p>
--	---

1.6 Sistematika Penulisan

Adapun sistematika penulisan dalam menyusun tugas akhir yaitu sebagai berikut:

BAB I = PENDAHULUAN

Pada BAB I, membahas mengenai latar belakang penulisan, tujuan kegiatan penelitian, manfaat penelitian, dan sistematika penulisan tugas akhir.

BAB II = TINJAUAN PUSTAKA

Pada BAB II, membahas dan menjelaskan teori yang dilakukan berdasarkan metode yang digunakan yaitu Regresi Linear Berganda, Regresi *Random Forest*, dan Regresi *Support Vector Machine*.

BAB III = METODE PENELITIAN

Pada BAB III, membahas pengumpulan data, jenis data, dan pembuatan diagram alir atau *Flow Chart* sebagai pemahaman pada data yang akan digunakan dan diolah untuk menentukan hasil yang ingin dicapai.

BAB IV = PEMBAHASAN DAN ANALISIS HASIL

Pada BAB IV, membahas cara pengolahan data dilakukan dengan Regresi Linear Berganda, dan Regresi *Random Forest*. Untuk mengetahui hasil korelasi yang didapatkan serta melakukan prediksi berdasarkan metode yang digunakannya.

BAB V = KESIMPULAN DAN SARAN

Pada BAB V, membahas kesimpulan berdasarkan penelitian yang telah dilaksanakan, dan memberikan saran berdasarkan pengalaman dan hasil analisa yang didapatkan.