

BAB II LANDASAN TEORI

2.1 Penelitian Terdahulu

Tabel 2. 1 Penelitian Terdahulu

Judul Jurnal	Nama Jurnal	Penulis/Tahun	Algoritma	Kesimpulan
Experimental measurement <i>and modeling</i> of water-based drilling mud density using adaptive boosting <i>Decision Tree</i> , support vector machine, <i>and K</i> -nearest neighbors: A case study from the South Pars gas field	Journal of Petroleum Science <i>and Engineering</i>	Abbas Hashemizadeh, Ph.D., Ahmad Maaref, Mohammadhadi Shateri, Aydin Larestani, Abdolhossein Hemmati-Sarapardeh, Ph.D. / 2021	<i>Decision Tree</i> , Support Vector Machine, dan K-Nearest Neighbors	Penelitian ini bertujuan memprediksi ketebalan lumpur pengeboran berdasarkan berbagai faktor seperti kedalaman, ukuran lubang, sudut kemiringan, dan sifat rheologi lumpur. Para peneliti menggunakan 815 data lapangan aktual untuk melatih lima model kecerdasan buatan. Model terbaik, yaitu ABR-DT, mampu memprediksi ketebalan lumpur dengan tingkat akurasi rata-rata 99,48% (AAPRE 0.5159%). Analisis sensitivitas menunjukkan faktor paling berpengaruh terhadap ketebalan lumpur adalah kekentalan plastis dan kedalaman lubang bor (kedalaman sejati vertikal). Sebaliknya, ukuran lubang bor, jenis lumpur, dan kehilangan cairan lumpur memiliki hubungan terbalik dengan ketebalan lumpur. Terakhir, analisis validitas data memastikan data dan model yang dikembangkan dapat diandalkan.

Judul Jurnal	Nama Jurnal	Penulis/Tahun	Algoritma	Kesimpulan
Analisis Kompleksitas Password Dengan Metode KNN, Naïve Bayes, <i>Decision Tree</i> , Ensemble Methods Dan <i>Linear Regression</i>	Digital Transformation Technology (Digitech)	Eri Mardiani, Nur Rahmansyah, Yunan Fauzi Wijaya, Annisa Amalia Fitri, Rayhan Mustafa, Muhammad Romadhoni Rizki, Komang Mustika Pramesti / 2023	KNN, Naïve Bayes, <i>Decision Tree</i> , Ensemble Methods dan <i>Linear Regression</i>	Pada penelitian ini dapat disimpulkan bahwa bahwa model-model tersebut mampu mencapai tingkat akurasi dan ketepatan prediksi yang tinggi. <i>Random Forest</i> dinilai memiliki kinerja tertinggi, diikuti oleh Naive Bayes, AdaBoost, Tree, dan k-NN. Oleh karena itu, <i>Random Forest</i> dianggap sebagai model algoritma paling akurat dalam studi ini, sementara <i>Linear Regression</i> tidak sesuai digunakan dalam <i>dataset</i> yang sama.
Analisis Pola Prediksi Data Time Series menggunakan <i>Support Vector Regression</i> , Multilayer Perceptron, dan Regresi Linear Sederhana	Jurnal Resti (Rekayasa Sistem dan Teknologi Informasi)	Ika Oktavianti, Ermatita, Dian Palupi Rini / 2019	Support Vector Regression, Multilayer Perceptron, dan Regresi Linear Sederhana	Penelitian ini memanfaatkan data time series dengan menggunakan <i>Support Vector Regression</i> (SVR) dan Multilayer Perceptron (MLP) untuk memprediksi pola perizinan. Hasil studi menunjukkan bahwa ketika jumlah data time series yang digunakan cukup besar, Metode MLP terbukti lebih akurat dibandingkan dengan SVR, dengan nilai akurasi -0.17 serta Mean Absolute Error (MAE) dan Root Mean Square Error (RMSE) sebesar 11.45 dan 15.84 secara berturut-turut. Namun, ketika data time series yang digunakan sedikit, metode SVR memberikan hasil prediksi yang lebih akurat daripada MLP, dengan akurasi -0.014 dan nilai MAE serta RMSE berturut-turut 32.80 dan 42.89. Dalam keseluruhan, SVR dan MLP-ANN terbukti sebagai pilihan yang baik untuk meramalkan pola perizinan, terutama

Judul Jurnal	Nama Jurnal	Penulis/Tahun	Algoritma	Kesimpulan
				dengan dukungan data time series yang besar guna mendapatkan prediksi yang lebih tepat. Selain itu, melalui hasil peramalan menggunakan Simple <i>Linear Regression</i> (SLR), disimpulkan bahwa diperlukan penambahan petugas pelayanan untuk meningkatkan kualitas layanan perizinan dan Indeks Kepuasan Masyarakat (IKM).
Analisis Prediksi Pendapatan Penduduk dengan Metode K-Nearest Neighbor, <i>Decision Tree</i> , Naive Bayes, Ensemble Methods, dan <i>Linear Regression</i>	INNOVATIVE: Journal Of Social Science Research	Eri Mardiani, Nur Rahmansyah, Endah Tri Esti Handayani, Sari Ningsih, Deny Hidayatullah, Dhieka Avrilia Lantana, Yuni Latifah, Alica Dwi Fahira, Keysha Belynda Tyva Panggabean, Imelta Natalia Ginting / 2023	K-Nearest Neighbor, <i>Decision Tree</i> , Naive Bayes, Ensemble Methods, dan <i>Linear Regression</i>	Dalam konteks ini, prediksi digunakan untuk mengolah informasi dari <i>dataset</i> yang besar, yang memerlukan pendekatan data mining. Data mining merupakan gabungan berbagai teknik dari <i>Machine Learning</i> , pengenalan pola, statistik, basis data, dan visualisasi untuk menggali informasi bernilai dari basis data yang luas. Pada penelitian ini algoritma K-NN memperoleh nilai AUC sebesar 94,4%, Naïve Bayes memperoleh nilai AUC sebesar 90,3%, <i>Decision Tree</i> dan Ensemble Tree memperoleh nilai AUC sebesar 71.3%, <i>Random Forest</i> memperoleh nilai AUC sebesar 89,6%, AdaBoost memperoleh nilai AUC sebesar 85%, dan <i>Linear Regression</i> memperoleh nilai MSE sebesar 6.618.
Implementation of <i>Decision Tree</i> Algorithm C4.5 and <i>Support Vector Regression</i> for Stroke Disease Prediction	MALCOM: Indonesian Journal of <i>Machine Learning and Computer Science</i>	Firman Akbar, Hanif Wira Saputra, Adhitya Karel Maulaya, Muhammad	<i>Decision Tree</i> dan <i>Support Vector Regression</i>	Penelitian ini dilakukan pada 5110 <i>dataset</i> pasien stroke dari Kaggle Open <i>Datasets</i> menggunakan pendekatan splitting data 70:30 dengan algoritma <i>Support Vector</i>

Judul Jurnal	Nama Jurnal	Penulis/Tahun	Algoritma	Kesimpulan
		Fikri Hidayat, Rahmaddeni / 2022		<i>Regression</i> dan <i>Decision Tree</i> C4.5 menunjukkan bahwa <i>Decision Tree</i> C4.5 secara signifikan lebih unggul dalam memprediksi <i>dataset</i> tersebut. Hasil penelitian menunjukkan bahwa nilai root mean square error (RMSE) dari <i>Decision Tree</i> C4.5 adalah 0.235, sementara algoritma <i>Support Vector Regression</i> menghasilkan RMSE sebesar 0.3999. Nilai RMSE yang lebih rendah, seperti yang diperoleh dari <i>Decision Tree</i> C4.5, menandakan ketepatan prediksi yang lebih baik, sehingga dapat disimpulkan bahwa algoritma ini efektif dalam memprediksi kejadian stroke pada pasien.
Perbandingan Metode Algoritma <i>Support Vector Regression</i> dan <i>Multiple Linear Regression</i> Untuk Memprediksi Stok Obat	Journal of Applied Technology and Informatics	Lit Malem Ginting, Marojahan MT. Sigiyo, Esra Delima Manurung, Juan Jasa Putra Sinurat / 2021	<i>Support Vector Regression</i> dan <i>Multiple Linear Regression</i>	Hasil penelitian menunjukkan bahwa <i>Machine Learning</i> menggunakan algoritma <i>Multiple Linear Regression</i> (MLR) dan <i>Support Vector Regression</i> (SVR) dapat efektif digunakan untuk memprediksi stok obat. Dalam eksperimen dengan data linear, diketahui bahwa <i>Multiple Linear Regression</i> memiliki nilai Mean Squared Error (MSE) yang lebih kecil daripada <i>Support Vector Regression</i> , dengan MLR memiliki MSE sebesar $1.4134835564042063e-29$ dan SVR sebesar 0.001357677202647385 . Kesimpulan dari penelitian ini adalah bahwa untuk jenis data linear, MLR lebih

Judul Jurnal	Nama Jurnal	Penulis/Tahun	Algoritma	Kesimpulan
				disukai daripada SVR dalam proses pemodelan.
Perbandingan Prediksi Harga Saham Menggunakan Metode SVR, RFR, dan DTR	Prosiding Pendidikan Matematika dan Matematika	Royan Agil Nugroho, St. Rahmawati, Diyah Kresna Eni A / 2022	<i>Decision Tree</i> , <i>Random Forest</i> , dan <i>Support Vector Regression</i>	Penelitian ini menggunakan tiga metode regresi, yaitu <i>Support Vector Regression</i> (SVR), <i>Random Forest Regression</i> (RFR), dan <i>Decision Tree Regression</i> (DTR), karena metode-metode tersebut dianggap sesuai untuk memprediksi harga saham. Kemudian, ketiga metode ini diterapkan pada <i>dataset</i> saham Kimia Farma Tbk untuk analisis. <i>Dataset</i> tersebut berisi informasi tentang fluktuasi harga saham perusahaan Kimia Farma Tbk setiap harinya. Tujuan akhir dari penelitian ini adalah untuk membandingkan tingkat keakuratan ketiga metode tersebut. Hasil penelitian menunjukkan bahwa model prediksi dengan akurasi terbaik adalah <i>Random Forest Regression</i> (RFR), dengan skor Mean Squared Error (MSE)-test sebesar 13443.61 dan R2 <i>score</i> -test sebesar 0.19.
<i>Predicting the Depression Level of Excessive Use of Mobile Phone: Decision Tree and Linear Regression Algorithm</i>	Patuakhali Science and Technology	Imrus Salehin, Iftakhar Mohammad Talha, Nazmun Nessa Moon, Mohd. Saifuzzaman, Fernaz Narin Nur, Mariom Akter / 2021	<i>Decision Tree</i> , <i>Naïve Bayes</i> dan <i>Linear Regression</i>	Penelitian ini mengonfirmasi bahwa salah satu penyebab depresi adalah penggunaan berlebihan ponsel. Semakin banyak waktu yang dihabiskan dengan ponsel, semakin banyak stres yang diciptakan dalam otak, yang kemudian menyebabkan depresi. Meskipun perkembangan sektor Teknologi Informasi dan Komunikasi (ICT) membuka horison baru bagi kita,

Judul Jurnal	Nama Jurnal	Penulis/Tahun	Algoritma	Kesimpulan
				remaja dan dewasa muda menjadi rentan terhadap teknologi baru ini. Depresi dapat menjadi penyebab pemikiran bunuh diri atau perasaan ketidakmampuan, kekacauan pribadi atau konflik dengan keluarga, kekerasan fisik, masalah tidur, nyeri kronis, kecemasan, dll. Hal ini karena remaja memiliki risiko depresi yang lebih tinggi, ini membuatnya lebih berbahaya. Frustrasi tidak memungkinkan mereka untuk maju dalam hidup. Mereka akan tertinggal. Kami akan membahas secara singkat dalam karya masa depan kami, bagaimana mengatasi masalah serius ini dan bagaimana kita dapat bertahan dengan mudah melalui situasi ini. Pada penelitian ini algoritma <i>Decision Tree</i> mendapatkan akurasi sebesar 86%, Naïve Bayes mendapatkan akurasi sebesar 81% dan <i>Linear Regression</i> mendapatkan akurasi sebesar 65%.
Stock Market Analysis Using <i>Linear Regression and Decision Tree Regression</i>	Chongqin University	Rezaul Karim, Md Khorshed Alam, Md Rezaul Hossain / 2021	<i>Linear Regression</i> dan <i>Decision Tree</i>	Penelitian ini hanya menggunakan dua algoritma regresi. Tujuan utama dari penelitian ini adalah untuk menggunakan algoritma <i>Machine Learning</i> berbasis supervisi, yaitu <i>Linear Regression</i> dan <i>Decision Tree Regression</i> , untuk memprediksi harga saham. Hasil penelitian menunjukkan bahwa <i>Linear Regression</i> memberikan akurasi yang lebih baik untuk kedua jenis <i>dataset</i> , baik

Judul Jurnal	Nama Jurnal	Penulis/Tahun	Algoritma	Kesimpulan
				yang kecil maupun besar. Di sisi lain, <i>Decision Tree</i> Regression menunjukkan prediksi harga yang kurang akurat tergantung pada ukuran <i>dataset</i> yang digunakan.
Perbandingan Prediksi Kemiskinan di Indonesia Menggunakan Support Vector Machine (SVM) dengan Regresi Linear	Jurnal Sains dan Matematika Statistika	Abdul Karim / 2020	Support Vector Machine (SVM) dan <i>Linear Regression</i>	Penelitian ini menegaskan bahwa Support Vector Machine (SVM) merupakan salah satu metode yang efektif, terutama untuk data yang memiliki distribusi yang tidak diketahui. SVM menggunakan fungsi kernel yang memungkinkannya untuk menangani berbagai jenis data. Melalui tahap <i>plotting</i> data, ditemukan bahwa sebaran data membentuk pola linier. Hasil perbandingan antara SVM dan regresi linear menunjukkan bahwa <i>plotting</i> titik yang diprediksi oleh model SVM hampir sempurna mengikuti pola data yang sebenarnya. Dari hasil ini, disimpulkan bahwa dalam memprediksi kemiskinan di Indonesia, SVM memberikan hasil yang lebih baik daripada regresi linear.

Berdasarkan 10 penelitian terdahulu di atas yang mana membandingkan algoritma regresi didalamnya dapat terlihat bahwa setiap model algoritma memiliki kemampuannya masing masing sesuai dengan *dataset* yang diuji. Pada penelitian [5],[6],[7],[8] didapatkan hasil bahwa algoritma *Decision Tree* merupakan algoritma dengan tingkat prediksi yang paling baik dibandingkan oleh algoritma lainnya. Sementara pada penelitian [9],[10] algoritma yang memiliki performa terbaik adalah *Linear Regression* dengan hasil akurasi yang cukup tinggi. Disisi lain, *Support Vector Regression* juga mendapatkan hasil yang tidak kalah baik dari *Decision Tree* dan *Linear Regression* dimana pada penelitian [11],[12] SVR memperoleh tingkat akurasi yang lebih baik dibandingkan dengan *Linear Regression* dan *Decision Tree*. Berdasarkan hal ini, dapat dikatakan bahwa ketiga algoritma disini memiliki performa yang baik tetapi harus disesuaikan lagi kedalam *dataset* yang akan diuji. *Decision Tree* sendiri sangat sering mendapatkan performa terbaik dalam hal ini namun *Linear Regression* dan *Support Vector Regression* juga tidak dapat dikatakan buruk karena dengan uji coba *dataset* lain kedua algoritma ini seringkali mendapatkan performa yang baik juga.

Penelitian ini memiliki berbagai perbedaan dari penelitian terdahulu yang digunakan yaitu perbedaan pada *dataset* dan algoritma yang digunakan. Penelitian ini menggunakan *dataset* pemain *NBA* sebagai acuan untuk memprediksi siapa pemain yang dapat memperoleh gelar *MVP* pada musim 2022-2023. Pada penelitian terdahulu jurnal keempat menggunakan Neural dan Model Regresi untuk memprediksi hasil olahraga sedangkan pada penelitian ini digunakan Model Regresi dengan tiga algoritma berbeda yaitu *Decision Tree*, *Linear Regression*, *Support Vector Regression*. Pada penelitian ini variabel yang digunakan untuk mengukur pemain dalam mendapatkan gelar *MVP* adalah *Fantasy points*. Hal ini dilakukan karena *Fantasy points* merupakan sebuah nilai statistik yang didapatkan dari setiap pemain dari seluruh statistik mereka dalam beberapa kategori.

2.2 Olahraga Basket

Olahraga Basket merupakan salah satu olahraga yang digemari oleh seluruh kalangan umur mulai dari anak kecil hingga orang dewasa. Permainan bola basket disini merupakan cabang olahraga yang mana dimainkan oleh dua regu didalam pertandingannya. Setiap regu atau tim disini terdiri dari 5 orang pemain didalamnya. Permainan bola basket disini termasuk kedalam tipe olahraga bola besar yang mana para pemain harus memasukan bola ke dalam ring untuk mendapatkan poin (skor). [13]

Bola basket pertama kali ditemukan oleh Dr. James Naismith yang mana merupakan anggota dari sekolah pelatihan *Young Mans Christian Association* (YMCA) di Springfield, Massachusetts pada Desember tahun 1891. Untuk dapat memainkan olahraga bola basket dengan benar diperlukan teknik yang harus dikuasai terlebih dahulu. Menurut Kosasih (1993) teknik dasar permainan bola basket terdiri dari beberapa jenis antara lain: [14]

1. *Dribbling*

Dribbling disini merupakan gerakan membawa bola yang sesuai dengan peraturan dalam permainan bola basket. Teknik ini dilakukan dengan cara memantulkan bola basket ke lantai dengan satu tangan atau dua tangan secara bergantian. *Dribbling* disini merupakan teknik dasar yang penting untuk dikuasai saat bermain basket karena tidak hanya digunakan untuk membawa atau menggiring bola saja tetapi juga digunakan untuk mengamankan bola dari lawan dengan cara melakukan berbagai variasi *dribbling* yang dapat membantu pemain untuk melewati lawan.

2. *Passing*

Passing (Operan) merupakan Gerakan melempar bola ke sesama regu atau biasa dikatakan mengoper bola. Teknik ini dilakukan dengan cara melakukan lemparan yang terarah kepada teman satu regu. *Passing* dapat dilakukan dengan menggunakan satu tangan ataupun dua tangan. Teknik *passing* disini memiliki

beberapa tipe, berikut merupakan tipe passing yang paling sering digunakan didalam permainan bola basket

- *Bounce Pass* (Operan Pantul)
- *Chest Pass* (Operan setinggi dada)
- *Overhead Pass* (Operan di atas kepala)

3. *Catching*

Catching atau menangkap merupakan teknik yang harus dikuasai untuk dapat mengambil atau menangkap bola dari team atau lawan. Teknik catching dapat dilakukan dengan menggunakan satu ataupun dua tangan. Pemain basket harus menguasai teknik catching agar dapat menerima operan dari team dan juga melakukan transisi dengan optimal.

4. *Shooting*

Shooting atau menembak adalah teknik yang digunakan untuk memasukan bola kedalam ring basket. Teknik ini berguna untuk memperoleh skor. Teknik satu ini harus dapat dikuasai oleh setiap pemain basket karena dengan menguasai teknik shooting yang baik maka pemain akan lebih mudah untuk mencetak skor dan memenangkan pertandingan. Dalam pertandingan bola basket terdapat garis perimeter yang sudah ditentukan dan pemain yang memasukan bola didalam garis perimeter akan mendapatkan 2 poin sedangkan untuk pemain yang menembak dan memasukan bola diluar garis perimeter yang ditentukan akan mendapatkan skor 3 poin (*3point shoot*).

5. *Rebound*

Rebound merupakan sebuah teknik yang digunakan untuk mengambil atau meraih bola yang dilemparkan oleh tim ataupun lawan. Rebound disini dilakukan dengan cara merebut bola yang terpantul karena tim atau lawan tidak berhasil memasukan bola ke dalam ring. Rebound disini dibutuhkan untuk dapat mendapatkan kembali kekuasaan terhadap bola dan dapat menciptakan peluang

baru untuk mencetak skor. Rebound terdiri dari 2 macam yaitu offensive rebound dan defensive rebound.

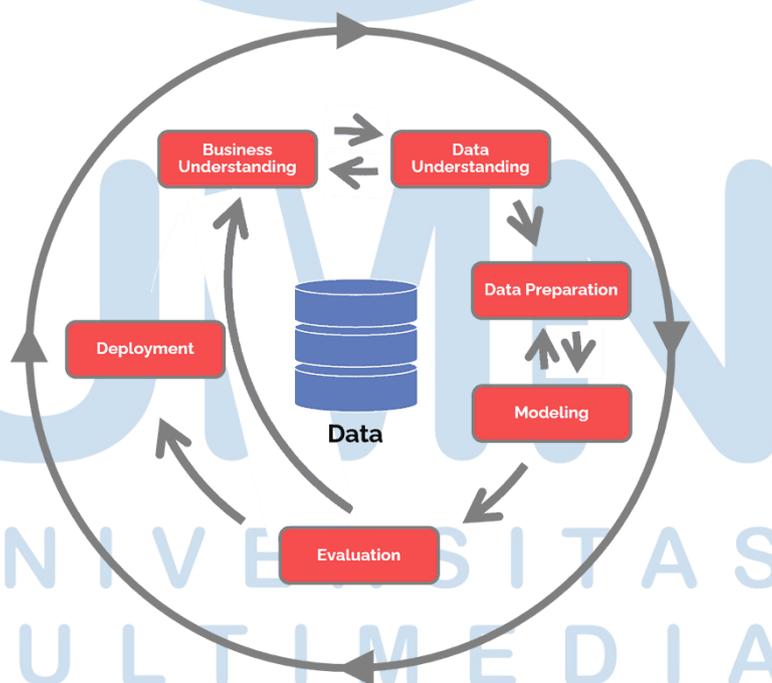
6. *Pivot*

Pivot merupakan sebuah teknik yang dilakukan untuk melakukan pergerakan atau perubahan arah badan tanpa melakukan dribbling. Hal ini dilakukan dengan cara mengubah arah badan ke segala arah dengan satu kaki bergerak dan satu kaki menetap di posisi sebagai poros.

2.3 Algoritma dan Framework

2.3.1 CRISP-DM

CRISP-DM merupakan sebuah model proses independent yang digunakan dan dimanfaatkan dalam proses *data mining*. *CRISP-DM* sendiri terdiri dari 6 tahapan yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*. [15]



Gambar 2. 1 CRISP - DM

2.3.1.1 Business Understanding

Business Understanding adalah tahapan pertama yang ada didalam *CRISP-DM*. Tahapan ini fokus pada proses pemahaman pada suatu bisnis dan penggunaan data mining didalamnya mulai dari pemahaman mengenai rencana, tujuan, hingga jenis data mining yang akan digunakan.

2.3.1.2 Data Understanding

Data Understanding merupakan tahapan kedua yang ada didalam *CRISP-DM*. Tahap ini berisikan proses pemahaman pada data yang akan digunakan dalam melakukan penelitian. Didalam tahap ini proses pengumpulan data akan dilakukan. Selain itu, Data juga akan dikelola dan diperiksa kualitasnya.

2.3.1.3 Data Preparation

Data Preparation merupakan tahap yang memiliki fokus dalam menyiapkan data menjadi data yang siap untuk digunakan. Dalam tahapan ini data akan disiapkan secara bertahap mulai dari pembersihan data (*data cleansing*) hingga melakukan format ulang pada data untuk mendapatkan data yang siap digunakan dan tentunya memiliki kualitas yang baik.

2.3.1.4 Modeling

Modeling merupakan tahap dilakukannya pemilihan teknik pemodelan yang akan digunakan serta melakukan proses pembentukan model menggunakan algoritma yang sudah ditentukan hingga melakukan penilaian pada model (*assess model*) dengan cara melakukan perbandingan antara beberapa jenis model yang akan digunakan.

2.3.1.5 Evaluation

Evaluation merupakan tahap dilakukannya evaluasi pada model yang sudah dibentuk dalam tahapan *modeling*. Dalam tahap evaluasi ini

model akan diulas secara lebih terperinci. Hal ini dilakukan agar dapat menilai dan melihat model mana yang paling sesuai dengan bisnis dan tujuan yang dibutuhkan.

2.3.1.6 Deployment

Deployment merupakan tahapan terakhir didalam *CRISP-DM*. Tahapan ini berisikan proses akhir seperti penerapan model yang sudah dibuat dan juga melakukan pemantauan (*monitoring*) serta pemeliharaan secara menyeluruh untuk mendapatkan hasil yang maksimal serta mengurangi resiko terjadinya masalah selama melakukan tahap operasional.

2.3.2 Machine Learning

Machine Learning merupakan sebuah teknologi yang dikembangkan untuk dapat belajar dan mengembangkan wawasannya sendiri tanpa membutuhkan arahan dari pengguna. *Machine Learning* disini sudah mulai berkembang secara pesat dimana *Machine Learning* sudah menjadi pilihan untuk melakukan pengembangan sebuah *software* untuk fitur pengenalan suara, pemrosesan bahasa alami (*natural language processing*), pengontrolan pada robot, dan aplikasi lainnya yang berhubungan dengan teknologi yang advance. [16]

Istilah *Machine Learning* pertama kali dibuat oleh beberapa ilmuwan matematika seperti Thomas Bayes, Adrein Marie Legendre, dan Andrey Markow. Istilah ini dikemukakan pada tahun 1920an yang mana disertai dengan konsep dan dasar dari *Machine Learning* itu sendiri. [17] *Machine Learning* pada umumnya digunakan untuk memberikan kemampuan pada mesin untuk menangani data dengan lebih efisien. *Machine Learning* memiliki 2 tipe pembelajaran yaitu:

1. Supervised Learning

Supervised Learning merupakan tipe pembelajaran *Machine Learning* yang memerlukan data pelatihan *input* dan *output*. Tipe pembelajaran ini biasanya

digunakan untuk melakukan regresi pada data yang tidak terlihat ke dalam kategori yang ditentukan dan juga melakukan prediksi. [5]

2. *Unsupervised Learning*

Unsupervised Learning merupakan tipe pembelajaran *Machine Learning* yang dapat memproses data mentah atau data yang tidak memiliki label (unlabelled data). Tipe pembelajaran ini biasanya digunakan untuk melakukan identifikasi pola dan juga tren yang ada didalam kumpulan data mentah. Tipe pembelajaran ini juga seringkali digunakan untuk mengelompokkan data kedalam grup tertentu. [5]

2.3.3 **Decision Tree**

Decision Tree adalah algoritma regresi terawasi yang menggunakan seperangkat aturan untuk membentuk struktur seperti pohon. Algoritma *Decision Tree* adalah bagian dari keluarga algoritma pembelajaran terawasi. Algoritma *Decision Tree* mirip dengan proses pengambilan keputusan manusia dan itu membuat algoritma ini mudah dipahami. Algoritma ini dapat memecahkan dalam kedua situasi apakah seseorang memiliki data diskrit atau kontinu sebagai input.

Decision Tree adalah jenis metode statistik yang dilakukan melalui representasi grafis dari proses pengambilan keputusan dalam kondisi tertentu. Biasanya digunakan untuk menentukan apakah keputusan yang dibuat berdasarkan data yang diberikan sudah optimal atau belum. *Decision Tree* memiliki banyak aplikasi di hampir setiap area di mana pilihan harus dibuat dari alternatif yang tersedia. Selain industri manufaktur, aplikasi ini mencakup masalah sosial, pengaturan bisnis, dan industri jasa seperti perbankan, perpustakaan dan rumah sakit, industri farmasi, semua di semua perusahaan kecil dan besar. [18]

2.3.4 Linear Regression

Linear Regression merupakan algoritma yang seringkali digunakan untuk data science dan merupakan sebuah algoritma supervised learning. Algoritma *Linear Regression* sendiri menggunakan prinsip regresi yang mana digunakan untuk membuat sebuah model prediksi untuk target variabel yang didasari oleh variabel lainnya atau variabel bebasnya. Algoritma ini seringkali digunakan untuk mencari korelasi atau hubungan antar variabel yang ada dan menghasilkan sebuah prediksi yang akurat. Tujuan utama dari penggunaan algoritma ini adalah untuk melakukan sebuah prediksi pada variabel terikat (Y) berdasarkan variabel bebas yang diberikan (X). Prediksi ini akan menampilkan hasil akhir berupa model hubungan linear antara variabel input / independen (X) dengan variabel output / dependen (Y). Pada akhirnya akan ditemukan garis lurus terbaik yang dapat menggambarkan hubungan antara variabel-variabel yang digunakan. Dalam *Linear Regression*, garis lurus akan diwakili oleh persamaan matematika seperti $Y = \beta_0 + \beta_1 X + \epsilon$. Pada persamaan ini Y merupakan variabel dependen atau dapat diartikan sebagai variabel yang ingin diprediksi hasilnya, X adalah variabel independen, β_0 adalah intercept (nilai Y ketika $X = 0$), β_1 merupakan koefisien regresi yang menunjukkan perubahan didalam Y yang mana diakibatkan oleh perubahan satu unit pada X, kemudian ϵ merupakan kesalahan acak. [19]

Algoritma *Linear Regression* disini merupakan salah satu algoritma paling umum dan populer untuk digunakan dalam analisis regresi. Hal ini dikarenakan oleh sifatnya yang sederhana, kemampuan dalam memberikan prediksi yang cukup baik dan akurat dalam berbagai studi kasus, dan juga interpretasi yang mudah.

2.3.5 Support Vector Regression

Support Vector Regression merupakan sebuah algoritma didalam *Machine Learning* yang digunakan untuk membuat sebuah model hubungan antar variabel dependen dan satu atau lebih variabel independen. Tujuan dari

digunakannya algoritma ini adalah untuk melakukan prediksi nilai variabel dependen berdasarkan nilai-nilai variabel independen yang diberikan. Algoritma *Support Vector Regression* sendiri merupakan sebuah algoritma yang seringkali digunakan untuk tugas regresi. Algoritma ini adalah varian dari Support Vector Machine (SVM) dan dirancang untuk melakukan prediksi nilai numerik berkelanjutan. Hal ini membuat algoritma *Support Vector Regression* cocok untuk studi kasus seperti perkiraan harga sahan dan lainnya yang memiliki tipe data numerik. [20]

Tujuan dari digunakannya algoritma *Support Vector Regression* adalah untuk menemukan sebuah model regresi dengan margin disekitar nilai prediksi yang ingin dicari. Penggunaan algoritma ini memungkinkan adanya keseimbangan antara penyesuaian data dan menghindari overfitting. Hal ini sangat berguna dan tentunya berfungsi dengan baik untuk mendapatkan hasil prediksi yang akurat. Implementasi dasar *Support Vector Regression* adalah menggunakan *Python* dengan menggunakan *library Machine Learning* scikit learn. [21]

2.4 Tools

2.4.1 Microsoft Excel

Microsoft Excel merupakan *software* yang dimiliki oleh perusahaan *Microsoft*. Excel diluncurkan oleh *Microsoft* pada tahun 1985. *Software* Excel memiliki kegunaan untuk membantu organisasi atau perusahaan bisnis dalam melakukan penyimpanan data dalam bentuk tabel atau *spreadsheet*. [22] Selain itu, Excel juga memiliki kemampuan untuk melakukan pengolahan dan perhitungan pada data yang bersifat numerik. Kemampuan ini dilakukan dengan menggunakan berbagai rumus yang tersedia didalam *software Microsoft Excel*. Penggunaan *Microsoft Excel* juga tergolong cukup fleksible dan dapat dihubungkan kedalam aplikasi atau *software Microsoft* lainnya. [23]

2.4.2 Python

Python adalah sebuah bahasa pemrograman yang cukup populer dikalangan para programmer. Bahasa *Python* merupakan salah satu bahasa pemrograman yang berorientasi pada object (*object-oriented*). *Python* sendiri pada umumnya digunakan untuk membuat sebuah aplikasi dan juga melakukan analisa pada data. *Python* menyediakan struktur data tingkat tinggi serta syntax yang sangat sederhana dan mudah untuk dipahami. Selain itu, *Python* sendiri bersifat *open-source* dan dapat dijalankan secara mudah dan praktis pada setiap komputer modern. Penggunaan *Python* pastinya akan meningkatkan produktivitas para pengguna. Hal ini dapat terjadi karena *Python* menyediakan *Library* yang lengkap dan juga pengoperasiannya yang dapat dilakukan diberbagai *platform* seperti Windows, Mac, dan juga Linux. [24]

2.4.3 Visual Studio Code

Visual Studio Code merupakan sebuah aplikasi editor teks yang dikembangkan oleh *Microsoft* serta mendapatkan nominasi sebagai Most Popular Development Environment pada Stack Overflow Developer Survey di tahun 2019. *Visual Studio Code* sendiri dirancang dan dikembangkan untuk menjadi sebuah aplikasi code editor yang ringan, cepat, dan mudah untuk digunakan beberapa *user* dengan bahasa pemrograman yang beragam. Selain itu, *Visual Studio Code* juga dapat digunakan secara gratis, bersifat open source dan cross-platform Hal ini tentunya menjadikan *Visual Studio Code* sebagai salah satu code editor yang paling populer untuk digunakan saat ini dikarenakan oleh beberapa kemampuan dan kelebihan yang dimilikinya. Selain itu, *Visual Studio Code* juga memiliki keunggulan yang sangat berguna yaitu ekosistem ekstensi yang luas dan besar. Hal ini tentunya memungkinkan *user* untuk menyesuaikan dan memperluas fungsionalitas editor sesuai dengan keperluan pengguna seseuai dengan kebutuhannya. *Visual Studio Code* dapat digunakan secara gratis diberbagai platform antara lain Windows, Linux, dan juga macOS.[25]