



PROPOSAL TEKNIS

IMPLEMENTASI SISTEM ANALISA DESKRIPTIF DAN PREDIKTIF DENGAN BIG DATA BERBASIS CLOUD



KRAKATAU
INFORMATION TECHNOLOGY



DAFTAR ISI

1. Pendahuluan

- 1.1. Latar Belakang Masalah
- 1.2. Maksud dan Tujuan

2. Ruang Lingkup Proyek

- 2.1. Kesesuaian Tenaga Ahli
- 2.2. Kesesuaian Ruang Lingkup

3. Metode Pengembangan, Manajemen & Penjadwalan Proyek

- 3.1. Metode Analisis dan Machine Learning
- 3.1. Metode Pengembangan Agile SDLC
- 3.2. Manajemen Proyek

4. Hasil Pekerjaan



1. Pendahuluan

Penggunaan laporan dan *dashboard* merupakan hal yang penting bagi suatu perusahaan dalam memahami kondisi bisnisnya. INALUM membutuhkan solusi untuk pengelolaan data yang terintegrasi dengan berbagai platform dan tersentralisasi mengikuti perkembangan kegiatan usaha di berbagai bidang sehingga dapat dilakukan analisa dalam pengambilan keputusan strategis. Saat ini, pemrosesan data di PT INALUM untuk kebutuhan laporan dan *dashboard* masih belum memenuhi kebutuhan yang diharapkan, khususnya *dashboard* yang bisa menampilkan hasil analisis deskriptif dan prediktif terkait kebutuhan stok dan prediksi harga.

Seiring dengan kebutuhan tersebut, maka PT INALUM berencana akan membangun suatu Big Data berbasis cloud yang terbentuk dari berbagai sumber, baik dari sumber data eksternal berupa API dan atau flat file dari penyedia data terkemuka seperti CRU, Bloomberg, Wood MacKenzie dan MBR, dan dan sumber data internal PT INALUM yang berasal dari Aplikasi Core Bisnis (Operasional), Akuntansi, Investasi, dan aplikasi lainnya untuk mendukung proses analisis deskriptif dan prediktif yang lebih efisien, efektif, dan komprehensif dengan menggunakan Machine Learning dan AI sehingga informasi yang disajikan pada *dashboard* dapat disajikan secara tepat dan akurat dan sesuai dengan apa yang dibutuhkan oleh PT INALUM.

1.1. Latar Belakang Masalah

PT Inalum, sebagai entitas bisnis yang beroperasi dalam berbagai sektor, menghadapi tantangan dalam mengelola data yang tersebar dari berbagai platform. Proses bisnis, terutama dalam konteks Purchase Requisition (PR), pengelolaan vendor, dan pemantauan kesepakatan, sering kali mengalami keterlambatan akibat kekurangan sistem untuk mengintegrasikan data secara efektif. Hal ini mengakibatkan kurangnya visibilitas dan kesulitan dalam mengambil keputusan strategis yang cepat dan tepat. Selain itu, ketidakmampuan dalam meramalkan fluktuasi harga bahan baku seperti LME dan MJP berdasarkan data historis menyebabkan ketidakpastian dalam perencanaan keuangan, mempengaruhi efisiensi operasional dan daya saing perusahaan.

Dalam konteks ekspansi bisnis dan peningkatan kompleksitas operasional, perlunya solusi manajemen data terintegrasi menjadi semakin mendesak. Tantangan dalam pemantauan dan analisis data, kurangnya transparansi keuangan, dan ketergantungan pada sumber eksternal untuk prediksi harga adalah hambatan utama yang perlu diatasi. Oleh karena itu, pengembangan solusi yang dapat menyatukan, menganalisis, dan meramalkan data dari berbagai sumber menjadi imperatif untuk mendukung PT Inalum dalam mencapai efisiensi operasional yang optimal dan meningkatkan daya saing di pasar yang terus berubah.



1.2. Maksud dan Tujuan

Maksud:

PT INALUM ingin membangun Business Intelligence yang sesuai dengan kebutuhan bisnis terkini. Kerja sama ini bertujuan untuk menciptakan solusi yang mudah dioperasikan, efisien, serta memiliki skalabilitas yang optimal.

Tujuan:

Pembangunan sistem Descriptive & Predictive Analytic dengan Big Data berbasis cloud di PT INALUM diarahkan oleh IT Strategic Plan perusahaan, dengan fokus pada mencapai tujuan-tujuan berikut:

1. **Pemberian Informasi Manajemen:** Menghasilkan informasi yang bermanfaat bagi manajemen PT INALUM sesuai dengan tugas dan tanggung jawab mereka masing-masing. Informasi yang dihasilkan akan membantu manajemen dalam mengambil keputusan strategis yang lebih terinformasi.
2. **Perencanaan Bisnis yang Lebih Akurat:** Meningkatkan akurasi dalam perencanaan bisnis operasional pelayanan PT INALUM dengan memberikan akses ke data terstruktur dan aktual. Ini akan membantu perusahaan untuk merumuskan rencana bisnis yang lebih tepat dan efektif.
3. **Pengambilan Keputusan yang Cepat dan Akurat:** Mendukung proses pengambilan keputusan yang lebih cepat dan lebih tepat dengan memberikan akses langsung ke informasi bisnis yang relevan. Dengan adanya data yang terstruktur dan mudah diakses, keputusan dapat diambil dengan dasar yang lebih kuat.
4. **Pemutakhiran Informasi Terpusat:** Menciptakan sumber informasi terpusat yang diperbarui secara konsisten. Ini memastikan bahwa semua pihak yang berwenang di perusahaan memiliki akses ke informasi yang terbaru, mengurangi risiko kesalahan atau informasi yang tidak konsisten.

Dengan membangun Descriptive & Predictive Analytic dengan Big Data sesuai dengan rencana ini, PT INALUM berharap dapat mengoptimalkan pengelolaan data, meningkatkan efisiensi operasional, dan meningkatkan kemampuan pengambilan keputusan dalam mendukung pertumbuhan dan keberlanjutan perusahaan.



2. Ruang Linkup Proyek

Dalam melaksanakan proyek pengadaan dan pengembangan sistem Descriptive & Predictive Analytic dengan Big Data berbasis cloud untuk PT INALUM, PT. Bumi Madani akan menjalankan serangkaian kegiatan yang mencakup aspek-aspek berikut:

1. **Rancangan Sistem:** PT. Bumi Madani akan merancang sistem Descriptive & Predictive Analytic dengan Big Data yang sesuai dengan kebutuhan PT INALUM. Ini mencakup desain arsitektur, skema basis data, alur data, dan komponen-komponen sistem.
2. **Inventarisasi dan Persiapan Data:** PT. Bumi Madani akan melakukan inventarisasi terhadap sumber data yang diperlukan untuk mendukung proyek ini. Proses ini termasuk membangun dan menyiapkan metadata yang menjadi dasar untuk proses Business Intelligence.
3. **Pengembangan Sistem Pendukung Pengambilan Keputusan:** PT. Bumi Madani akan mengembangkan sistem yang mendukung pengambilan keputusan berbasis Business Intelligence. Ini termasuk pengembangan fitur-fitur yang memungkinkan analisis dan visualisasi data untuk keperluan manajemen.
4. **Pemodelan dan Analisis Data:** PT. Bumi Madani akan mengembangkan model data yang sesuai dengan kebutuhan dan tujuan PT INALUM. Ini melibatkan analisis mendalam terhadap data yang akan digunakan, serta pemodelan yang mengakomodasi kebutuhan pengguna.
5. **Pengadaan Perangkat Lunak:** PT. Bumi Madani akan menyediakan perangkat lunak yang diperlukan untuk pengembangan sistem, termasuk alat-alat Business Intelligence dan komponen lain yang relevan.
6. **Pengembangan Laporan dan Dashboard:** PT. Bumi Madani akan mengembangkan laporan-laporan dan dashboard yang diperlukan oleh PT INALUM. Ini mencakup desain visual, format, dan fungsionalitas dari laporan dan dashboard tersebut.
7. **Pelatihan Pengguna:** PT. Bumi Madani akan memberikan pelatihan kepada pegawai PT INALUM agar mereka dapat mengoperasikan sistem yang dikembangkan dengan mandiri. Pelatihan akan mencakup pengguna akhir, administrator, serta troubleshooting dan maintenance.
8. **Capacity Planning:** PT. Bumi Madani akan memberikan rekomendasi mengenai infrastruktur dan analisa pendukung yang diperlukan untuk aplikasi ke depan.
9. **Pemeliharaan Sistem:** Setelah implementasi, PT. Bumi Madani akan memberikan layanan pemeliharaan sesuai dengan timeline yang telah ditetapkan. Ini mencakup perbaikan bug dan patch, peningkatan dan perbaikan produk, serta layanan dukungan teknis atas permasalahan yang timbul.



Keseluruhan ruang lingkup pekerjaan ini dirancang untuk mencapai tujuan akhir pembuatan sistem Descriptive & Predictive Analytic dengan BigData berbasis cloud yang efisien, akurat, dan dapat mendukung pengambilan keputusan yang lebih baik serta pemutakhiran informasi terpusat bagi PT INALUM.

2.1. Kesesuaian Tenaga Ahli

Uraian penugasan untuk setiap Tenaga Ahli yang dilibatkan dalam pekerjaan ini dapat dilihat pada tabel sebagai berikut :

Staf Profesional		
Tenaga Ahli	Jumlah	Penugasan
Project Manager	1 orang	<ul style="list-style-type: none">- Perencanaan Proyek: Mengembangkan rencana proyek yang jelas, termasuk jadwal, anggaran, dan sumber daya yang diperlukan untuk mengelola tim Data Scientist, AI Engineer, Data Analyst, Business Intelligence, Data Engineer, dan Cloud Engineer dalam mencapai tujuan proyek.- Koordinasi Tim: Memastikan kolaborasi yang efisien antara semua anggota tim, mengidentifikasi dan mengatasi hambatan yang mungkin muncul, serta memantau kemajuan proyek secara keseluruhan.- Pelaporan dan Komunikasi: Bertanggung jawab untuk berkomunikasi secara efektif dengan pemangku kepentingan proyek, menyediakan pembaruan berkala tentang kemajuan proyek, dan mengelola perubahan yang mungkin terjadi selama proyek.
Data Scientist	2 orang	<ul style="list-style-type: none">- Pemahaman Bisnis dan Data: Memahami tujuan bisnis proyek dan kebutuhan data yang diperlukan untuk pengembangan model analitik. Melakukan analisis data, eksplorasi, transformasi, dan pemahaman awal.- Analisis Data Mendalam: Menggunakan BigQuery dan alat analitik lainnya untuk melakukan analisis data yang mendalam, mengidentifikasi tren, pola, dan wawasan yang relevan.- Model Machine Learning: Membangun model machine learning dengan BigQuery ML berdasarkan hasil analisis data. Melakukan validasi dan fine-tuning model jika diperlukan.- Dokumentasi dan Pelaporan: Membuat



		dokumentasi lengkap tentang model, proses, hasil, dan rekomendasi. Berkontribusi dalam pembuatan laporan yang menjelaskan temuan dan solusi kepada tim dan pemangku kepentingan.
AI Engineer	1 orang	<ul style="list-style-type: none">- Pemahaman Bisnis dan Model AI: Memahami tujuan bisnis proyek dan mengidentifikasi kebutuhan model AI. Merancang, mengembangkan, dan mengimplementasikan model AI.- Pengembangan Model: Mengembangkan dan melatih model machine learning dan deep learning, serta memastikan integrasi yang tepat dengan infrastruktur IT.- Validasi Model AI: Menguji model AI, melakukan fine-tuning, dan memastikan kinerja yang optimal. Menerapkan model ke Vertex AI Prediction untuk prediksi real-time.- Dokumentasi dan Pelaporan: Membuat dokumentasi lengkap tentang model AI, proses pengembangan, hasil, dan rekomendasi. Berkontribusi dalam pembuatan laporan yang menjelaskan temuan dan solusi kepada tim dan pemangku kepentingan.
Data Analyst	1 orang	<ul style="list-style-type: none">- Analisis Data: Mengimpor, menganalisis, dan memvisualisasikan data menggunakan alat analitik seperti BigQuery. Memberikan wawasan awal dari data dan mengidentifikasi pola atau tren yang relevan.- Analisis Mendalam: Melakukan analisis lebih mendalam untuk mengidentifikasi pola atau tren yang lebih kompleks dalam data. Menggunakan alat analitik untuk menganalisis data dengan detail.- Rekomendasi: Menyusun rekomendasi berdasarkan hasil analisis data, yang dapat digunakan untuk mengambil keputusan yang lebih baik dalam bisnis.- Dokumentasi: Membuat dokumentasi lengkap tentang analisis yang dilakukan dan hasilnya.
Business Intelligence	1 orang	<ul style="list-style-type: none">- Pemahaman Kebutuhan Bisnis: Memahami kebutuhan bisnis dan tujuan proyek BI. Merancang dan mengembangkan solusi BI yang

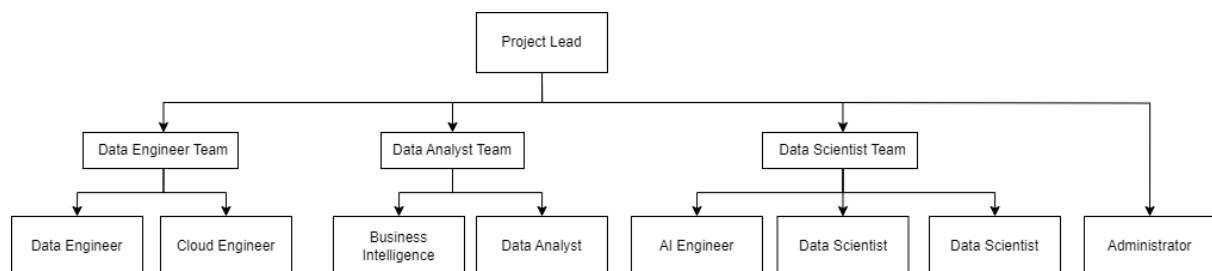


		<p>sesuai dengan tujuan tersebut.</p> <ul style="list-style-type: none">- Pemodelan Data: Merancang skema data dan model data yang mendukung proses pelaporan dan analisis bisnis. Mempersiapkan data yang dibutuhkan untuk solusi BI.- Pengembangan Dashboard dan Laporan: Mengembangkan tampilan Dashboard yang informatif dan laporan yang menjelaskan hasil analisis. Menggunakan alat seperti Tableau untuk visualisasi data.- Pelaporan dan Monitoring: Mengatur pelaporan rutin menggunakan solusi BI untuk memberikan pemahaman berkelanjutan kepada tim dan pemangku kepentingan.- Pelaporan Akhir dan Komunikasi: Mengintegrasikan hasil analisis ke dalam laporan akhir yang merangkum kinerja dan hasil selama periode tertentu. Menyampaikan temuan, rekomendasi, dan visualisasi yang relevan kepada pemangku kepentingan.
Data Engineer	1 orang	<ul style="list-style-type: none">- Desain Data Pipeline: Merancang data pipeline dengan Dataflow, mengidentifikasi sumber data, dan membangun infrastruktur yang dibutuhkan.- Pembangunan Data Pipeline: Menerapkan data pipeline dalam mode streaming dengan Dataflow, menghubungkan sumber data ke BigQuery dan penyimpanan Cloud Storage, serta memastikan transformasi data.- Monitoring dan Maintenance: Melakukan pemantauan rutin, pemecahan masalah, dan menjaga dokumentasi terkait setup dan alur data pipeline. Mengoptimalkan kinerja dan ketersediaan data.
Cloud Engineer	1 orang	<ul style="list-style-type: none">- Infrastruktur Cloud: Memastikan infrastruktur cloud seperti Compute Engine, Pub/Sub, dan Cloud Functions berjalan dengan baik untuk mendukung kebutuhan tim Data Engineer, Data Scientist, dan AI Engineer.- Integrasi Layanan Cloud: Mengintegrasikan layanan cloud seperti Cloud Functions dengan data pipeline, memastikan sinkronisasi data yang baik, dan mengelola perubahan dan pembaruan sesuai kebutuhan.



		<ul style="list-style-type: none">- Kinerja dan Pengoptimalan Cloud: Menganalisis kinerja layanan cloud, mengidentifikasi area perbaikan, dan melakukan pengoptimalan untuk efisiensi dan kinerja yang lebih baik dalam seluruh lingkungan cloud.
--	--	---

2.1.1. Struktur Organisasi Proyek



2.2. Kesesuaian Ruang Lingkup

Kesesuaian ruang lingkup proyek Big Data dengan Google Cloud Platform (GCP) adalah langkah penting dalam memahami dan merancang proposal teknis yang efektif. Dalam konteks ini, ruang lingkup proyek mencakup pemahaman yang mendalam tentang infrastruktur GCP yang tersedia, alat dan layanan Big Data yang disediakan oleh GCP, serta bagaimana proyek akan mengintegrasikan dan memanfaatkan sumber daya ini untuk mencapai tujuan bisnis.

Kesesuaian ruang lingkup juga mencakup perencanaan yang tepat, pemilihan alat yang sesuai, dan alokasi sumber daya yang efisien, sehingga proyek Big Data dapat berjalan dengan lancar dan menghasilkan hasil yang optimal dengan memanfaatkan kapabilitas GCP. Dengan memastikan kesesuaian ruang lingkup proyek Big Data dengan GCP, kita dapat memastikan bahwa proyek akan berjalan sesuai harapan dan memberikan manfaat yang signifikan bagi bisnis atau organisasi yang bersangkutan.

Berikut Penjelasan Product / Service Google Cloud Platform:

1. Compute Engine

- 1 x Dataproc master node
 - Region: Jakarta
 - 30 total hours per month
 - Provisioning model: Regular
 - Instance type: n2-standard-2
 - Operating System / Software: Free
- 3 x Dataproc worker nodes
 - Region: Jakarta
 - 90 total hours per month



- Provisioning model: Regular
 - Instance type: n2-standard-8
 - Operating System / Software: Free
 - 1 x Dataproc master node
 - Region: Jakarta
 - 730 total hours per month
 - Provisioning model: Regular
 - Instance type: n1-standard-1
 - Operating System / Software: Free
 - 3 x Dataproc worker nodes
 - Region: Jakarta
 - 2,190 total hours per month
 - Provisioning model: Regular
 - Instance type: n1-standard-1
 - Operating System / Software: Free
2. Cloud Storage
- 1x Standard Storage
 - Location: Jakarta
 - Total Amount of Storage: 4,096 GiB
 - Always Free usage included: No
3. BigQuery General
- BigQuery On-Demand
 - Location: Jakarta
 - Active Logical Storage: 10,240 GiB
 - Queries: 100 TiB
4. BigQuery ML
- Location: Jakarta
 - Create Model: 5 TB
 - Prediction: 5 TB
 - Evaluation: 0 TB
5. Pub/Sub
- Message delivery type: Basic
 - Volume: 33.3 GiB
 - Subscriptions: 1
 - Topic retention cost: 8.991
6. Dataflow
- 1 x n1-standard-1 workers in Streaming Mode
 - Region: Jakarta
 - Total vCPU Hours: 730
 - Total Memory Hours: 2,737.5 GiB/Hours
 - PD Local Storage: 146,000 GiB/hours
7. Dataproc
- Region: Jakarta
 - Cluster size: 26 vCPUs
 - Time used: 30 hours
 - Master nodes: 1



- Normal worker nodes: 3
 - Spot worker nodes: 0
 - Storage:
 - (Per node): 100.000 GiB
 - (All nodes, 30 hours usage equivalent): 16.438 GiB
8. Cloud Function
- Region: Jakarta
 - Invocations: 30
 - RAM (GiB-seconds): 144,000 per month
 - CPU (GHz-seconds): 86,400 per month
9. Cloud Data Fusion
- Number of instances: 1
 - Edition: ENTERPRISE
 - Average hours per month instance is running: 30 hours
10. Cloud VPN
- Region: Jakarta
 - VPN: 1,460 tunnel hours per month
11. Persistent Disk (Accompanying)
- Product accompanying: Dataproc (x1)
 - Zonal standard PD: 16.438 GiB
 - Product accompanying: Dataproc (x1)
 - Zonal standard PD: 2000 GiB
12. Vertex AI Training
- Location: Jakarta
 - Instance type: n2-standard-32
 - Training jobs per week: 1
 - Average training job length (hour): 8
13. Vertex AI Prediction (x2)
- Location: Jakarta
 - Instance type: n2-standard-32
 - Number of models to serve: 1
 - Average hours per day each server is running: 1
 - Average days per week each server is running: 1
14. AutoML for Structured data with Vertex AI
- Training hours (Edge): 34.667
15. Sumber Data
- CRU
 - Metal Bulletin Research
 - Wood Mackenzie
 - Bloomberg
 - SAP HANA
 - Flat flies
 - Non-SAP
16. Variabel Tambahan
- Aluminium Price : Harga Aluminium (y)
 - Copper Price : Harga Copper



- BADI-2 : Baltic Dry Index
- MJP : Premium Main Japan Port (Mid-Point, Quarterly)
- USD Index : Indeks USD
- AL Prod (Global) : Produksi Aluminium Global
- AL Cons (Global) : Konsumsi Aluminium Global
- Aluminium Stocks : Stok Aluminium Dunia
- Alumina Index : Index Alumina
- Oil Prices : Harga Minyak

3. Metode Pengembangan, Manajemen & Penjadwalan Proyek

Dalam bagian Metode Pengembangan, Manajemen, dan Penjadwalan Proyek proposal teknis Big Data dengan Google Cloud Platform (GCP), kami akan merinci pendekatan yang akan kami ambil untuk menghadirkan solusi Big Data yang inovatif dan efisien. Kami akan menguraikan strategi pengembangan yang melibatkan pemilihan alat-alat dan layanan GCP yang optimal, serta merinci tahapan dan metode yang akan kami tempuh dalam proses pengembangan.

Selain itu, kami akan menjelaskan pendekatan manajemen proyek kami yang akan memastikan pemantauan, pengendalian, dan pelaporan yang ketat, sehingga proyek berjalan sesuai rencana dan anggaran. Kami juga akan menyampaikan penjadwalan proyek yang akan membantu memastikan setiap tahap proyek berjalan dengan efisien dan tepat waktu.

3.1. Metode Analisis dan Machine Learning

Dalam bagian Metode Analisis dan Machine Learning dari proposal teknis Big Data dengan Google Cloud Platform (GCP), kami akan menguraikan pendekatan kami untuk mengungkap wawasan yang berharga dari data besar yang terkumpul. Kami akan menjelaskan metode analisis yang akan digunakan untuk memproses dan menggali makna dari data, dengan penekanan pada teknik-teknik eksploratif dan statistik yang relevan. Selain itu, kami akan menguraikan bagaimana kami akan menerapkan Machine Learning untuk membangun model yang dapat memprediksi, mengklasifikasikan, dan memahami pola dalam data, sehingga dapat memberikan solusi cerdas untuk tantangan bisnis yang dihadapi. Pendekatan ini akan memungkinkan kami untuk mengoptimalkan penggunaan Big Data dalam proyek ini dan menghasilkan wawasan yang dapat memberikan manfaat strategis bagi PT Inalum.

3.1.1. Linear Models

Linear regression adalah salah satu metode statistik yang digunakan untuk memodelkan hubungan antara satu atau lebih variabel independen (disebut juga prediktor atau variabel input) dan variabel dependen (disebut juga variabel target atau variabel output) yang bersifat kontinu. Tujuan dari linear regression adalah untuk menemukan hubungan



linier antara variabel input dan output sehingga kita dapat memprediksi nilai output berdasarkan input yang diberikan.

Ada dua jenis linear regression yang umum digunakan:

Simple Linear Regression: Dalam kasus ini, hanya ada satu variabel independen yang digunakan untuk memprediksi variabel dependen. Modelnya memiliki bentuk seperti ini:

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

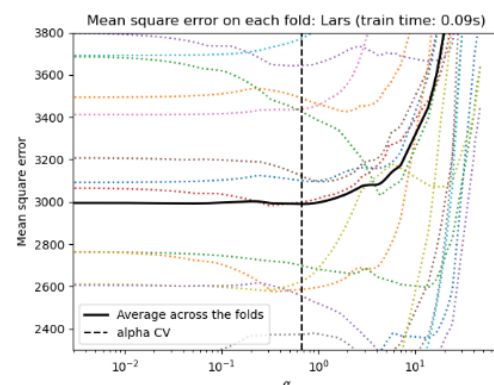
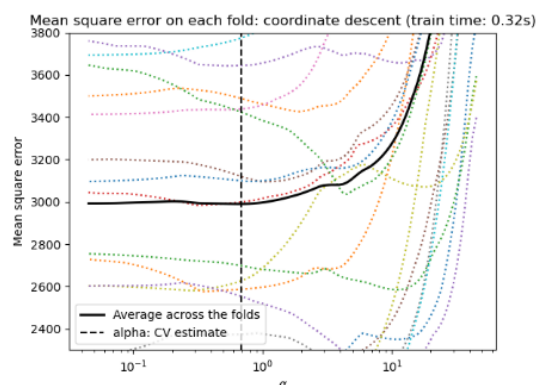
- Y adalah variabel dependen.
- X adalah variabel independen.
- β_0 adalah intercept (nilai Y ketika $X=0$)
- β_1 adalah koefisien regresi yang mengukur hubungan linier antara X dan Y.
- ϵ adalah kesalahan (error) yang merupakan perbedaan antara nilai aktual Y dan nilai yang diprediksi oleh model.

Multiple Linear Regression: Dalam kasus ini, terdapat lebih dari satu variabel independen yang digunakan untuk memprediksi variabel dependen. Modelnya memiliki bentuk seperti ini:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon$$

- Y adalah variabel dependen.
- X_1, X_2, \dots, X_p adalah variabel independen.
- β_0 adalah intercept.
- $\beta_1, \beta_2, \dots, \beta_p$ adalah koefisien regresi yang mengukur hubungan linier antara masing-masing variabel independen dan Y.
- ϵ adalah kesalahan.

Linear regression menggunakan metode least squares untuk menemukan nilai terbaik dari β yang sesuai dengan data. Metode ini mencari nilai β yang meminimalkan jumlah kuadrat kesalahan (ϵ^2).





Linear Models adalah keluarga metode statistik yang digunakan untuk menggambarkan hubungan linier antara variabel independen (prediktor) dan variabel dependen (hasil). Berikut adalah beberapa contoh penggunaan Linear Models untuk analisis data yang direkomendasikan:

1. Prediksi Harga Aluminium (y) berdasarkan Variabel LME dan MJP Price Approach

Dalam hal ini, penganalisis dapat menggunakan regresi linier untuk memahami bagaimana perubahan dalam LME (London Metal Exchange) dan MJP (Main Japan Port) Price memengaruhi harga aluminium. Tim proyek dapat membangun model linier yang memprediksi harga aluminium berdasarkan data historis LME, MJP, dan faktor-faktor lain seperti BADI-2, indeks USD, dan lainnya.

2. Analisis Persediaan Aluminium Dunia (Aluminium Stocks)

Pengembang dapat menggunakan regresi linier untuk mengidentifikasi faktor-faktor yang mempengaruhi stok aluminium dunia. Ini bisa melibatkan variabel seperti produksi aluminium global (AL Prod (Global)) dan konsumsi aluminium global (AL Cons (Global)). Model linier dapat membantu dalam memahami hubungan antara variabel-variabel ini dan stok aluminium dunia.

3. Prediksi Total Pengeluaran (Total Spend)

Pengembang dapat menggunakan regresi linier untuk memprediksi total pengeluaran perusahaan berdasarkan variabel-variabel seperti jumlah pemasok aktif (Active Supplier), jumlah pemasok terdaftar (Registered Suppliers), dan lainnya. Ini dapat membantu dalam perencanaan anggaran dan pengelolaan sumber daya perusahaan.

4. Pemahaman Pengaruh Harga Minyak (Oil Prices) terhadap Biaya Produksi dan Logistik

Pengembang dapat menggunakan regresi linier untuk memahami bagaimana perubahan harga minyak mempengaruhi biaya produksi dan logistik perusahaan. Ini dapat membantu dalam perencanaan dan pengelolaan biaya perusahaan.

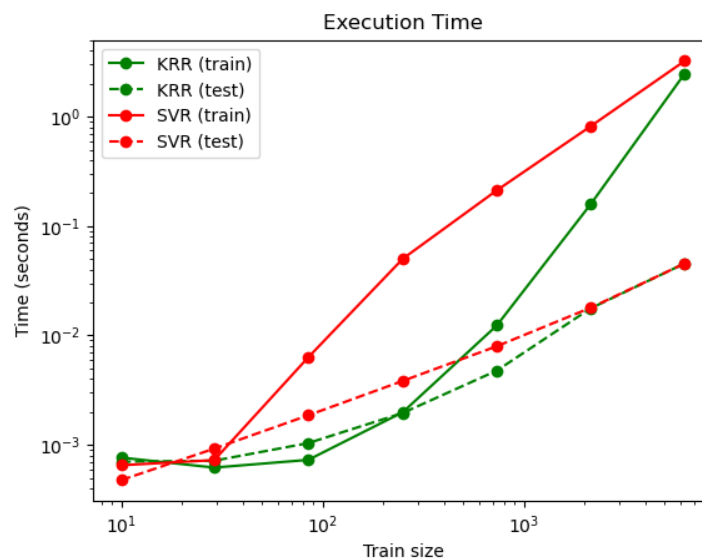
5. Prediksi Kinerja Vendor berdasarkan Sejarah Pengiriman Vendor

Untuk Vendor Delivery Performance, pengembang dapat menggunakan regresi linier untuk memprediksi kinerja vendor berdasarkan data historis pengiriman vendor, yang mungkin melibatkan faktor-faktor seperti wilayah (Region) dan pabrik (Plant).



3.1.2. Kernel Ridge Regression

Kernel Ridge Regression (KRR) adalah metode regresi yang merupakan perluasan dari regresi ridge dengan menggunakan kernel. Ini adalah teknik yang berguna untuk memodelkan hubungan antara variabel independen dan dependen, khususnya ketika hubungan tersebut tidak linier. KRR mencoba menemukan hubungan antara variabel input dan output dengan menggabungkan karakteristik dari regresi ridge dan kernel tricks yang umumnya digunakan dalam Support Vector Machines (SVM).



Berikut adalah penjelasan tentang Kernel Ridge Regression:

1. Kernel Trick: Kernel adalah fungsi matematis yang memungkinkan kita untuk mentransformasikan data ke ruang fitur yang memiliki dimensi lebih tinggi. Dalam KRR, kernel digunakan untuk menghitung jarak atau kesamaan antara pasangan titik data di ruang fitur yang lebih tinggi. Kernel umum yang digunakan termasuk kernel linier, kernel polynomial, kernel radial basis function (RBF/Gaussian), dan lain-lain.
2. Objektif KRR: Tujuan KRR adalah untuk menemukan fungsi regresi yang memiliki hubungan linier dalam ruang fitur yang lebih tinggi. Ini dapat diformulasikan sebagai berikut:

$$f(x) = \sum_{i=1}^n \alpha_i K(x, x_i)$$

- $f(x)$ adalah fungsi regresi yang memprediksi nilai variabel target y berdasarkan variabel input x .
- α_i adalah koefisien yang harus dipelajari dari data.
- $K(x, x_i)$ adalah kernel yang mengukur kesamaan antara x dan x_i dalam ruang fitur yang ditingkatkan.



3. Regularisasi Ridge: Seperti regresi ridge, KRR juga termasuk elemen regularisasi untuk mencegah overfitting. Ini dilakukan dengan menambahkan norma L2 dari koefisien α_i ke dalam objektif optimasi.

KRR adalah varian dari regresi linier yang memungkinkan pemodelan hubungan yang lebih kompleks antara variabel. Ini berguna ketika hubungan antara variabel independen dan dependen tidak bersifat linier, dan bisa lebih fleksibel dalam menangani data yang rumit. Dalam konteks proyek ini, berikut adalah beberapa cara KRR dapat digunakan:

1. Prediksi Harga Aluminium (y) berdasarkan berbagai faktor

Pengembang dapat menggunakan KRR untuk memodelkan hubungan antara Harga Aluminium dan semua variabel independen yang dibutuhkan, seperti Harga Copper, Baltic Dry Index (BADI-2), Premium Main Japan Port (MJP), Indeks USD (USD Index), Produksi Aluminium Global (AL Prod (Global)), Konsumsi Aluminium Global (AL Cons (Global)), Stok Aluminium Dunia (Aluminium Stocks), Index Alumina (Alumina Index), dan Harga Minyak (Oil Prices). KRR akan memungkinkan analisis untuk menangkap hubungan non-linier dan interaksi antara faktor-faktor ini.

2. Analisis Pengaruh Faktor-faktor pada Pengeluaran (Total Spend)

Metode analisis dapat menggunakan KRR untuk memahami bagaimana faktor-faktor seperti Active Supplier/Registered Suppliers ratio, Total Number of Invoice, Harga Minyak (Oil Prices), dan variabel lainnya mempengaruhi Total Spend perusahaan. KRR akan membantu dalam mengidentifikasi hubungan non-linier yang mungkin ada di antara faktor-faktor ini.

3. Prediksi Kinerja Vendor berdasarkan Sejarah Pengiriman Vendor

Untuk Vendor Delivery Performance, KRR dapat digunakan untuk memodelkan bagaimana faktor-faktor seperti wilayah (Region) dan pabrik (Plant) mempengaruhi kinerja vendor dengan lebih akurat daripada model linier.

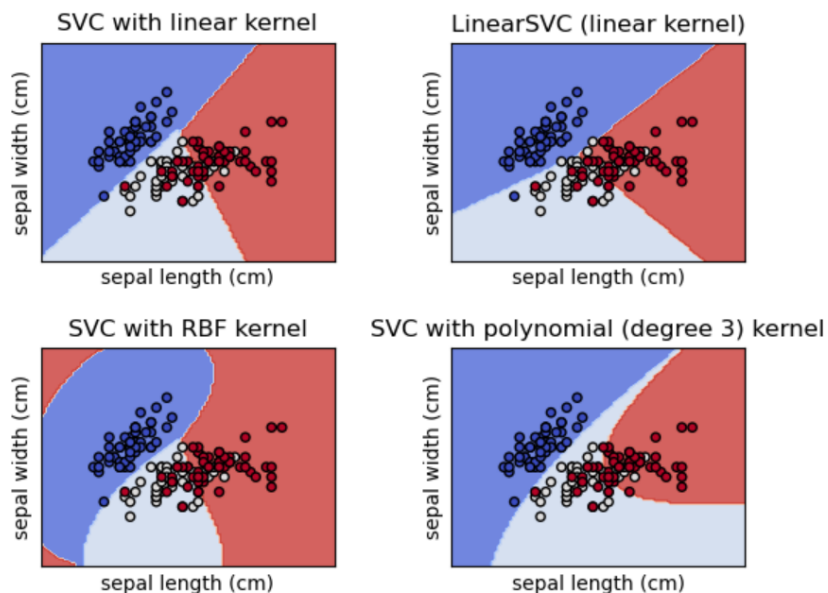
KRR memungkinkan penggunaan fungsi kernel yang memetakan data ke ruang dimensi yang lebih tinggi, yang dapat membantu dalam menangkap pola yang rumit dalam data. Ini sering digunakan dalam situasi di mana regresi linier konvensional tidak cukup untuk menjelaskan hubungan antara variabel.

3.1.3. Support Vector Machines

Support Vector Machines (SVM) adalah salah satu metode machine learning yang digunakan untuk tugas klasifikasi dan regresi. SVM bertujuan untuk menemukan batas keputusan (decision boundary) yang optimal untuk memisahkan dua kelas data atau memprediksi nilai regresi. Batas keputusan ini disebut "hyperplane" dalam ruang fitur yang



ditingkatkan. SVM dikenal karena kemampuannya dalam menangani data yang tidak linier melalui kernel tricks.



Berikut adalah penjelasan tentang Support Vector Machines:

1. **Hyperplane:** SVM mencari hyperplane yang memaksimalkan margin (jarak terdekat antara hyperplane dan titik data) antara dua kelas yang berbeda. Hyperplane ini adalah batas keputusan yang digunakan untuk mengklasifikasikan data ke dalam kelas yang sesuai.
2. **Kernel Tricks:** SVM memungkinkan penggunaan kernel functions, seperti kernel linier, kernel polynomial, kernel radial basis function (RBF/Gaussian), dan lain-lain, untuk mengubah data ke dalam ruang fitur yang memiliki dimensi lebih tinggi. Ini memungkinkan SVM untuk menangani data yang tidak linier dan menemukan hyperplane yang lebih kompleks untuk memisahkan kelas.
3. **Support Vectors:** Titik data yang berada paling dekat dengan hyperplane dan mempengaruhi posisi dan orientasi hyperplane disebut "support vectors." Support vectors adalah titik-titik penting dalam SVM karena mereka berkontribusi pada pembentukan batas keputusan.
4. **Regularisasi:** SVM juga mencakup elemen regularisasi yang membantu mengendalikan overfitting. C parameter dalam SVM mengontrol sejauh mana model SVM memperhatikan titik data yang memiliki kesalahan klasifikasi.

Berikut adalah beberapa cara SVM dapat digunakan dalam konteks proyek ini:

1. **Prediksi Harga Aluminium (y) berdasarkan Faktor-faktor Eksternal**

Alternatif metode analisis dapat menggunakan SVM untuk memodelkan hubungan antara Harga Aluminium (Aluminium Price) dan faktor-faktor eksternal seperti Harga Copper, Baltic Dry Index (BADI-2),



Premium Main Japan Port (MJP), Indeks USD (USD Index), Produksi Aluminium Global (AL Prod (Global)), Konsumsi Aluminium Global (AL Cons (Global)), Stok Aluminium Dunia (Aluminium Stocks), Index Alumina (Alumina Index), dan Harga Minyak (Oil Prices). SVM dapat membantu dalam membangun model yang memisahkan data harga aluminium berdasarkan kombinasi faktor-faktor ini.

2. Klasifikasi Agreement Status

SVM dapat digunakan untuk mengklasifikasikan Agreement Status (status perjanjian) berdasarkan berbagai fitur, seperti Vendor, Plant, Start Date, End Date, Remaining Days, dan Contract Utilization (%). Dengan SVM, hasil analisis dapat mengklasifikasikan perjanjian sebagai aktif atau tidak aktif berdasarkan data tersebut.

3. Analisis Vendor Delivery Performance

Pengembang dapat menggunakan SVM untuk menganalisis Vendor Delivery Performance berdasarkan faktor-faktor seperti wilayah (Region) dan pabrik (Plant). SVM dapat membantu dalam mengidentifikasi pola yang mengklasifikasikan vendor sebagai pengiriman yang baik atau buruk berdasarkan data historis.

4. Segmentasi Vendor

SVM dapat digunakan untuk mengklasifikasikan vendor ke dalam kelompok atau segmen berdasarkan karakteristik tertentu, seperti Platform, Category, atau Line of Business. Ini dapat membantu dalam mengidentifikasi vendor-vendor yang serupa dalam konteks tertentu.

5. Analisis Trend Expenditing PO per Minggu

SVM dapat digunakan untuk menganalisis tren pengeluaran (expenditing) Purchase Order (PO) per minggu berdasarkan faktor-faktor yang relevan. SVM dapat membantu dalam mengidentifikasi pola-pola dalam tren pengeluaran dan memprediksi perubahan di masa depan.

SVM adalah metode yang sangat fleksibel dan dapat digunakan baik untuk klasifikasi maupun regresi, tergantung pada tujuan analisis. Ini juga efektif dalam menangani data yang kompleks dan berdimensi tinggi, seperti yang mungkin dihadapi dalam proyek analisis Big Data. Penggunaan SVM memungkinkan untuk mengidentifikasi hubungan non-linier dan mengklasifikasikan data ke dalam kelompok atau segmen yang relevan, yang dapat membantu perusahaan dalam membuat keputusan strategis untuk meningkatkan kinerja.

3.1.4. Decision Trees

Decision Trees (Pohon Keputusan) adalah salah satu metode machine learning yang digunakan untuk tugas klasifikasi dan regresi. Decision tree adalah model prediksi yang



berbentuk pohon, di mana setiap simpul (node) dalam pohon mewakili suatu keputusan atau tes pada suatu atribut, cabang-cabang dari simpul mewakili hasil tes tersebut, dan daun (leaf) mewakili label kelas atau nilai regresi. Decision Trees membagi data menjadi subset yang semakin kecil berdasarkan serangkaian keputusan untuk memprediksi hasil.

Decision tree trained on all the iris features



Berikut adalah penjelasan tentang Decision Trees:

1. Simpul Akar (Root Node): Simpul awal dalam pohon disebut akar (root node). Ini mewakili seluruh dataset dan melakukan tes pada atribut untuk membagi data ke dalam subset berdasarkan hasil tes tersebut.
2. Simpul Daun (Leaf Node): Simpul yang tidak memiliki anak disebut simpul daun (leaf node). Setiap simpul daun mewakili kelas atau nilai regresi yang diprediksi untuk data yang mencapai simpul tersebut.
3. Simpul Internal: Simpul yang memiliki anak-anak dan melakukan tes pada atribut disebut simpul internal.
4. Pembelajaran Pohon: Decision tree dibangun melalui proses pembelajaran. Algoritma pembelajaran memilih atribut terbaik untuk diuji pada setiap simpul berdasarkan kriteria tertentu seperti Information Gain atau Gini Impurity. Proses ini terus berlanjut hingga tercapai kriteria berhenti (misalnya, maksimal kedalaman pohon atau jumlah data dalam simpul tertentu).
5. Keuntungan: Keuntungan dari decision trees termasuk kemudahan interpretasi, kemampuan menangani data kategori (categorical) dan numerik, serta kemampuan mengidentifikasi atribut yang penting dalam prediksi.

Decision Trees adalah algoritma yang memodelkan keputusan berdasarkan serangkaian aturan keputusan yang dibangun berdasarkan fitur-fitur data. Dalam konteks proyek ini, berikut adalah beberapa cara Decision Trees dapat digunakan:



1. Klasifikasi Purchase Requisition (PR) Composition by Status

Alternatif metode analisis dapat menggunakan Decision Trees untuk mengklasifikasikan Purchase Requisition (PR) berdasarkan statusnya. Ini dapat membantu dalam memahami bagaimana berbagai faktor seperti kategori PR, metode tender, dan lainnya mempengaruhi status PR, seperti "approved," "pending," atau "rejected."

2. Analisis Top 10 Vendor dan Faktor-faktor Terkait

Decision Trees dapat digunakan untuk mengidentifikasi faktor-faktor terkait yang mempengaruhi vendor menjadi salah satu dari Top 10 Vendor. Ini mungkin melibatkan faktor seperti platform, kategori, dan kinerja pengiriman vendor.

3. Prediksi Agreement Status

Metode analisis dapat menggunakan Decision Trees untuk memprediksi Agreement Status berdasarkan variabel-variabel seperti Vendor, Plant, Start Date, End Date, dan lainnya. Ini dapat membantu dalam mengidentifikasi pola yang mengarah ke perjanjian aktif atau tidak aktif.

4. Segmentasi Spend by Supplier dan Category

Decision Trees dapat digunakan untuk mengelompokkan supplier dan kategori pengeluaran ke dalam segmen berdasarkan karakteristik tertentu. Ini dapat membantu dalam mengidentifikasi pola pengeluaran yang relevan.

5. Analisis Spend Trend dan Faktor-faktor Pengaruhnya

Decision Trees digunakan untuk memahami faktor-faktor yang mempengaruhi Total Spend dan tren pengeluaran perusahaan. Ini dapat membantu dalam mengidentifikasi faktor yang berkontribusi pada kenaikan atau penurunan pengeluaran.

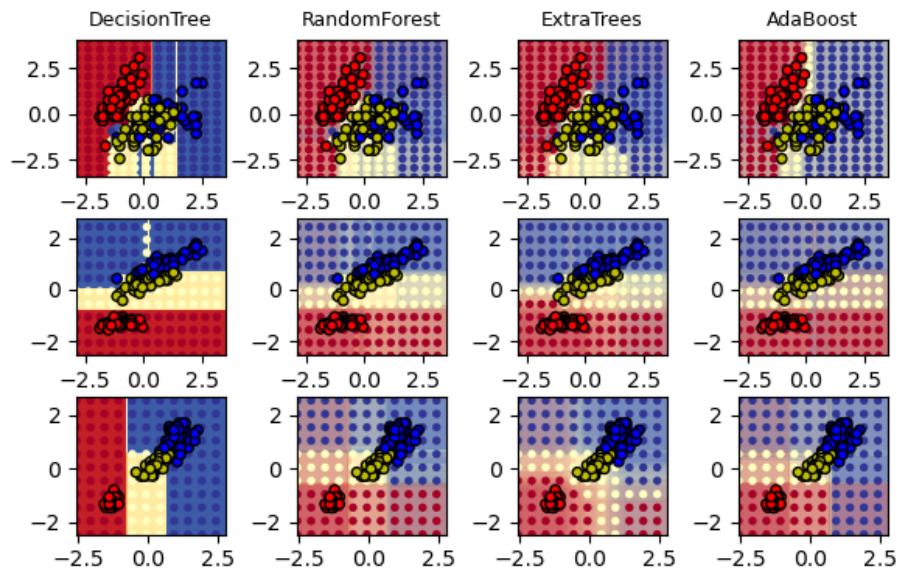
Decision Trees adalah alat yang kuat untuk analisis data karena mudah dipahami dan dapat mengungkap hubungan dan pola dalam data dengan jelas. Metode ini juga dapat digunakan untuk pengambilan keputusan yang lebih baik dengan mengidentifikasi faktor-faktor penting yang mempengaruhi hasil bisnis.

3.1.5. Ensembles (Random Forest dan Gradient Boosting)

Random Forest adalah sebuah teknik ensemble learning yang digunakan dalam machine learning untuk meningkatkan akurasi dan ketahanan model. Ini merupakan perluasan dari Decision Trees, di mana banyak pohon keputusan digunakan bersama-sama untuk membuat prediksi. Random Forest dikenal karena kemampuannya dalam mengatasi masalah overfitting yang sering terjadi pada Decision Trees.



Classifiers on feature subsets of the Iris dataset



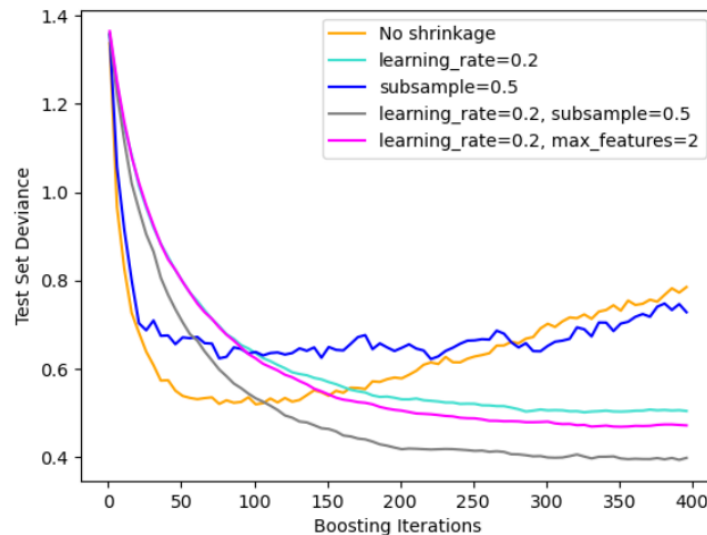
Berikut adalah penjelasan tentang Random Forest:

1. Ensemble Learning: Random Forest adalah bentuk ensemble learning di mana sejumlah besar pohon keputusan (yang dikenal sebagai pohon "anggota" atau "base") dibuat dan digabungkan. Setiap pohon memberikan prediksi, dan hasil akhir adalah agregasi prediksi dari semua pohon.
2. Bootstrap Sampling: Untuk setiap pohon anggota dalam Random Forest, data training diambil sebagai sampel secara acak dengan penggantian (bootstrap sampling). Ini berarti bahwa setiap pohon dibangun dengan dataset yang berbeda, dan ini membantu mengurangi overfitting.
3. Feature Randomization: Ketika membangun setiap pohon, hanya sebagian dari atribut (fitur) yang digunakan. Ini menghindari pohon-pohon menjadi terlalu spesifik terhadap atribut tertentu dan meningkatkan generalisasi.
4. Agregasi Prediksi: Prediksi dari setiap pohon digabungkan dengan berbagai cara tergantung pada jenis tugas. Misalnya, untuk klasifikasi, hasil yang paling umum diambil sebagai prediksi akhir (mode), sementara untuk regresi, rata-rata dari prediksi individu digunakan.
5. Keuntungan: Random Forest adalah algoritma yang sangat kuat untuk banyak tugas machine learning. Ini memiliki kemampuan untuk menangani dataset besar, menangani atribut dengan nilai yang hilang, dan mengatasi overfitting. Selain itu, Random Forest juga dapat memberikan informasi tentang kepentingan atribut dalam prediksi.

Gradient Boosting adalah salah satu teknik ensemble learning yang digunakan dalam machine learning untuk meningkatkan akurasi prediksi model. Metode ini memadukan sejumlah kecil model yang lebih sederhana (biasanya pohon keputusan) menjadi model



yang lebih kuat. Salah satu implementasi yang populer dari Gradient Boosting adalah Gradient Boosting Trees atau Gradient Boosting Machine (GBM).



Berikut adalah penjelasan tentang Gradient Boosting:

1. **Pohon Keputusan Sebagai Base Learner:** Dalam Gradient Boosting, pohon keputusan adalah base learner yang sering digunakan. Secara berurutan, pohon keputusan dibangun untuk mengoreksi kesalahan prediksi model sebelumnya.
2. **Kesalahan Residual:** Untuk setiap iterasi dalam algoritma Gradient Boosting, model mencoba memprediksi kesalahan (residual) dari model sebelumnya. Residual ini adalah perbedaan antara nilai sebenarnya dan prediksi sebelumnya.
3. **Pemberian Bobot:** Setiap model (pohon) yang dibangun memiliki bobot yang ditentukan berdasarkan sejauh mana model tersebut dapat memperbaiki kesalahan model sebelumnya. Model yang lebih baik dalam mengurangi kesalahan akan diberi bobot lebih besar.
4. **Agregasi Prediksi:** Prediksi dari semua model (pohon) digabungkan dengan mengambil jumlah atau rata-rata prediksi dari setiap pohon. Prediksi akhir adalah hasil agregasi dari semua iterasi.
5. **Keuntungan:** Gradient Boosting adalah teknik yang sangat kuat untuk berbagai tugas machine learning, termasuk klasifikasi dan regresi. Ia mampu mengatasi masalah overfitting, menghasilkan model yang kuat, dan memiliki fleksibilitas dalam penggunaan berbagai fungsi kerugian.

Random Forest adalah metode ensemble yang menggabungkan beberapa pohon keputusan (decision trees) menjadi "hutan" untuk menghasilkan prediksi yang lebih akurat. Ini dapat digunakan dalam berbagai konteks proyek analisis Big Data. Berikut adalah beberapa cara untuk dapat menggunakan Random Forest:



1. Prediksi Harga Aluminium (y) berdasarkan Faktor-faktor Eksternal

Random Forest digunakan untuk memprediksi Harga Aluminium berdasarkan faktor-faktor seperti Harga Copper, Baltic Dry Index, Premium Main Japan Port, Indeks USD, Produksi Aluminium Global, Konsumsi Aluminium Global, Stok Aluminium Dunia, Index Alumina, dan Harga Minyak. Random Forest akan menggabungkan sejumlah pohon keputusan untuk menghasilkan prediksi yang lebih stabil.

2. Klasifikasi Agreement Status

Random Forest dapat digunakan untuk mengklasifikasikan Agreement Status (status perjanjian) berdasarkan berbagai fitur seperti Vendor, Plant, Start Date, End Date, Remaining Days, dan Contract Utilization (%). Ini akan membantu dalam mengidentifikasi faktor-faktor yang mempengaruhi status perjanjian.

3. Analisis Spend Dashboard

Random Forest dapat digunakan untuk menganalisis Total Spend, Ratio Active Supplier/Registered Suppliers, dan faktor-faktor lain yang mempengaruhi pengeluaran perusahaan. Ini akan membantu dalam pemahaman pola pengeluaran.

Gradient Boosting adalah metode ensemble lain yang bekerja dengan cara menggabungkan sejumlah model prediktif secara berurutan dengan mengoreksi kesalahan model sebelumnya. Ini adalah teknik yang kuat untuk memperbaiki model secara berulang untuk meningkatkan akurasi. Berikut adalah beberapa contoh penggunaan Gradient Boosting:

1. Prediksi Harga Aluminium (y) berdasarkan Faktor-faktor Eksternal

Gradient Boosting digunakan untuk memodelkan hubungan antara Harga Aluminium dan faktor-faktor seperti Harga Copper, Baltic Dry Index, Premium Main Japan Port, Indeks USD, Produksi Aluminium Global, Konsumsi Aluminium Global, Stok Aluminium Dunia, Index Alumina, dan Harga Minyak. Gradient Boosting akan memperbaiki model secara berulang dan menghasilkan prediksi yang lebih akurat.

2. Analisis Vendor Delivery Performance

Gradient Boosting digunakan untuk menganalisis kinerja pengiriman vendor berdasarkan faktor-faktor seperti wilayah (Region) dan pabrik (Plant). Ini akan membantu dalam mengidentifikasi faktor-faktor kunci yang mempengaruhi pengiriman vendor.

3. Klasifikasi PR Composition by Status

Gradient Boosting dapat digunakan untuk mengklasifikasikan Purchase Requisition (PR) berdasarkan statusnya, seperti "approved," "pending," atau

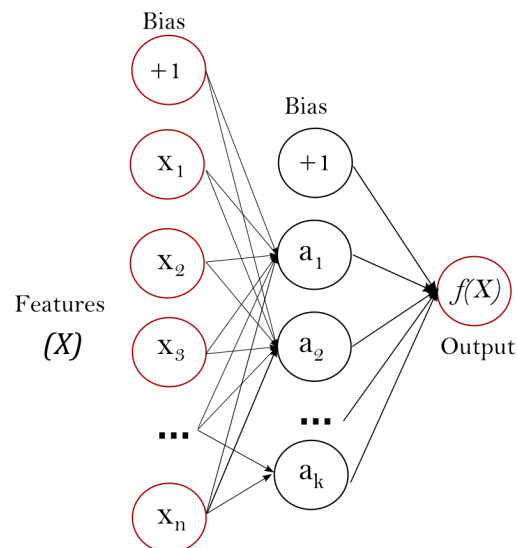


"rejected." Ini akan membantu dalam mengidentifikasi faktor-faktor yang mempengaruhi status PR.

Sementara Random Forest menggabungkan sejumlah model keputusan, Gradient Boosting fokus pada perbaikan berurutan dari model-model sebelumnya. Kedua metode ini adalah pilihan yang kuat untuk pemodelan dalam analisis Big Data, terutama ketika hubungan antara variabel tidak linier dan kompleks.

3.1.6. Neural Network Models (Supervised)

Neural Network Models (Model Jaringan Saraf) adalah sebuah pendekatan dalam machine learning dan deep learning yang terinspirasi oleh cara kerja otak manusia. Ini terdiri dari lapisan-lapisan (neuron) yang saling terhubung dan dapat mempelajari representasi fitur yang kompleks dari data. Neural network models dapat digunakan untuk berbagai tugas seperti klasifikasi, regresi, pengenalan gambar, pemrosesan bahasa alami, dan banyak lagi.



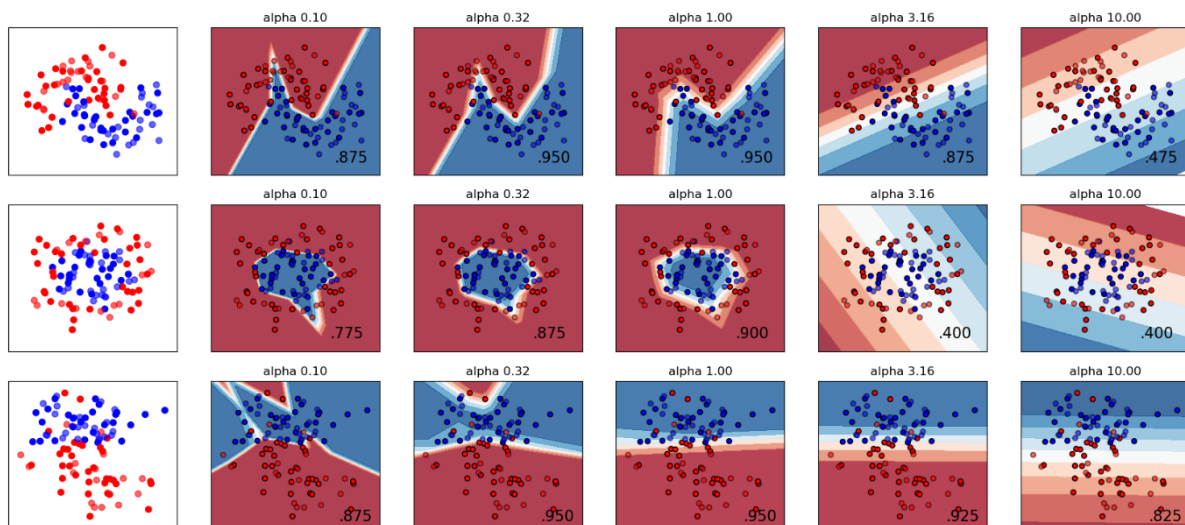
Berikut adalah penjelasan tentang Neural Network Models:

1. **Lapisan-Lapisan Neuron:** Neural network terdiri dari beberapa lapisan neuron, termasuk lapisan input, lapisan tersembunyi (hidden layers), dan lapisan output. Lapisan input menerima data masukan, lapisan tersembunyi berfungsi untuk memproses dan mempelajari fitur-fitur yang kompleks, dan lapisan output menghasilkan prediksi atau output.
2. **Bobot dan Bias:** Setiap koneksi antara neuron memiliki parameter yang disebut bobot (weight) dan bias. Selama pelatihan, bobot dan bias disesuaikan sehingga jaringan dapat mempelajari hubungan antara fitur input dan target output.
3. **Fungsi Aktivasi:** Setiap neuron memiliki fungsi aktivasi yang memutuskan apakah neuron tersebut akan aktif (mengirimkan sinyal) berdasarkan masukan yang diterimanya. Beberapa fungsi aktivasi umum termasuk sigmoid, ReLU (Rectified Linear Unit), dan tanh.



4. Pelatihan: Pelatihan jaringan saraf melibatkan proses meminimalkan kesalahan antara prediksi jaringan dan nilai target yang sebenarnya. Ini dilakukan dengan algoritma pembelajaran seperti backpropagation, yang menyesuaikan bobot dan bias untuk meminimalkan kesalahan.
5. Deep Learning: Deep learning adalah varian dari neural network yang memiliki banyak lapisan tersembunyi (deep neural network). Ini telah berhasil digunakan dalam banyak tugas, seperti pengenalan gambar, pengenalan wicara, dan pemrosesan bahasa alami.

Berikut adalah contoh visualisasi performance hasil klasifikasi untuk setiap nilai alpha:



Metode Neural Network Models (dalam konteks yang diawasi) adalah salah satu pendekatan yang kuat untuk analisis Big Data dan pemodelan prediktif yang kompleks. Dalam proyek ini, Neural Network Models dapat digunakan untuk berbagai tujuan analisis. Berikut adalah beberapa cara yang dapat diterapkan menggunakan Neural Network Models:

1. Prediksi Harga Aluminium (y) berdasarkan Faktor-faktor Eksternal

Neural Network Models digunakan untuk memodelkan hubungan antara Harga Aluminium dan faktor-faktor seperti Harga Copper, Baltic Dry Index, Premium Main Japan Port, Indeks USD, Produksi Aluminium Global, Konsumsi Aluminium Global, Stok Aluminium Dunia, Index Alumina, dan Harga Minyak. Ini akan memungkinkan untuk menghasilkan prediksi yang akurat dan menangkap hubungan non-linier yang mungkin ada di antara variabel-variabel ini.

2. Klasifikasi Agreement Status

Neural Network Models dapat digunakan untuk mengklasifikasikan Agreement Status berdasarkan berbagai fitur seperti Vendor, Plant, Start Date, End Date, Remaining Days, dan Contract Utilization (%). Dengan model ini, Anda dapat mengidentifikasi faktor-faktor yang mempengaruhi status perjanjian.



3. Analisis Vendor Delivery Performance

Neural Network Models dapat digunakan untuk menganalisis kinerja pengiriman vendor berdasarkan faktor-faktor seperti wilayah (Region) dan pabrik (Plant). Ini akan membantu dalam mengidentifikasi faktor-faktor kunci yang mempengaruhi kinerja vendor dalam pengiriman.

4. Prediksi Total Spend dan Analisis Pengeluaran

Neural Network Models dapat digunakan untuk memprediksi Total Spend dan menganalisis faktor-faktor yang mempengaruhi pengeluaran perusahaan, seperti rasio aktif supplier/registered suppliers, jumlah faktur, dan lainnya.

5. Klasifikasi PR Composition by Status

Neural Network Models dapat membantu dalam mengklasifikasikan Purchase Requisition (PR) berdasarkan statusnya, seperti "approved," "pending," atau "rejected." Ini dapat membantu dalam memahami faktor-faktor yang mempengaruhi status PR.

6. Analisis Trend dan Peramalan

Neural Network Models juga dapat digunakan untuk menganalisis tren dan peramalan berbagai metrik yang terdapat dalam dashboard, seperti Trend of Expenditing PO per Week, PR Trend, dan lainnya.

Dalam implementasi Neural Network Models, pengembang perlu melakukan praproses data, memilih arsitektur jaringan yang sesuai, dan melatih model dengan data historis yang relevan. Neural Network Models dapat menangani data yang kompleks dan berdimensi tinggi serta dapat mengidentifikasi pola dan hubungan yang rumit dalam data.

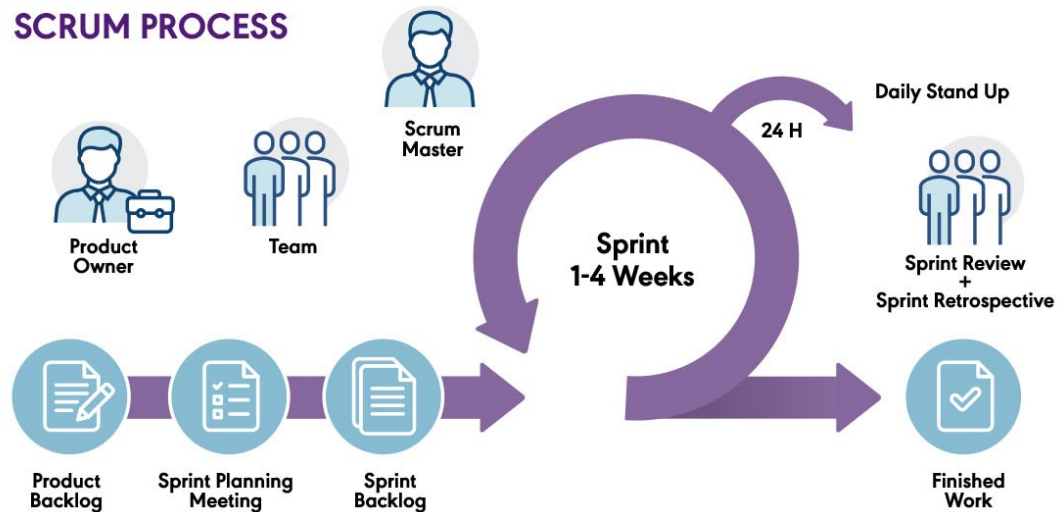
3.2. Metode Pengembangan Agile SDLC

Agile Scrum adalah kerangka kerja yang sangat relevan dalam pengembangan proyek Big Data, karena memungkinkan tim untuk mengatasi kompleksitas dan ketidakpastian yang sering terjadi dalam proyek-proyek ini. Dalam proposal teknis untuk proyek Big Data, penyusun dapat menjelaskan bagaimana Agile Scrum akan digunakan untuk mengelola dan mengembangkan solusi Big Data, serta bagaimana peran berbagai peran dalam tim sesuai dengan timeline yang telah diberikan.

Dalam proyek Big Data, tim akan mengikuti kerangka kerja Agile Scrum untuk mengelola pengembangan solusi Big Data yang kompleks. Pendekatan ini memungkinkan untuk merespons dengan cepat terhadap perubahan dan menyesuaikan solusi Big Data sesuai kebutuhan dan temuan selama proyek berlangsung.



SCRUM PROCESS



1. Iterasi Agile Scrum

a. Sprint 1-2: Persiapan Awal (Minggu 1-2)

Dalam Sprint ini, tim akan fokus pada pemahaman awal proyek Big Data dan kebutuhan data. Data Scientists dan AI Engineers akan mengidentifikasi sumber data dan mengumpulkannya di Cloud Storage. Ini adalah tahap awal yang penting untuk memastikan bahwa tim memiliki akses ke data yang diperlukan.

b. Sprint 3-4: Eksplorasi Data dan Transformasi (Minggu 3-4)

Pada Sprint ini, Data Scientists dan AI Engineers akan mulai menganalisis data dengan BigQuery. Mereka juga dapat menggunakan Cloud Functions untuk transformasi data jika diperlukan. Langkah ini membantu tim memahami data dan membangun dasar untuk model Machine Learning.

c. Sprint 5-6: Validasi dan Penyesuaian Model (Minggu 5-6)

Data Scientists akan menguji model yang telah dibangun dan melakukan fine-tuning jika diperlukan. Validasi model dengan data dari Pub/Sub dan Dataflow memastikan bahwa model dapat bekerja di lingkungan produksi.

d. Sprint 7-9: Implementasi dan Evaluasi (Minggu 7-9)

Tim akan menerapkan model ke Vertex AI Prediction untuk prediksi real-time. Selama tahap ini, tim akan menganalisis hasil prediksi dan mengukur kinerja model. Perbaikan model akan diterapkan berdasarkan hasil evaluasi.

e. Sprint 10-12: Dokumentasi dan Pelaporan (Minggu 10-12)



Sprint terakhir akan berfokus pada dokumentasi lengkap tentang seluruh proses proyek Big Data. tim akan membuat laporan yang menjelaskan temuan, solusi, dan kinerja model selama periode tiga bulan.

2. Kolaborasi Tim

Agile Scrum memungkinkan kolaborasi erat antara anggota tim yang berperan sebagai Data Scientists, AI Engineers, Data Analysts, Business Intelligence, dan Data Engineers. Masing-masing tim akan berkontribusi pada tahap-tahap yang berbeda dalam siklus pengembangan.

3. Fleksibilitas dalam Perubahan

Agile Scrum memungkinkan tim untuk merespons dengan cepat terhadap perubahan kebutuhan data atau solusi selama proyek berlangsung. Tim dapat melakukan penyesuaian dan fine-tuning berdasarkan temuan yang muncul selama pengembangan.

4. Transparansi dan Pelaporan Rutin

Agile Scrum mendorong transparansi dalam pengembangan proyek. Tim Data Analysts dan Business Intelligence akan mengatur pelaporan rutin untuk memberikan pemahaman berkelanjutan kepada pemangku kepentingan.

5. Monitoring dan Optimalisasi Kinerja

Data Engineers akan memonitor dan mengoptimalkan data pipeline yang kritis untuk suksesnya proyek Big Data. Mereka akan memastikan ketersediaan data yang terus menerus dan melakukan pengoptimalan jika diperlukan.

3.3. Manajemen Proyek

Rencana Tahapan Pekerjaan Pengadaan dan Pengembangan Sistem Data Warehouse dan Business Intelligence di PT INALUM:

1. Tahap I – Desain Pipeline dan Pengenalan Data (2 Minggu):
 - a. Mengidentifikasi sumber data, mengumpulkan dan menyimpan data awal.
 - b. Mempersiapkan data, mengimpor data ke BigQuery, membuat tampilan dashboard di Tableau.
 - c. Merancang pipeline data, mengidentifikasi sumber data, membangun infrastruktur di Compute Engine.
2. Tahap II – Konstruksi Pipeline dan Analisis Awal (2 Minggu):
 - a. Menganalisis data, melakukan transformasi jika diperlukan, dan membangun model Machine Learning dengan BigQuery ML.
 - b. Menganalisis data, memvisualisasikan hasil analisis di Tableau.
 - c. Menerapkan pipeline data, menghubungkan sumber data, memastikan transformasi dan pembersihan data.
3. Tahap III – Validasi dan Pemantauan (2 Minggu):
 - a. Menguji dan menyempurnakan model Machine Learning.



- b. Mengatur pelaporan rutin, memonitor data streaming, menyediakan laporan berkala.
 - c. Melakukan pemantauan rutin, memecahkan masalah, dan memelihara dokumentasi pipeline.
4. Tahap IV – Integrasi Data dan Analisis Tingkat Lanjut (3 Minggu):
 - a. Menerapkan model ke Vertex AI untuk prediksi waktu nyata, menganalisis hasil, dan melakukan evaluasi.
 - b. Melakukan analisis yang lebih mendalam, membuat rekomendasi berdasarkan data.
 - c. Mengintegrasikan pipeline dengan Cloud Functions untuk transformasi tambahan, menjaga sinkronisasi data.
5. Tahap V – Optimalisasi Hasil Akhir dan Pelaporan (3 Minggu):
 - a. Buat dokumentasi lengkap, laporan, dan sampaikan hasilnya kepada tim dan pemangku kepentingan.
 - b. Mengintegrasikan hasil analisis ke dalam laporan akhir, mengkomunikasikan temuan dan visualisasi kepada para pemangku kepentingan.
 - c. Menganalisis kinerja pipeline, melakukan optimasi, dan memberikan laporan mengenai kinerja dan perbaikan yang dilakukan

Dengan rencana tahapan ini, Penyedia Jasa akan menjalankan proses pengadaan dan pengembangan Descriptive and Predictive Analytic sesuai dengan jangka waktu **3 bulan** sejak tanggal kesepakatan kerja ditandatangani. Setiap tahap dirancang untuk mencapai tujuan tertentu dan memastikan kelancaran proyek serta keberlanjutan sistem setelah implementasi.

4. Hasil Pekerjaan

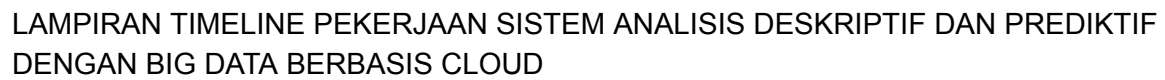
Hasil output dari proyek ini adalah sebuah dashboard analisis yang dirancang secara khusus untuk proposal teknis. Dashboard ini menyajikan laporan penjelasan yang komprehensif tentang hasil analisis yang telah dilakukan. Dengan menggunakan dashboard ini, pemangku kepentingan dapat dengan mudah melihat dan memahami data serta temuan kunci dari proposal teknis tersebut.

Dashboard ini memberikan visualisasi grafis yang informatif, menampilkan metrik kunci, dan mungkin juga menyediakan kemampuan untuk mengeksplorasi data lebih lanjut. Dengan adanya dashboard ini, proses evaluasi dan pengambilan keputusan terkait proposal teknis akan menjadi lebih efisien dan transparan.

No	Jenis Dashboard	Detail Analisis
1	Requisition Dashboard	<ul style="list-style-type: none">- Average Days PR to Award<ul style="list-style-type: none">- Average Days All Type- Average Service- Average Material- Average Days PR to Award Per Buyer- Purchase Requisition IDR- Purchase Requisition USD<ul style="list-style-type: none">- Purchase Requisition Quantity



		<ul style="list-style-type: none"> - Purchase Requisition Budget - PR Composition by Status - PR Composition by Category - PR Composition by Tender Method - PR Trend
2	Vendor Dashboard	<ul style="list-style-type: none"> - Top 10 Vendor <ul style="list-style-type: none"> - Platform - Category - Vendor Delivery Performance <ul style="list-style-type: none"> - Region - Plant - Vendor CSMS <ul style="list-style-type: none"> - Line of Business
3	Agreement Dashboard	<ul style="list-style-type: none"> - Agreement per Platform <ul style="list-style-type: none"> - USD - IDR - Agreement Status <ul style="list-style-type: none"> - Vendor - Plant - Agreement - Start Date - End Date - Remaining Days - Contract Utilization (%)
4	Expediting Dashboard	<ul style="list-style-type: none"> - Purchase Order <ul style="list-style-type: none"> - PO This Week - PO Last Week - PO Difference from Last Week - Trend of Expenditing PO per Week - Outstanding Maintenance Work Order <ul style="list-style-type: none"> - MWO 30 Days - MWO 60 Days - MWO 90 Days - Average lead Time Region
5	Spend Dashboard	<ul style="list-style-type: none"> - Total Spend - Ratio Active Supplier/Registered Suppliers - Active Suppliers - Registered Suppliers - Top 10 Spend by Supplier - Top 10 Spend by Category - Top 10 Spend by Non-PO - Total Number of Invoice - Monthly Number of Invoice - Total Count Spend

[illegible]



LAMPIRAN TOPOLOGI PEKERJAAN SISTEM ANALISIS DESKRIPTIF DAN PREDIKTIF DENGAN BIG DATA BERBASIS CLOUD

