

# Sistem Pendukung Keputusan Berbasis Metode Korelasi untuk Menganalisis Penyebab Tidak Tercapainya Target Produksi Batu Bara di PT XYZ

Wa Ode Zalmawati<sup>1)</sup>, Muhammad Nur Yasir Utomo<sup>2)</sup> Rini Nur<sup>3)</sup>

<sup>1,2,3</sup> Teknik Elektro, Politeknik Negeri Ujung Pandang  
waodezalmawati.dokumen@gmail.com<sup>1)</sup>, yasirutomo@poliupg.ac.id<sup>2)</sup>, rini@poliupg.ac.id<sup>3)</sup>

## Abstrak

Kegagalan pencapaian target produksi menjadi fenomena yang hampir tidak dapat dihindari dalam dunia bisnis, dimana hal tersebut tentunya akan bermuara pada kerugian bagi perusahaan terkait. Sangat penting bagi perusahaan untuk mengetahui penyebab kegagalan tercapainya target produksi agar dapat meminimalisir terjadinya potensi kerugian dikemudian hari. Untuk di PT XYZ sendiri terdapat departemen khusus yang bertugas untuk menganalisis penyebab tidak tercapainya target produksi. Namun berdasarkan hasil wawancara yang dilakukan bahwa hingga saat ini *insight* yang diperoleh tidak begitu mendukung, dikarenakan kualitas data yang tidak baik dan metode olah data yang diterapkan masih belum maksimal. Penelitian ini bertujuan untuk meng-*improve* kualitas data dengan serangkaian tahapan *data gathering*, *data preparation*, dan konsep ETL untuk mengungkap faktor-faktor penyebab tidak tercapainya target produksi batu bara pada PT XYZ. Penelitian ini berhasil meng-*improve* kualitas data pada perusahaan, dan banyak memperoleh temuan menarik dari data, sehingga berhasil mengungkap faktor-faktor penyebab tidak tercapainya target produksi batu bara di PT XYZ. Faktor-faktor penyebab tersebut dikategorikan berdasarkan nilai *feature importance*, koefisien korelasi, dan signifikansi (*P-Value*) pada variabel data yang berjumlah 28 *sublead*. Hasil dari seluruh tahapan analisis dituangkan dalam bentuk visualisasi data melalui *website* dengan *tool* Streamlit dan Python dengan data yang bersumber dari *graph database* Neo4J.

**Keywords:** *Data Preparation, ETL, Neo4J, Streamlit, Analisis Faktor Berpengaruh*

## I. PENDAHULUAN

PT XYZ merupakan salah satu perusahaan tambang batubara terbesar di Indonesia dengan berbagai mitra kontraktor yang bekerja untuk kegiatan operasional pertambangannya [1]. Para mitra kontraktor tersebut memiliki komitmen *plan/target* produksi yang telah disepakati bersama antara pihak perusahaan dengan mitra kontraktor itu sendiri, dimana target tersebutlah yang sedapat mungkin dicapai oleh mitra kontraktor pada periode tertentu sesuai dengan hasil kesepakatan di atas. Dalam kegiatan produksi tidak jarang target produksi harian yang telah ditetapkan tidak tercapai [2]. Target produksi dikatakan tidak tercapai ketika kondisi *plan* produksi lebih besar dari *actual* produksi, yang tentunya hal ini akan bermuara pada kerugian. Sangat penting bagi perusahaan untuk mengetahui penyebab tidak tercapainya target produksi agar dapat meminimalisir potensi kerugian terulang lagi di periode berikutnya.

Khusus di perusahaan batu bara sendiri, salah satu cara mengetahui penyebab tidak tercapainya target produksi adalah dengan menganalisis nilai *Effective Utilization* (EU) yang bertujuan untuk mengetahui seberapa efektifkah penggunaan alat yang digunakan dalam beroperasi [3]. Nilai EU dipengaruhi oleh dua variabel yaitu *Physical Availability* (PA) dan *Used Of Availability* (UA) [2]. Di PT. XYZ, yang bertanggung jawab untuk menganalisis penyebab ketidaktercapaian target produksi adalah *Production Control Departement* (*user*) yang hasil analisisnya rutin dipaparkan melalui *meeting* evaluasi mingguan bersama dengan manajernya setiap pekan.

Menurut hasil wawancara bersama dengan *subject matter expert* dari *user* yang dilakukan pada tanggal 7 Maret 2022 secara *daring*, diperoleh fakta bahwa “Untuk di PT. XYZ, belum melakukan analisis EU karena tidak lengkapnya data PA dan UA pada *database*. Sampai saat ini, *user* telah melakukan analisis deskriptif pada data UA, namun *insight* yang diperoleh tidak begitu mendukung terakik “mengapa target produksi tidak tercapai?”. Ternyata setelah diidentifikasi lebih lanjut, *insight* yang belum mendukung tersebut disebabkan oleh kualitas data *user* yang tidak baik, seperti tidak seragamnya kategori atribut *sulead indicator*, terlalu banyak *missing value*, terdapat data di beberapa kontraktor yang tidak lengkap, dan terdapat data keseluruhan pada periode 2019-2020 yang tidak lengkap. Perlu diketahui bahwa data UA terdiri atas 28 *sublead indicator* (penghambat produksi), diantaranya *Rain*, *Blasting*, dan *Slippery*. Seberapa besar dan signifikannya pengaruh variabel UA terhadap target produksi menjadi *goals* awal bagi *user* agar menjadi dasar dalam pengambilan keputusan yang lebih tepat guna.

Menurut [2] untuk mengetahui kuatnya hubungan antar variabel maka dapat dilihat dari besarnya angka *r* atau korelasi. Sedangkan untuk melihat seberapa signifikan pengaruh suatu variabel terhadap variabel lainnya dapat dilakukan dengan cara melihat nilai *P-Value* dari koefisien korelasi yang didapatkan [4]. Namun dengan kondisi data yang ada pada *user*, maka terlebih dahulu dilakukan serangkaian tahapan *data preparation* pada data. Oleh karena pada penelitian ini akan dilakukan perbaikan kualitas data sehingga dapat menghasilkan hasil analisis yang dituangkan pada *website* yang diharapkan mampu memudahkan *user* dalam memantau penyebab

target produksi tidak tercapai dan dapat membantu user dalam mengambil keputusan terkait proses bisnisnya.

## II. KAJIAN LITERATUR

### A. Data Preparation

Data Preparation adalah proses membersihkan dan mengubah data mentah sebelum diproses dan dianalisis. Tahapan ini adalah langkah terpenting dan sangat krusial sebelum pemrosesan dan sering kali melibatkan pemformatan ulang data, membuat koreksi pada data, dan menggabungkan kumpulan data untuk memperkaya insight awal data itu sendiri.

### B. Perhitungan Analisis Korelasi

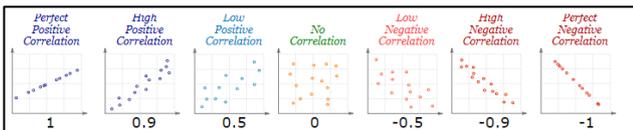
Analisis korelasi adalah salah satu teknik untuk mengetahui ada tidaknya hubungan yang menarik antar variabel dan membantu proses identifikasi relevansi antar atribut [7]. Dalam dunia bisnis, analisis korelasi adalah dalam proses bisnis yang harus diperbaiki (*improvement area*). Analisis korelasi membantu para pemimpin bisnis dalam membuat keputusan yang lebih berdampak pada bisnisnya berdasarkan pola dalam data [6][7].

Berikut adalah persamaan menghitung korelasi :

$$r = \frac{n \sum XY - (\sum X)(\sum Y)}{\sqrt{\{n \sum X^2 - (\sum X)^2\} \{n \sum Y^2 - (\sum Y)^2\}}} \quad (1)$$

Dengan :  $r$  = Koefisien Korelasi,  $n$  = Jumlah Data,  $x$  = Variabel Bebas,  $Y$  = Variabel Terikat.

Berikut ditampilkan gambar yang menunjukkan keeratan hubungan dan arah dari korelasi.



Gambar 1. Arah dan Hubungan Korelasi

Berdasarkan Gambar 1 diketahui bahwa nilai koefisien korelasi berada pada range -1 sampai 1. -1 berarti terdapat hubungan negatif sempurna, 0 berarti tidak memiliki hubungan sama sekali, dan 1 berarti memiliki hubungan positif sempurna. Semakin mendekati 1 nilai dari  $r$  maka semakin kuat hubungan antar variabel [2]. Korelasi memiliki kemungkinan pengujian secara dua arah. Apabila koefisien korelasi bernilai positif dikatakan korelasi searah yang bermakna jika variabel X naik maka variabel Y akan turun, dan sebaliknya jika koefisien korelasi bernilai negatif maka dikatakan korelasi tidak searah dengan artian jika variabel X naik maka variabel Y juga akan naik [2].

### C. Perhitungan Analisis Signifikansi

Menurut [4] dalam artikelnya menuliskan bahwa secara umum angka signifikansi sebesar 0,01; 0,05 dan 0,1. Kebanyakan peneliti menggunakan angka signifikansi 0,05, dengan ketentuan sebagai berikut.

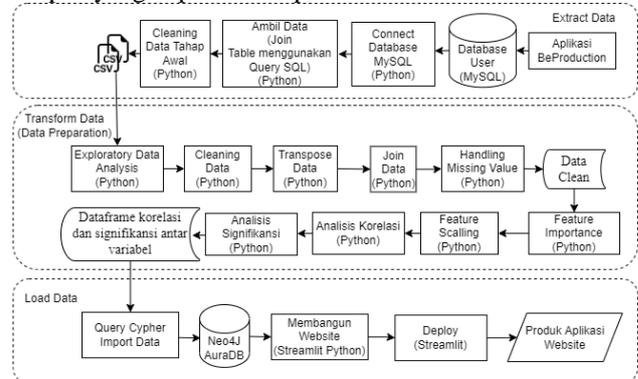
1. Jika angka signifikansi ( $\alpha$ ) < 0,05, maka hubungan kedua variabel signifikan.
2. Jika angka signifikansi ( $\alpha$ ) > 0,05, maka hubungan kedua variabel tidak signifikan.

### D. Graph Database dan Neo4J

Graph Database adalah database yang terdiri atas sekumpulan *node* dan *edge* yang merupakan salah satu bentuk dari *database Non-Relational* yang bersifat dinamis, dan mampu menyimpan, mengambil, serta memanipulasi data kompleks dengan waktu komputasi yang singkat [8].

## III. METODE PENELITIAN

Prosedur penelitian dilakukan melalui beberapa tahapan yang dapat dilihat pada Gambar 2 berikut.



Gambar 2. Prosedur Penelitian.

### A. Extract Data

#### 1. Connect Server Database MySQL

Tahapan ini bertujuan untuk menghubungkan *server database MySQL* aplikasi *BeProduction* ke perangkat lokal peneliti untuk mengakses data-data produksi user. Selain itu, pada tahapan ini juga peneliti akan melihat, mengamati, menelusuri, dan mempelajari struktur tabel pada data produksi.

#### 2. Ambil Data (Join Table Menggunakan Query SQL)

Pengambilan data dilakukan sekaligus dengan *join table* pada tabel *TU* dan *Prod*, dengan menggunakan *query SQL*. Perlu diketahui bahwa data *TU* dan *Prod* menjadi data utama pada penelitian.

#### 3. Cleaning Data Tahap Awal

Prosedur *cleaning* yang dilakukan adalah menghapus beberapa kolom yang *duplicate*, memperbaiki nama kolom, serta memperbaiki struktur data awal, kemudian diekspor menjadi data berformat *.csv*.

### B. Transform Data (Data Preparation)

#### 1. Exploratory Data Analysis (EDA)

Melakukan analisis deskriptif untuk memperoleh *insight* awal terhadap data olah.

#### 2. Cleaning Data

Menghapus duplikat data atau data yang tidak perlu dan memperbaiki struktur data. Prosedur data *cleaning* ini dilakukan untuk memastikan kualitas data yang digunakan[9][10].

#### 3. Transpose Data

Pada tahapan ini akan dilakukan penyesuaian atribut data dengan analisis yang dilakukan.

#### 4. Join Data

Menggabungkan data *TU* hasil *cleaning* dengan *Prod* berdasarkan atribut yang ditentukan.

#### 5. Handling Missing Value

Menghilangkan atau mengisi atribut yang nilainya masih kosong atau bernilai NULL atau NaN.

6. *Feature Importance*

Tahapan ini dilakukan untuk memilih fitur yang dianggap penting dari fitur-fitur yang ada.

7. *Feature Scalling*

*Feature Scalling* merupakan teknik statistik yang berfungsi untuk menormalisasikan *range* pada fitur-fitur data sehingga seluruh fitur berada pada *range* yang sama [11].

C. *Analisis Korelasi dan Signifikansi*

1. *Analisis Korelasi*

Mencari nilai koefisien korelasi untuk dijadikan sebagai patokan penentuan kategori *sublead*.

2. *Analisis Signifikansi*

Menentukan tingkatan signifikansi pengaruh *sublead* terhadap hasil produksi, dengan cara melihat nilai *P-Value* dari koefisien korelasi yang didapatkan.

D. *Load/Visualisasi Data*

1. *Import Data* ke Neo4J Desktop dan Neo4J AuraDB

*Import* data dilakukan pada dua platform Neo4J yakni *Desktop* dan *AuraDB*. Data yang di-*import* ke *database* adalah dataframe dengan pengelompokan korelasi dan signifikansi data pada masing-masing CG dan OB, yang akan ditandai dengan 6 node labels masing-masing CG dan OB. Selain itu, juga akan di-*import* dataframe detail *timeseries* kejadian *sublead* masing-masing CG dan OB.

2. *Membangun Website* dengan *Streamlit*

*Streamlit* dipilih karena cocok digunakan untuk studi kasus visualisasi data pada *web*.

3. *Deploy* Menggunakan *Streamlit*

Untuk melakukan proses *deploy* pada *streamlit*, terlebih dahulu melakukan *push* kode program ke github, kemudian mendaftarkan *link* github ke dalam *streamlit*.

IV. HASIL DAN PEMBAHASAN

A. *Hasil Extract Data*

Tahapan ini menghasilkan data TU dan Prod yang masing-masing data nya dapat dilihat pada Gambar 3 dan 4 berikut.

date	name_contractor	name_site	id_site	id_con	gain	lag_indicator	lead_indicator	sublead_indicator
2021-01-07	BUMA B7	BMO2	2	4	Loss	OB	UA	Rain
2021-01-07	BUMA B7	BMO2	2	4	Gain	OB	UA	Slippery

Gambar 3. Hasil Join Data TU

name_kontraktor	name_site	name_pit	t_date	id_site	id_con	id_pit	a_ob	a_cg
0	SIS 56	BMO1	2019-12-31	1	3	8	0.0	0.0
1	SIS 56	BMO1	2019-12-30	1	3	8	0.0	0.0

Gambar 4. Hasil Join Data Prod

B. *Hasil Transform Data*

1. *Hasil Exploratory Data Analysis (EDA)*

Tahapan ini dilakukan beberapa langkah seperti pengecekan jumlah baris dan kolom pada masing-masing dataframe TU dan Prod, memperbaiki tipe data atribut, memilih atribut penting pada data, memfilter kontraktor

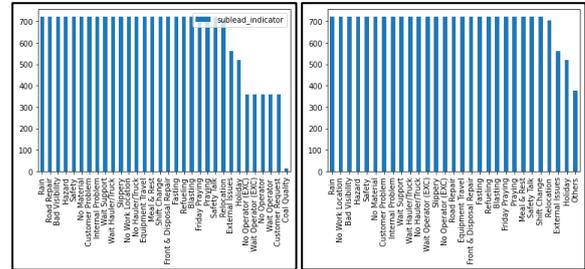
(id 4), melakukan identifikasi kecocokkan *id\_site* dengan *nama\_site*. Berikut adalah hasil EDA yang diperoleh pada salah satu data TU.

	date	id_site	id_con	name_site	name_contractor	lag_indicator	sublead_indicator	a_hr
0	2021-01-07	2	4	BMO2	BUMA B7	OB	Rain	7.341963
1	2021-01-07	2	4	BMO2	BUMA B7	OB	Slippery	0.965241

Gambar 5. Hasil EDA Data TU

2. *Hasil Cleaning Data*

Tahapan ini bertujuan untuk memperbaiki kualitas data yang digunakan. Berikut adalah hasil yang diperoleh.



Gambar 6. Perbandingan Sebelum (Kiri) dan Sesudah (Kanan) Penyeragaman Sublead

3. *Hasil Transpose Data*

Dalam tahapan ini dilakukan *transpose* 28 baris *sublead indicator* dijadikan sebagai atribut, dan nilai untuk masing-masing 28 atribut tersebut adalah lamanya *sublead* terjadi dalam satuan jam. Berikut adalah hasil yang diperoleh.

date	id_site	id_con	name_site	name_contractor	lag_indicator	sublead_indicator	Bad Visibility	Blasting	Customer Problem	Equipment Travel
2021-01-01	2	4	BMO2	BUMA B7	CG	OB	0.0	0.000000	0.237281	0.000000
							0.0	0.000000	0.000000	0.000000

Gambar 7. Hasil Transpose Data

4. *Hasil Join Data*

Pada tahapan ini dilakukan join data antara hasil pada proses *transpose* data TU dengan hasil proses EDA data Prod. Berikut adalah hasil yang diperoleh.

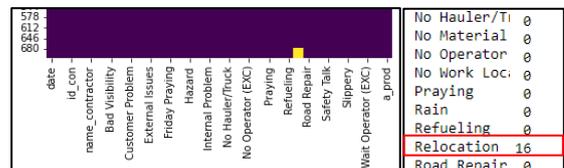
Road Repair	Safety	Safety Talk	Shift Change	Slippery	Hauler/Truck	Wait Operator (EXC)	Wait Support	a_prod
0.000000	0.0	0.000000	0.548246	0.000000	0.073684	0.015789	0.000000	17453.150
0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000

Gambar 8. Hasil Join Data

Berdasarkan Gambar 7 dapat dilihat bahwa setiap hari terdapat dua *lag indicator* (CG da OB) pada data, maka hasil dari *a\_prod* akan terbagi mengikut ke masing-masing *lag indicator*-nya, yakni CG dan OB. Sehingga jumlah data Prod sekarang sama dengan jumlah data TU, yakni 710 baris x 35 kolom, dari yang sebelumnya data Prod berjumlah 4.367 baris x 31 kolom.

5. *Hasil Handling Missing Value*

Berikut adalah hasil pengecekan *missing value* pada data.



Gambar 9. Pengecekan Missing Value

Berdasarkan Gambar 9 dapat dilihat bahwa terdapat atribut *sublead indicator* “Relocation” dengan *missing value* sejumlah 16 baris. Jika dikaitkan dengan proses bisnis yang ada, maka teknik *imputasi missing value* yang cocok adalah *imputasi nol*, dengan alasan adanya indikasi bahwa baris yang kosong pada *sublead* “Relocation” bermakna memang *sublead* “Relocation” dihari itu tidak terjadi (0 jam). Berikut adalah hasil proses *imputasi nol* yang diperoleh.

Gambar 10. Hasil *Imputasi Nol*

6. Hasil *Feature Importance*

Tahapan ini dilakukan untuk memilih 10 *sublead* penting dari total 28 *sublead* yang ada. Berikut adalah hasil dari pencarian nilai *importance* 28 *sublead* yang diperoleh.

```
[0.01986954 0.0049813 0.09537007 0.00027575 0.01008703 0.02917692
0.01218592 0.01567807 0. 0. 0.10480447 0.03443778 0.04979438
0.02888378 0.08378116 0.02091425 0.0048416 0.01061822 0.19820671
0.00545335 0.00365098 0.00278819 0. 0.02256369 0.05596732
0.0672062 0.03242679 0.0788626 0.00717395]
```

Gambar 11. Hasil Nilai *Importance* pada Data CG

```
[0.00675959 0.01126756 0.0008918 0.03737579 0.00105305 0.00622033
0.00420246 0.01351777 0.00139711 0.00816238 0.02534833 0.04770656
0.01030866 0.00117248 0.00930909 0.00239767 0.19845502 0.39781003
0.00911411 0. 0.00702541 0. 0.01066361 0.0745646
0.06642388 0.02021619 0.01296138 0.01647515]
```

Gambar 12. Hasil Nilai *Importance* pada Data OB

Berikut ditampilkan Tabel hasil 10 *sublead* yang memiliki nilai *importance* yang tinggi.

Tabel 1. Hasil *Feature Importance* CG

No	Sublead	Nilai <i>Feature Importance</i>
1.	Rain	0.19820671
2.	Holiday	0.10480447
3.	Customer Problem	0.09537007
4.	No Material	0.08378116
5.	Internal Problem	0.03443778
6.	Wait Operator (EXC)	0.0788626
7.	Slippery	0.0672062
8.	Shift Change	0.05596732
9.	Meal & Rest	0.04979438
10.	Wait Hauler/Truck	0.03242679

Table 2. Hasil *Feature Importance* OB

No	Sublead	Nilai <i>Feature Importance</i>
1.	Rain	0.39701003
2.	Praying	0.19845502
3.	Shift Change	0.0745646
4.	Slippery	0.06642388
5.	Meal & Rest	0.04770656
6.	Equipment Travel	0.03737579
7.	Internal Problem	0.02534833
8.	Wait Hauler/Truck	0.02021619
9.	Wait Support	0.01647515
10.	Front & Disposal Repair	0.01351777

Berdasarkan hasil pada tahapan *feature importance* yang diperoleh, dapat dilihat bahwa untuk kategori CG yang signifikan penting nilai *sublead*-nya adalah Rain dengan nilai fitur sebesar 0.19820671, kemudian menyusul Holiday dengan nilai fitur 0.10480447, dan

Customer Problem dengan nilai fitur sebesar 0.09537007, dst. Sedangkan untuk kategori OB, *sublead* yang signifikan penting adalah juga Rain dengan nilai fitur sebesar 0.39701003, kemudian Praying dengan nilai fitur sebesar 0.19845502, dan Shift Change dengan nilai fitur sebesar 0.0745646, dst.

7. Hasil *Feature Scalling*

*Scalling* dilakukan dengan metode *MinMaxScaler*. Tahapan ini bertujuan untuk menyamakan skala pada data. Berikut adalah perbandingan data sebelum dan sesudah dilakukan *scalling*.

Wait Operator (EXC)	Slippery	Shift Change	Meal & Rest	Internal Problem	Wait Hauler/Truck	a_prod
0.015789	0.000000	0.548246	0.551316	0.038842	0.073684	17453.15
0.664035	0.039474	1.325000	1.530702	0.185088	0.673246	19747.05

Gambar 13. Data Sebelum Dilakukan *Scalling*

Wait Operator (EXC)	Slippery	Shift Change	Meal & Rest	Internal Problem	Wait Hauler/Truck	a_prod
0.007367	0.000000	0.112482	0.189625	0.014820	0.009565	0.678086
0.309832	0.008907	0.271847	0.526483	0.074455	0.087398	0.767208

Gambar 14. Data Sebelum Dilakukan *Scalling*

C. Hasil Analisis Korelasi dan Signifikansi

1. Hasil Analisis Koefisien Korelasi

Berikut ditampilkan Tabel matriks hasil koefisien korelasi yang diperoleh.

Tabel 3. Matrik Hasil Koefisien Korelasi CG

No.	Sublead	Koefisien Korelasi
1.	Bad Visibility	0.03
2.	Blasting	0.01
3.	Customer Problem	-0.23
4.	Equipment Travel	0.07
5.	External Issues	-0.09
6.	Fasting	-0.15
7.	Friday Praying	-0.02
8.	Front & Disposal Repair	-0.02
9.	Hazard	0
10.	Holiday	-0.24
11.	Internal Problem	0.21
12.	Meal & Rest	0.15
13.	No Hauler/Truck	0.04
14.	No Material	-0.05
15.	No Operator (EXC)	0.16
16.	No Work Location	-0.02
17.	Praying	-0.01
18.	Rain	-0.34
19.	Refueling	0.04
20.	Relocation	0.01
21.	Road Repair	0.08
22.	Safety	0
23.	Safety Talk	-0.01
24.	Shift Change	0.23
25.	Slippery	-0.22
26.	Wait Hauler/Truck	0.21
27.	Wait Operator (EXC)	0.32
28.	Wait Support	-0.00

Berikut adalah hasil korelasi berdasarkan matrik yang didapatkan pada data CG.

- 1.) Korelasi Negatif = Rain (-0.34); Holiday (-0.24); Customer Problem (-0.23); Slippery (-0.22); Fasting (-0.15).

2.) Korelasi positif = Wait Operator (EXC) (0.32); Shift Change (0.23); Internal Problem (0.21); Wait Hauler/Truck (0.21); No Operator (0.16).

Tabel 4. Matrik Hasil Koefisien Korelasi OB

No.	Sublead	Koefisien Korelasi
1.	Bad Visibility	-0.12
2.	Blasting	0.00
3.	Customer Problem	0.06
4.	Equipment Travel	0.51
5.	External Issues	-0.08
6.	Fasting	-0.11
7.	Friday Praying	-0.05
8.	Front & Disposal Repair	0.06
9.	Hazard	0.09
10.	Holiday	-0.19
11.	Internal Problem	0.27
12.	Meal & Rest	0.46
13.	No Hauler/Truck	0.17
14.	No Material	0.07
15.	No Operator (EXC)	-0.05
16.	No Work Location	0.00
17.	Praying	0.60
18.	Rain	-0.74
19.	Refueling	0.03
20.	Relocation	0
21.	Road Repair	-0.09
22.	Safety	0
23.	Safety Talk	0.06
24.	Shift Change	0.55
25.	Slippery	-0.58
26.	Wait Hauler/Truck	0.14
27.	Wait Operator (EXC)	0.21
28.	Wait Support	-0.13

Berikut adalah hasil korelasi berdasarkan matrik yang didapatkan pada data OB.

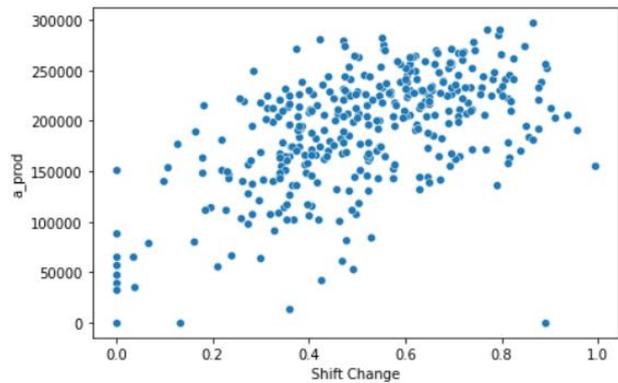
1.) Korelasi negatif

- Korelasi tinggi negatif ( di atas 0.5 ) : Rain (-0.74), dan Slippery (-0.58)
- Korelasi rendah negatif (di bawah 0.5) : Holiday (-0.19).

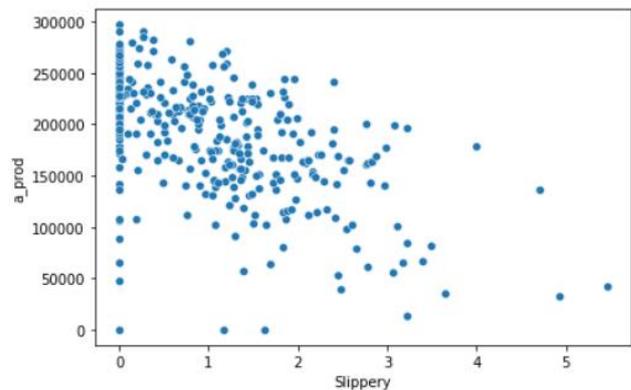
2.) Korelasi positif

- Korelasi tinggi positif ( di atas 0.5 ) : Praying (0.60), Shift Change (0.55), Equipment Travel (0.51).
- Korelasi rendah positif (di bawah 0.5) : Meal & Rest (0.46), Internal Problem (0.27).

Untuk lebih jelasnya dilakukan *plotting scatter* yang salah satu hasilnya seperti berikut.



Gambar 15. Gambar Visualisasi Scatter Plot Korelasi Positif



Gambar 16. Gambar Visualisasi Scatter Plot Korelasi Negatif

Berdasarkan gambar visualisasi di atas, dapat ditarik *insight* bahwa semakin tinggi lama kejadian *Shift Change*, maka semakin tinggi pula hasil produksi. Sama halnya dengan *Meal & Rest* pada data CG di atas bahwa hal tersebut tentunya menjadi janggal karena tidak mungkin semakin lama waktu pergantian *shift*, maka semakin tinggi pula hasil produksi. Sedangkan untuk *sublead Slippery*, semakin lama kejadian jalanan licin (*Slippery*), maka hasil produksi semakin rendah. Hal tersebut selaras dengan proses bisnis yang ada.

2. Hasil Analisis Signifikansi

Pada tahapan ini dilakukan pengkategorian *sublead* berdasarkan nilai korelasi dan signifikansi (nilai *P-Value*) pada masing-masing *lag indicator* CG dan OB terhadap hasil produksi batu bara dengan alur seperti berikut.

- 1.) Jika koefisien korelasi bernilai negatif dan nilai *P-Value* bernilai signifikan (kurang dari atau sama dengan 0.05), maka dimasukkan dalam kategori *sublead* signifikan.
- 2.) Jika koefisien korelasi bernilai positif dan nilai *P-Value* bernilai signifikan (kurang dari atau sama dengan 0.05), maka dimasukkan dalam kategori *sublead* kurang signifikan.
- 3.) Jika koefisien korelasi bernilai negatif dan nilai *P-Value* bernilai tidak signifikan (lebih dari atau sama dengan 0.05), maka dimasukkan dalam kategori tidak signifikan.

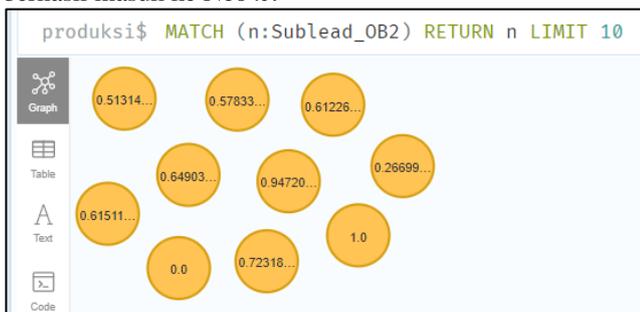
Tabel 5 di bawah ini menunjukkan hasil kategori *sublead*.

Tabel 5. Hasil Kategori *Sublead*

No.	Lag Indikator	Sublead		
		Signifikan	Kurang Signifikan	Tidak Signifikan
1.	CG	Rain, Holiday, Customer Problem, Slippery	Wait Operator, Shift Change, Internal Problem, Wait Hauler, Meal & Rest	No Material
2.	OB	Rain, Slippery, Wait Support	Praying, Shift Change, Equipment Travel, Meal & Rest, Internal Problem, Wait Hauler	Front & Disposal Repair

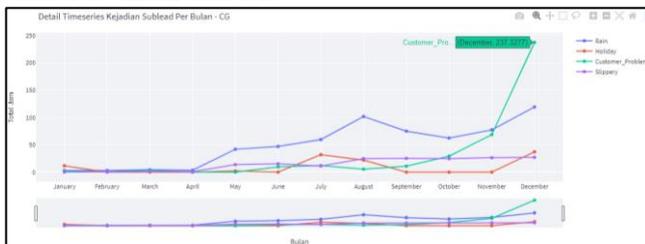
D. Hasil Load Data/Visualisasi Data

Berikut adalah hasil dari salah satu *node label* yang berhasil masuk ke Neo4J.



Gambar 17. Output Query Import Data pada Neo4J

Berikut adalah salah satu hasil visualisasi dataframe *timeseries* untuk data CG.



Gambar 18. Hasil Visualisasi *Timeseries*

V. KESIMPULAN

1. Penelitian ini berhasil melakukan seluruh tahapan *data preparation* yang terbukti mampu meningkatkan kualitas data yang digunakan.
2. Diperoleh nilai koefisien korelasi masing-masing data CG dan OB untuk melihat arah dan hubungan antara *sublead* dengan hasil produksi
3. Diperoleh 10 *sublead* yang paling berperan penting dalam mempengaruhi target produksi berdasarkan nilai *importance*-nya pada masing-masing *lag indicator* CG dan OB.
4. Diperoleh tiga kategori signifikansi pengaruh *sublead* terhadap target produksi pada masing-masing *lag indicator* CG dan OB.
5. Penelitian ini berhasil membangun dan mengimplementasikan penggunaan *graph database* sebagai media penyimpanan data
6. Berhasil membuat sistem aplikasi *website* menggunakan *Streamlit Python* dengan sumber data yang terhubung pada *graph database*.

UCAPAN TERIMA KASIH

Ucapan terima kasih penulis kepada orang tua, teman penulis, kedua dosen pembimbing, seluruh dosen dan staff terkhusus di Teknik Elektro Program Studi Teknik Komputer dan Jaringan Politeknik Negeri Ujung Pandang.

REFERENSI

- [1] H. Maraden, T. Redeb, K. Berau, and P. Batubara, "PENAMBANGAN DI PT BERAU COAL," vol. 1, pp. 138–144, 2021.
- [2] M. A. Alfarizi, N. Nurhakim, and R. N. Hakim, "Hubungan Physical Availability Dan Used of Availability Terhadap Overburden Removal Di Pt Semesta Centramas," *J. Himasapta*, vol. 5, no. 2, p. 29, 2020, doi: 10.20527/jhs.v5i2.2338.
- [3] A. Suryadi, N. Nurhakim, and R. N. Hakim, "Analisis Pengaruh Effective Utilization Terhadap Ketercapaian Produksi Batubara Pt Semesta Centramas," *Jurnal GEOSAPTA*, vol. 5, no. 1. p. 19, 2019, doi: 10.20527/jg.v5i1.5674.
- [4] Binus, "Alpha dan p Value – Accounting," 2020. <https://accounting.binus.ac.id/2020/12/19/alpha-dan-p-value/>.
- [5] S. Kumar and I. Chong, "Correlation analysis to identify the effective data in machine learning: Prediction of depressive disorder and emotion states," *Int. J. Environ. Res. Public Health*, vol. 15, no. 12, 2018, doi: 10.3390/ijerph15122907.
- [6] www.researchoptimus.com, "Correlation and Regression Analysis for Decision Making," 2022. <https://www.researchoptimus.com/article/what-is-correlation.php>.
- [7] I. Syamsuddin, R. Nur, H. Nirwana, I. Abduh, and D. Al-Dabass, "Decision making analysis of video streaming algorithm for private cloud computing infrastructure," *Int. J. Electr. Comput. Eng.*, vol. 7, no. 6, pp. 3529–3535, 2017, doi: 10.11591/ijece.v7i6.pp3529-3535.
- [8] J. J. Miller, "Graph database applications and concepts with Neo4j," *Proc. South. Assoc. Inf. Syst. Conf. Atlanta, GA, USA*, vol. 2324, p. 36, 2013.
- [9] R. Rosdiana, T. Eddy, S. Zawiyah, and N. Y. U. Muhammad, "Analisis Sentimen pada Twitter terhadap Pelayanan Pemerintah Kota Makassar," *Proceeding SNTEI*, no. June 2020, pp. 87–93, 2019.
- [10] M. N. Y. Utomo, M. Sudaryanto, and K. Saddhono, "Tools and Strategy for Distance Learning to Respond COVID-19 Pandemic in Indonesia," *Ing. des Syst. d'Information*, vol. 25, no. 3, pp. 383–390, 2020, doi: 10.18280/isi.250314.
- [11] Fiana, "Memahami Numerical Feature Scaling Untuk Meningkatkan Performace Machine Learning," 2021. <https://medium.com/easyread/memahami-semua-tentang-numerical-feature-scaling-dalam-8-menit-d79c65e0ccb6>.