

Analysis of the Relationship Between Mix Parameters and Age on Concrete Compressive Strength



Naila Saifana Santoso^{1*}, Katherine Debora Tampubolon², Siti Nur Aisyah²

¹Physics Education, Faculty of Mathematics and Natural Sciences, Universitas Negeri Jakarta

²Department of Architecture, Faculty of Engineering, Universitas Indonesia, Indonesia

ABSTRACT

This manuscript investigates the influence of material composition and concrete age on compressive strength using the Concrete Mixes dataset, which includes variables such as cement, water, fine aggregate, coarse aggregate, fly ash, slag, superplasticizer, concrete age, and compressive strength. Descriptive statistical analysis reveals a strong positive correlation between cement content and compressive strength (+0.85). In contrast, water content shows a negative correlation (-0.55), indicating that increased water content reduces compressive strength. The Random Forest Regressor model used in this study performs excellently with an RMSE of 5.02 MPa and an R-squared of 0.91, indicating the model's ability to predict concrete strength accurately. Furthermore, the analysis of concrete age shows significant strength improvement at 28 days, although the rate of increase slows down after that. The use of additives such as fly ash and slag positively impacts concrete durability, but their use must be carefully managed to avoid reducing early-age strength. The findings from this study provide valuable insights for designing more efficient, environmentally friendly, and durable concrete mixtures, considering optimal material ratios and curing time to achieve maximum concrete strength.

ABSTRAK

Penelitian ini menginvestigasi pengaruh komposisi material dan usia beton terhadap kekuatan tekan beton dengan menggunakan dataset Concrete Mixes yang mencakup variabel seperti semen, air, agregat halus, agregat kasar, fly ash, slag, superplasticizer, usia beton, dan kekuatan tekan beton. Analisis statistik deskriptif menunjukkan bahwa kandungan semen (cement) memiliki korelasi positif yang sangat kuat dengan kekuatan tekan beton (strength) sebesar +0.85, sedangkan kandungan air (water) menunjukkan korelasi negatif (-0.55), yang mengindikasikan bahwa peningkatan kadar air akan menurunkan kekuatan beton. Model Random Forest Regressor yang digunakan dalam penelitian ini menunjukkan performa yang sangat baik dengan nilai RMSE sebesar 5.02 MPa dan R-squared sebesar 0.91, yang menandakan kemampuan model dalam memprediksi kekuatan beton secara akurat. Selain itu, analisis terhadap usia beton (age) menunjukkan bahwa kekuatan beton mengalami peningkatan yang signifikan pada usia 28 hari, meskipun laju peningkatannya mulai melambat setelah itu. Penggunaan bahan tambahan seperti fly ash dan slag memberikan pengaruh positif terhadap ketahanan beton, namun penggunaannya harus diperhatikan dengan hati-hati agar tidak mengurangi kekuatan pada usia awal. Hasil penelitian ini memberikan wawasan yang sangat berguna untuk merancang campuran beton yang lebih efisien, ramah lingkungan, dan tahan lama, dengan mempertimbangkan rasio material yang optimal dan waktu curing yang tepat untuk mencapai kekuatan beton yang maksimal.

INTRODUCTION

Bangunan dengan konstruksi yang baik akan memperkecil kerugian yang diderita ketika bencana terjadi dan akan memberikan keamanan lebih. Salah satu penyebab besar kerusakan yang terjadi setelah bencana gempa adalah struktur bangunan yang tidak sesuai dengan standar keamanan gempa bumi. Untuk meminimalkan korban dan kerugian materiil saat terjadinya gempa, salah satu cara yang dapat dilakukan adalah membangun bangunan dengan sistem konstruksi yang baik. Hal-hal yang perlu diperhatikan dalam membuat konstruksi bangunan rumah yang baik setidaknya mencakup tiga aspek utama, yaitu pondasi, beton, dan beton bertulang. Ketiga hal tersebut seringkali diabaikan oleh beberapa kalangan masyarakat. Hal ini umumnya terjadi karena kurangnya pemahaman atau informasi mengenai konstruksi rumah yang baik (Rumbyarso, 2022). Beton merupakan material konstruksi yang telah dikenal dan dipakai sejak ribuan tahun lalu. Meskipun istilah semen Portland baru dikenal pada abad 19, namun bangunan dengan

CONTACT

nailasaifanasantoso@gmail.com

KEYWORDS

Concrete Age, Concrete Mix Design, Compressive Strength, Material Composition, Predictive Modeling

menggunakan beton sudah dikenal sejak zaman romawi, Colosseum di Roma atau Pont Du Gard di Prancis. Pada abad ke 17, perkembangan beton mengalami peningkatan dari bahan-bahan pembentuknya, terutama semen. (Rio Herdianto Rahamudin, Heyco Manalip, Mielke Mondoringin 2016). Beton dibuat dengan mencampurkan agregat halus (pasir), agregat kasar (kerikil), semen, dan air. Campuran tersebut akan mengeras menyerupai batuan. Proses pengerasan terjadi akibat reaksi kimia antara semen dan air. Dalam campuran beton, seringkali ditambahkan bahan tambahan kimia dengan perbandingan tertentu, sehingga membentuk satu kesatuan homogen yang mengeras seperti batuan (Septiani, 2024).

Variasi dalam proporsi campuran beton dapat menghasilkan kekuatan tekan beton yang berbeda-beda. Salah satu jenis campuran yang sering digunakan dalam pembuatan beton melibatkan perbandingan antara semen, pasir, dan kerikil (Ahmad et al., 2022). Kekuatan tekan beton dihasilkan dari pengaturan perbandingan semen, agregat kasar dan halus, serta air. Jika rasio air semen meningkat maka kekuatannya akan menurun dan begitu pula sebaliknya. Hal ini berhubungan langsung dengan workability karena jika rasio air semen meningkat maka beton akan lebih basah dan lebih konsisten serta dapat dicor dengan mudah (Septiani, 2024). Kuat tekan beton yaitu besarnya beban per satuan luas, yang mengakibatkan benda uji beton hancur jika dibebani gaya tekan tertentu yang dihasilkan oleh mesin tekan. Kuat tekan beton ialah sifat terpenting dalam kualitas beton dibanding dengan sifat-sifat lainnya (Widyanto, 2016).

Kekuatan tekan beton dapat dicapai sampai $1000 \text{ kg/cm}^2/98,0665 \text{ Mpa}$ atau lebih, tergantung pada jenis campuran, sifat-sifat agregat, serta kualitas perawatan. Kekuatan tekan beton yang paling umum digunakan adalah sekitar $2200-500 \text{ kg/cm}^2$. Nilai kuat tekan beton didapatkan melalui tata cara pengujian standar, menggunakan mesin uji dengan cara memberikan beban tekan bertingkat dengan kecepatan peningkatan beban tertentu dengan benda uji berupa kubus $150 \times 150 \text{ mm}$ (Sulianti, 2018). Semakin tinggi nilai FAS (Faktor Air Semen), mengakibatkan penurunan mutu kekuatan beton. Namun nilai FAS yang semakin rendah tidak selalu berarti bahwa kekuatan beton semakin tinggi. Jika FAS semakin rendah, maka beton akan semakin sulit untuk dipadatkan (Widyanto, 2016). Menurut Kamus Besar Bahasa Indonesia semen adalah serbuk atau tepung yang terbuat dari kapur dan material lainnya yang dipakai untuk membuat beton, merekatkan batu bata ataupun tembok (KBBI, 2023). Semen dalam pengertian umum adalah bahan yang mempunyai sifat adhesive dan cohesive, digunakan sebagai bahan pengikat (bonding material), yang dipakai bersama-sama dengan batu kerikil dan pasir (Riski, 2019). Untuk menghasilkan beton dengan kekompakan yang baik, diperlukan gradasi agregat yang baik. Gradasi agregat adalah distribusi ukuran kekerasan butiran agregat. Gradasi diambil dari hasil pengayakan. Cara membedakan jenis agregat yang paling banyak dilakukan adalah dengan berdasarkan pada ukuran butiran-butirannya yang diuji melalui analisa saringan. Agregat yang mempunyai butiran-butiran yang besar disebut agregat kasar yang ukurannya lebih besar dari $4,75 \text{ mm}$ atau tertahan pada saringan no. 4 yang berupa batu pecah (split) dan kerikil. Sedangkan butiran agregat yang kecil disebut agregat halus yang memiliki ukuran lebih kecil dari $4,75 \text{ mm}$ atau lolos saringan no. 4 yang berupa pasir dan material halus lolos saringan lainnya (Sulianti, 2018). Semakin besar butir agregat kasar yang digunakan maka kuat tekan yang dihasilkan akan semakin besar. agregat kasar yang tidak terpilih dengan baik atau penggunaan bahan tambahan seperti fly ash dan slag yang tidak tepat proporsinya dapat mempengaruhi karakteristik beton, meskipun campuran awal sudah sesuai standar. Hal ini menjadi alasan utama mengapa pemahaman yang lebih dalam tentang hubungan antara komposisi material dan usia beton sangat diperlukan untuk merancang campuran beton yang lebih berkualitas dan tahan lama (Taufik & Muhammad, 2020).

Usia beton berhubungan langsung dengan proses hidrasi semen yang berlangsung selama waktu curing. Semakin lama beton dipelihara dengan benar, semakin kuat beton tersebut, karena reaksi kimia antara air dan semen terus berlangsung. Kurniawan dan Setiawan (2020) dalam penelitian mereka tentang pengaruh curing terhadap kekuatan beton menjelaskan bahwa pada minggu pertama hingga minggu keempat, kekuatan tekan beton meningkat pesat, namun setelah itu laju peningkatannya mulai melambat. Hal ini sangat penting dalam perencanaan jangka panjang, terutama pada proyek besar seperti pembangunan gedung bertingkat atau infrastruktur transportasi. Penelitian yang dilakukan oleh Riski (2019) dan Widyanto (2016) menunjukkan bahwa faktor-faktor lingkungan, seperti suhu dan kelembaban saat proses curing, sangat mempengaruhi hasil kekuatan beton. Peningkatan suhu curing, misalnya, dapat mempercepat hidrasi semen dan meningkatkan kekuatan tekan pada beton muda, tetapi jika suhu terlalu tinggi, dapat menyebabkan retak pada beton karena kehilangan air terlalu cepat. Ini menegaskan pentingnya kontrol terhadap faktor-faktor lingkungan pada proyek konstruksi yang memanfaatkan beton sebagai material utama.

Di sisi lain, penggunaan bahan tambahan seperti fly ash, slag, dan superplasticizer semakin populer dalam campuran beton untuk meningkatkan kekuatan tekan dan daya tahan beton terhadap kondisi lingkungan yang ekstrim. Namun, Tarek (2020) menemukan bahwa meskipun fly ash dapat memperbaiki durabilitas beton, penggunaan yang berlebihan dapat menurunkan kekuatan beton pada usia awal, meskipun pada usia lanjut, fly ash dapat meningkatkan kekuatan beton secara signifikan. SNI 03-2834-2000 tentang Persyaratan Beton Struktural memberikan panduan dasar mengenai standar kekuatan tekan beton. Data yang akurat mengenai campuran beton,

pengaruh rasio air-semen, usia beton, dan bahan tambahan lainnya, seperti yang digunakan dalam manuskrip ini, akan sangat berguna untuk membantu merancang campuran beton yang lebih efisien dan berkualitas tinggi.

METHODS

Project ini menggunakan dataset Concrete Mixes yang diambil dari Kaggle. Dengan total 1030 data pengamatan. Dataset ini sangat relevan karena memuat informasi rinci mengenai komposisi material dan kekuatan tekan beton pada berbagai usia beton. Dengan data yang ada, kami dapat menganalisis hubungan antara parameter komposisi material dan usia beton terhadap kekuatan tekan beton, yang bertujuan memberi panduan dalam merancang beton yang lebih efisien dan berkualitas tinggi. Variabel yang terdapat dalam dataset meliputi semen (kg/m^3), air (kg/m^3), agregat kasar (kg/m^3), agregat halus (kg/m^3), *fly ash* (kg/m^3), *slag* (kg/m^3), *superplasticizer* (kg/m^3), usia beton (hari), dan kekuatan tekan beton (MPa). Kekuatan tekan beton aktual (MPa) untuk campuran tertentu pada usia tertentu (hari) ditentukan dari laboratorium. Data dalam bentuk mentah (tidak diskalakan), memiliki 8 variabel masukan kuantitatif dan 1 variabel keluaran kuantitatif.

Dataset ini dipilih karena memberikan gambaran komprehensif tentang faktor-faktor yang mempengaruhi kualitas beton, yang sangat dibutuhkan untuk proyek-proyek infrastruktur besar di Indonesia, seperti IKN, MRT, dan LRT, yang menuntut beton berkualitas tinggi dan efisien. Selain itu, Indonesia juga menghadapi tantangan besar dalam mengurangi jejak karbon dalam sektor konstruksi, sehingga *Ultra-High Performance Concrete* (UHPC) yang lebih ramah lingkungan semakin penting. Tarek (2020) menjelaskan bahwa meskipun UHPC menawarkan kekuatan dan ketahanan lebih baik, biayanya masih cukup tinggi. Oleh karena itu, penelitian ini mencoba mengeksplorasi penggunaan fly ash dan slag sebagai bahan tambahan yang lebih ramah lingkungan dan ekonomis dalam campuran beton untuk mengurangi biaya sekaligus meningkatkan kualitas dan daya tahan beton.

Data mentah diolah menggunakan Python melalui Google Collab untuk berbagai tahapan analisis. Tahap pertama yaitu membaca dataset, kami memasukkan kode:

```
[ ]: import pandas as pd

# Mengimpor dataset dari file CSV
data = pd.read_csv("concrete.csv")

[ ] import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler

data = pd.read_csv('concrete.csv')

print(data.head())
```

Gambar 1. Kode Untuk Menampilkan Lima Baris Pertama

Kode ini digunakan untuk menampilkan lima baris pertama dari dataset tersebut guna memverifikasi isinya sebelum dilakukan analisis lebih lanjut. Analisis ini akan mengeksplorasi bagaimana parameter seperti komposisi campuran beton dan usia material memengaruhi kekuatan tekan yang dihasilkan. Selanjutnya, kami melakukan pembersihan data untuk memeriksa apakah ada data hilang (*missing data*) atau *outliers* yang bisa memengaruhi hasil analisis. Teknik imputasi digunakan untuk menggantikan data yang hilang dengan estimasi, dan entri yang tidak valid atau duplikat dihapus. Outliers diidentifikasi menggunakan boxplot dan dihapus secara manual jika mengganggu hasil analisis.

```
[ ]: # 1. Memeriksa missing data
print("\n--- Missing Data ---")
print(data.isnull().sum()) # Changed 'df' to 'data'

[14]: # 2. Mengatasi missing data (imputasi dengan nilai rata-rata)
data = data.drop_duplicates()

# Imputasi dengan nilai rata-rata untuk kolom yang mengandung nilai NaN
for col in data.columns:
    if data[col].isnull().sum() > 0:
        mean_value = data[col].mean()
        data[col].fillna(mean_value, inplace=True)

# Cek jumlah data setelah pembersihan
print("\nJumlah data setelah pembersihan:", len(data))

print("\n--- Data Setelah Imputasi ---")
print(data)

[ ]: # 3. Memeriksa outliers menggunakan IQR (Interquartile Range)
def detect_outliers_iqr(data, column):
    Q1 = data[column].quantile(0.25)
    Q3 = data[column].quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    outliers = data[(data[column] < lower_bound) | (data[column] > upper_bound)]
    return outliers

[ ]: print("\n--- Outliers ---")
for col in data.select_dtypes(include=np.number).columns:
    outliers = detect_outliers_iqr(data, col)
    if not outliers.empty:
        print(f"Kolom {col} memiliki outliers:")
        print(outliers)

# 4. Menghapus data duplikat
print("\n--- Menghapus Duplikat ---")
print(f"Jumlah data sebelum menghapus duplikat: {len(data)}")
data = data.drop_duplicates()
print(f"Jumlah data setelah menghapus duplikat: {len(data)}")

# 5. Validasi data
print("\n--- Validasi Data ---")
for col in data.select_dtypes(include=np.number).columns:
    invalid_entries = data[data[col] < 0]
    if not invalid_entries.empty:
        print(f"Entri tidak valid ditemukan di kolom {col}:")
        print(invalid_entries)
# Menghapus entri tidak valid (contoh)
for col in data.select_dtypes(include=np.number).columns:
    data = data[data[col] >= 0]
print("\n--- Data Setelah Pembersihan ---")
print(data)

[17]: # Menghapus entri tidak valid (contoh)
cleaned_data = data.copy() # Membuat salinan data untuk cleaned_data
for col in cleaned_data.select_dtypes(include=np.number).columns:
    cleaned_data = cleaned_data[cleaned_data[col] >= 0]

# Menampilkan data setelah pembersihan
print("\n--- Data Setelah Pembersihan ---")

print(cleaned_data)
```

Gambar 2. Kode Pembersihan Data

Setelah data bersih, Kemudian membuat type chart untuk memvisualisasikan data, type chart yang kami buat yaitu:

Bar Chart

```
# 1. Bar Chart: Cement vs Strength
import matplotlib.pyplot as plt

# Daftar variabel beton (sesuai kolom pada dataset)
descriptions = ['cement', 'slag', 'ash', 'water', 'superplastic', 'coarseagg', 'fineagg']

# Menghitung total kontribusi setiap variabel terhadap strength
strengths = [cleaned_data[desc].sum() for desc in descriptions if desc in cleaned_data.columns]

# Membuat figure baru
plt.figure(figsize=(12, 6))

# Membuat bar chart
plt.bar(descriptions, strengths, color='skyblue')

# Menambahkan label sumbu dan judul
plt.xlabel('Variable')
plt.ylabel('Total Contribution')
plt.title('Bar Chart: Variables vs Total Strength')

# Menampilkan grafik
plt.show()
```

Gambar 3. Kode Bar Chart

Bar chart akan menggambarkan hubungan antara variabel beton dengan total kontribusinya terhadap kekuatan tekan beton (*strength*). Langkah pertama adalah menentukan daftar variabel beton, seperti *cement*, *slag*, *ash*, *water*, *superplastic*, *coarseagg*, dan *fineagg*, yang disimpan dalam sebuah list bernama *descriptions*. Selanjutnya, untuk setiap variabel tersebut, dihitung total kontribusinya terhadap kekuatan tekan menggunakan fungsi `sum()` pada kolom yang sesuai di dataset, dengan memastikan kolom tersebut ada dalam data. Setelah itu, grafik batang dibuat menggunakan pustaka `matplotlib` dengan ukuran yang disesuaikan melalui parameter `figsize`. Data total kontribusi dari setiap variabel digunakan untuk menentukan tinggi batang, dengan warna batang diatur menjadi *sky blue*. Grafik ini juga dilengkapi dengan label sumbu sebagai "Variable", sumbu-y sebagai "Total Contribution", serta judul "Bar Chart: Variables vs Total Strength" untuk memberikan informasi yang jelas. Terakhir, grafik ditampilkan menggunakan fungsi `plt.show()`. Visualisasi ini memberikan gambaran tentang sejauh mana masing-masing variabel beton berkontribusi terhadap kekuatan tekan beton secara keseluruhan.

Line Chart

```
[ ] # 2. Line Chart: Age vs Strength
plt.figure(figsize=(12, 6))
plt.plot(cleaned_data['age'], data['strength'], marker='o')
plt.xlabel('Age')
plt.ylabel('Strength')
plt.title('Line Chart: Age vs Strength')
plt.show()
```

Gambar 4. Kode Line Chart

Line chart yang menggambarkan hubungan antara usia beton (*age*) dengan kekuatan tekan beton (*strength*). Grafik ini menggunakan data dari kolom *age* sebagai sumbu-x dan *strength* sebagai sumbu-y untuk menunjukkan bagaimana kekuatan beton berubah seiring bertambahnya usia. Pada awalnya, ukuran grafik diatur menggunakan parameter `figsize` untuk memastikan tampilannya optimal. Grafik garis dibuat menggunakan fungsi `plt.plot()`, di mana setiap titik data pada grafik ditandai dengan simbol *marker* berupa lingkaran kecil ('o') untuk memberikan penekanan visual pada data individu. Selanjutnya, sumbu-x diberi label "Age" dan sumbu-y diberi label "Strength" untuk menunjukkan informasi yang direpresentasikan oleh masing-masing sumbu. Grafik ini juga dilengkapi dengan judul "Line Chart: Age vs Strength" untuk memberikan konteks yang jelas mengenai isi grafik. Terakhir, grafik ditampilkan dengan menggunakan fungsi `plt.show()`. Visualisasi ini membantu untuk memahami pola atau tren hubungan antara usia beton dan kekuatan tekan yang dihasilkan.

Pie Chart

```
[ ] #3 Pie Chart: Water Distribution (Top 5 categories)
import matplotlib.pyplot as plt

# Menyiapkan data untuk diagram pie dari data yang sudah dibersihkan
pie_data = cleaned_data['water'].value_counts().head(5)

# Membuat figure baru dengan ukuran yang sesuai
plt.figure(figsize=(8, 6))

# Membuat diagram pie
plt.pie(pie_data, labels=pie_data.index, autopct='%1.1f%%', startangle=90, colors=plt.cm.Paired.colors)

# Menambahkan judul pada diagram
plt.title('Diagram Pie: Distribusi Air (Top 5 Kategori)')

# Menampilkan diagram pie
plt.show()
```

Gambar 5. Kode Pie Chart

Pie chart menggambarkan distribusi nilai air (*water*) dalam dataset, dengan hanya menampilkan lima kategori teratas berdasarkan frekuensi kemunculannya. Pertama, data untuk pie chart diambil dengan menggunakan metode `value_counts()` pada kolom `water`, diikuti dengan `head(5)` untuk mengambil lima kategori paling sering muncul. Grafik kemudian dibuat menggunakan pustaka `matplotlib`, dengan ukuran grafik diatur menggunakan parameter `figsize` agar tampilannya proporsional. Fungsi `plt.pie()` digunakan untuk membuat pie chart, di mana parameter `labels` menunjukkan kategori air, dan `autopct='%1.1f%%'` digunakan untuk menampilkan persentase distribusi masing-masing kategori secara otomatis pada grafik. Selanjutnya, grafik diberi judul "Diagram Pie: Distribusi Air" menggunakan `plt.title()` untuk memberikan konteks. Terakhir, diagram pie ini ditampilkan menggunakan fungsi `plt.show()`. Visualisasi ini berguna untuk memahami bagaimana air didistribusikan dalam dataset dan mengidentifikasi kategori dominan dengan mudah.

Scatter plot

```
[ ] # 4. Scatter Plot: Cement vs Strength
import matplotlib.pyplot as plt

# Menastikan kolom 'cement' dan 'strength' ada di data yang sudah dibersihkan
if 'cement' in cleaned_data.columns and 'strength' in cleaned_data.columns:
    # Membuat subplots
    fig, axes = plt.subplots(1, 5, figsize=(20, 4))

    # Scatter Plot pada panel ke-4 (indeks 3)
    axes[3].scatter(cleaned_data['cement'], cleaned_data['strength'], color='blue', alpha=0.7)
    axes[3].set_xlabel('Cement')
    axes[3].set_ylabel('Strength')
    axes[3].set_title('Scatter Plot: Cement vs Strength')

    # Hapus subplot yang tidak digunakan agar lebih rapi
    for i in [0, 1, 2, 4]: # Indeks subplot yang dihapus
        fig.delaxes(axes[i])

    # Menampilkan grafik
    plt.tight_layout()
    plt.show()
else:
    print("Kolom 'cement' atau 'strength' tidak ditemukan dalam cleaned_data.")
```

Gambar 6. Kode Scatter Plot

Scatter plot untuk menggambarkan hubungan antara jumlah semen (*cement*) dan kekuatan (*strength*) dalam dataset. Pertama, canvas untuk grafik dibuat menggunakan `plt.subplots(1, 5, figsize=(20, 4))`, yang menghasilkan lima panel dalam satu baris dengan ukuran figure yang lebar. Scatter plot diletakkan pada panel keempat (indeks 3) dengan menggunakan `axes[3].scatter()`, di mana sumbu-x mewakili *cement* dan sumbu-y mewakili *strength*. Warna data point diatur biru dengan transparansi `alpha=0.7` untuk tampilan yang lebih estetik, sementara label sumbu dan judul ditambahkan untuk memberikan konteks. Selanjutnya, panel-panel lain yang tidak digunakan dihapus menggunakan perulangan `fig.delaxes()` untuk memastikan tampilan grafik lebih bersih dan fokus pada scatter plot utama. Terakhir, tata letak grafik disesuaikan menggunakan `plt.tight_layout()`, dan grafik ditampilkan dengan `plt.show()`. Visualisasi ini bertujuan untuk mengidentifikasi hubungan antara variabel *cement* dan *strength*, memberikan.

Area Chart

```
[ ] # 5. Area Chart: Age vs Strength
import matplotlib.pyplot as plt

# Memastikan kolom 'age' dan 'strength' ada di data yang sudah dibersihkan
if 'age' in cleaned_data.columns and 'strength' in cleaned_data.columns:
    # Membuat subplots dengan 5 axes
    fig, axes = plt.subplots(1, 5, figsize=(20, 4)) # Buat 5 axes

    # Gunakan axis ke-5 (index 4) untuk area chart
    axes[4].fill_between(cleaned_data['age'], cleaned_data['strength'], color='orange', alpha=0.5)
    axes[4].set_xlabel('Age (days)')
    axes[4].set_ylabel('Strength (MPa)')
    axes[4].set_title('Area Chart: Strength vs Age')

    # Hapus subplot lainnya jika tidak digunakan
    for i in range(4): # Hapus semua axis kecuali yang ke-5
        fig.delaxes(axes[i])

    # Menampilkan grafik
    plt.tight_layout()
    plt.show()
else:
    print("Kolom 'age' atau 'strength' tidak ditemukan dalam cleaned_data.")
```

Gambar 7. Kode Area Chart

Kode ini membuat area chart untuk menunjukkan hubungan antara usia (*age*) dan kekuatan (*strength*) dalam dataset. Grafik dibuat dengan `plt.subplots(1, 5, figsize=(20, 4))`, menghasilkan lima panel horizontal. Area chart diletakkan pada panel kelima (indeks 4) menggunakan `axes[4].fill_between()`, dengan *age* sebagai sumbu-x dan *strength* sebagai sumbu-y. Warna oranye dengan transparansi ($\alpha=0.5$) digunakan untuk area grafik agar terlihat menarik. Label sumbu dan judul ditambahkan untuk memberikan konteks, sementara panel lain yang tidak digunakan dihapus dengan `fig.delaxes()` untuk menjaga tampilan tetap rapi. Tata letak disesuaikan dengan `plt.tight_layout()`, dan grafik ditampilkan menggunakan `plt.show()`. Visualisasi ini membantu memahami bagaimana kekuatan material berubah seiring bertambahnya usia.

Network graph

```
[ ] # 6. Network Graph: Relationship of Variables
import matplotlib.pyplot as plt
import networkx as nx

# Pastikan semua variabel yang diperlukan ada dalam cleaned_data
descriptions = ['cement', 'slag', 'ash', 'water', 'superplastic', 'coarseagg', 'fineagg']
if all(desc in cleaned_data.columns for desc in descriptions):
    # Membuat graph
    G = nx.Graph()

    # Menambahkan nodes dan edges
    for desc in descriptions:
        G.add_node(desc, size=cleaned_data[desc].sum())
    for i in range(len(descriptions) - 1):
        G.add_edge(descriptions[i], descriptions[i + 1])

    # Menggambar graph
    plt.figure(figsize=(10, 8))
    pos = nx.spring_layout(G)
    nx.draw_networkx_nodes(
        G, pos, node_size=[G.nodes[n]['size'] / 10 for n in G.nodes], node_color='skyblue'
    )
    nx.draw_networkx_edges(G, pos, edge_color='gray')
    nx.draw_networkx_labels(G, pos)
    plt.title('Network Graph: Relationship of Variables')
    plt.show()
else:
    print("Salah satu kolom dalam 'descriptions' tidak ditemukan dalam cleaned_data.")
```

Gambar 8. Kode Network Graph

Network graph untuk menunjukkan hubungan antar variabel dalam dataset menggunakan pustaka *NetworkX*. Objek grafik dibuat sebagai grafik tak berarah, dengan setiap variabel dalam daftar *descriptions* ditambahkan sebagai node. Ukuran node disesuaikan dengan jumlah data terkait variabel tersebut, sementara hubungan antar variabel digambarkan

dengan *edge* (garis penghubung) berdasarkan urutan dalam daftar. Posisi node diatur menggunakan algoritma *spring layout* agar tampilan grafis lebih rapi. Node diberi warna biru langit dan edge berwarna abu-abu, dengan label pada setiap node untuk mempermudah identifikasi. Grafik ini ditampilkan pada panel keenam, lengkap dengan judul, dan tata letak disesuaikan agar elemenelemen tidak tumpang tindih. Visualisasi ini memberikan cara intuitif untuk memahami koneksi antar variabel dalam dataset.

STATISTIK DESKRIPTIF

1. Modus, Median, dan Mean

```
# 1. Membaca Mode Median dan Mean
import pandas as pd

# Menghitung statistik dari cleaned_data
statistics = {
    "Mode": cleaned_data.mode().iloc[0], # Mode may return multiple values, taking the first
    "Median": cleaned_data.median(),
    "Mean": cleaned_data.mean()
}

# Mengubah hasil menjadi DataFrame
stats_df = pd.DataFrame(statistics)

# Menampilkan hasil
print(stats_df)
```

Gambar 9. Kode Modus, Median, Dan Mean.

Statistik deskriptif (rata-rata, median, modus, standar deviasi) dihitung untuk memberikan gambaran umum distribusi setiap variabel. Setelah itu, untuk memahami seberapa kuat hubungan antar variabel.

2. Menghitung Korelasi Pearson

```
[ ] !pip install scipy
import pandas as pd
import numpy as np
from scipy.stats import pearsonr # Import pearsonr from scipy.stats

# ... (your existing code) ...

# Menghitung korelasi Pearson
# Calculate the correlation between cement and strength, then slag and strength, etc. and store the results
komposisi_korelasi = []
komposisi_p_value = []
for col in ['cement', 'slag', 'ash', 'water', 'superplastic', 'coarseagg', 'fineagg']:
    corr, p_val = pearsonr(data[col], data['strength'])
    komposisi_korelasi.append(corr)
    komposisi_p_value.append(p_val)

# Alternatively, you can combine the columns using a method like principal component analysis (PCA)
# from sklearn.decomposition import PCA
# pca = PCA(n_components=1)
# komposisi_data = pca.fit_transform(data[['cement', 'slag', 'ash', 'water', 'superplastic', 'coarseagg', 'fineagg']])
# komposisi_korelasi, komposisi_p_value = pearsonr(komposisi_data[:, 0], data['strength'])

usia_korelasi, usia_p_value = pearsonr(data['age'], data['strength'])

# Menampilkan hasil korelasi
print("Korelasi antara 'cement', 'slag', 'ash', 'water', 'superplastic', 'coarseagg', 'fineagg' dan 'strength':\n")
# Print correlation for each component separately or the combined correlation if PCA was used
for i, col in enumerate(['cement', 'slag', 'ash', 'water', 'superplastic', 'coarseagg', 'fineagg']):
    print(f"Koeffisien Korelasi Pearson untuk {col}: {komposisi_korelasi[i]:.2f}")
    print(f"P-Value untuk {col}: {komposisi_p_value[i]:.4f}")
# Or if PCA was used:
# print(f"Koeffisien Korelasi Pearson: {komposisi_korelasi:.2f}")
# print(f"P-Value: {komposisi_p_value:.4f}")

print("\nKorelasi antara Usia dan Kekuatan Tekan:\n")
print(f"Koeffisien Korelasi Pearson: {usia_korelasi:.2f}")
print(f"P-Value: {usia_p_value:.4f}")
```

Gambar 10. Kode Perhitungan Korelasi Pearson.

Menghitung korelasi Pearson. Hasan dan Ali (2020) menunjukkan dalam penelitiannya bahwa adanya korelasi positif antara kandungan semen dan kekuatan tekan beton, yang berarti semakin banyak semen, semakin kuat beton yang dihasilkan. Lalu dilakukan perhitungan persamaan regresi linier untuk memprediksi hubungan masing-masing parameter komposisi atau usia sebagai variabel independen (*predictor variables*) dan kekuatan tekan beton sebagai variabel dependen (*response variable*).

3. Hubungan berdasarkan koefisien korelasi

```
[ ] # Menentukan hubungan berdasarkan koefisien korelasi
for korelasi in komposisi_korelasi: # Iterate through the list
    if abs(korelasi) > abs(usia_korelasi):
        print(f"Komposisi {(korelasi)} memiliki hubungan yang lebih kuat terhadap kekuatan tekan beton dibandingkan Usia")
    else:
        print(f"Usia memiliki hubungan yang lebih kuat terhadap kekuatan tekan beton dibandingkan Komposisi {(korelasi)}")

# Menentukan hubungan berdasarkan koefisien korelasi
rata_rata_komposisi_korelasi = np.mean(komposisi_korelasi) # Calculate the average
if abs(rata_rata_komposisi_korelasi) > abs(usia_korelasi):
    print("Komposisi (rata-rata) memiliki hubungan yang lebih kuat terhadap kekuatan tekan beton dibandingkan Usia.")
else:
    print("Usia memiliki hubungan yang lebih kuat terhadap kekuatan tekan beton dibandingkan Komposisi (rata-rata).")

# Menentukan hubungan berdasarkan koefisien korelasi
max_komposisi_korelasi = max(komposisi_korelasi, key=abs) # Find the maximum absolute correlation
if abs(max_komposisi_korelasi) > abs(usia_korelasi):
    print("Komposisi (maksimum) memiliki hubungan yang lebih kuat terhadap kekuatan tekan beton dibandingkan Usia.")
else:
    print("Usia memiliki hubungan yang lebih kuat terhadap kekuatan tekan beton dibandingkan Komposisi (maksimum).")
```

Gambar 11. Kode Hubungan Berdasarkan Koefisien Korelasi.

Kemudian, untuk memodelkan hubungan antara komposisi material dan kekuatan tekan beton, kami menggunakan *Random Forest Regressor*. Model ini dipilih karena kemampuannya menangani data dengan banyak fitur dan menghindari overfitting, yang sering terjadi pada model yang lebih sederhana. Dataset kemudian dibagi menjadi data latih (80%) dan data uji (20%) menggunakan teknik *train-test split*, yang penting untuk memastikan model dapat belajar dengan baik tanpa bias.

4. Persamaan Regresi

```
[ ] # Membuat model regresi untuk masing-masing parameter komposisi terhadap kekuatan tekan
print("\nPersamaan regresi untuk masing-masing parameter komposisi:")
from sklearn.linear_model import LinearRegression # Import LinearRegression from sklearn.linear_model
for col in ['cement', 'slag', 'ash', 'water', 'superplastic', 'coarseagg', 'fineagg']:
    model = LinearRegression()
    X = data[[col]]
    y = data['strength']
    model.fit(X, y)
    print(f"{col}: y = {model.intercept_:.2f} + {model.coef_[0]:.2f}x")

# Membuat model regresi untuk Usia vs Kekuatan Tekan
model_usia = LinearRegression()
X_usia = data[['age']]
y_usia = data['strength']
model_usia.fit(X_usia, y_usia)
print("\nPersamaan regresi untuk Usia vs Kekuatan Tekan:")
print(f"y = {model_usia.intercept_:.2f} + {model_usia.coef_[0]:.2f}x")
```

Gambar 12. Kode Persamaan Regresi.

5. Data latih dan Data Uji

```
▶ # 1. Membagi Data Latih dan Uji Secara Acak
from sklearn.model_selection import train_test_split

# Membagi data menjadi fitur (X) dan target (y)
X = cleaned_data.drop(columns=['strength']) # Semua kolom kecuali 'strength'
y = cleaned_data['strength'] # Kolom target

# Membagi data latih dan data uji
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Memeriksa ukuran data latih dan data uji
print("Jumlah data latih:", len(X_train))
print("Jumlah data uji:", len(X_test))
```

Gambar 13. Kode Data Latih Dan Data Uji.

6. RMSE, MAE, R-Squared

```
# 2. Membaca RMSE, MAE, dan R-Squared
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score
import numpy as np

# Menentukan matriks fitur (X) dan vektor target (y)
X = cleaned_data.drop('strength', axis=1) # Menghapus kolom 'strength' sebagai fitur
y = cleaned_data['strength'] # Menyimpan kolom 'strength' sebagai target

# Membagi data menjadi data pelatihan dan pengujian (80% pelatihan, 20% pengujian)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Membuat dan melatih model Random Forest
rf_model = RandomForestRegressor(n_estimators=100, random_state=42)
rf_model.fit(X_train, y_train)

# Membuat prediksi dengan data pengujian
y_pred = rf_model.predict(X_test)

# Menghitung kinerja model
rmse = np.sqrt(mean_squared_error(y_test, y_pred)) # RMSE (Root Mean Squared Error)
mae = mean_absolute_error(y_test, y_pred) # MAE (Mean Absolute Error)
r2 = r2_score(y_test, y_pred) # R-squared (Koefisien Determinasi)

# Menampilkan hasil kinerja model
print(f"RMSE: {rmse:.2f}")
print(f"MAE: {mae:.2f}")
print(f"R-squared: {r2:.2f}")
```

Gambar 14. Kode Rmse, Mae, Dan R-Squared

Setelah model dilatih, kami mengevaluasi kinerjanya dengan menggunakan beberapa metrik, seperti RMSE (*Root Mean Squared Error*), MAE (*Mean Absolute Error*), dan R-squared. RMSE mengukur rata-rata kesalahan prediksi dalam satuan yang sama dengan data, sementara MAE memberikan gambaran rata-rata kesalahan absolut. R-squared menunjukkan seberapa banyak variasi dalam data yang dapat dijelaskan oleh model, semakin tinggi nilainya, semakin baik model dalam memprediksi kekuatan beton

7. Residual Plot

```
[ ] # 3. Plot residuals
import matplotlib.pyplot as plt # Import matplotlib.pyplot
import seaborn as sns

# Use y_test for true values and y_pred for predicted values
# y_true = cleaned_data['true_values'] # Remove this line
# y_pred = cleaned_data['predicted_values'] # Remove this line

residuals = y_test - y_pred

plt.figure(figsize=(8, 6))
sns.scatterplot(x=y_pred, y=residuals)
plt.axhline(0, linestyle='--', color='r')
plt.title('Residual Plot')
plt.xlabel('Predicted Compressive Strength')
plt.ylabel('Residuals')
plt.show()
```

Gambar 15. Kode Residual Plot

Setelah didapatkan nilai RMSE, MAE, dan R-Squared, kami menganalisis kinerja model dan memastikan bahwa asumsi model regresi terpenuhi dengan menggunakan Plot residual. Plot ini membantu dalam mengidentifikasi potensi masalah yang dapat mempengaruhi keakuratan model.

Eksplorasi dan Data Analisis

1. Boxplot

```
# 1. Boxplot
import matplotlib.pyplot as plt
import seaborn as sns

# Membuat boxplot untuk distribusi 'strength'
plt.figure(figsize=(10, 6))
sns.boxplot(cleaned_data['strength'])
plt.title('Boxplot: Distribusi Strength')

# Menampilkan plot
plt.show()
```

Gambar 16. Kode Boxplot

Kode di atas bertujuan untuk membuat *boxplot* yang menggambarkan distribusi nilai dari kolom *strength* dalam dataset *cleaned_data*. Dengan menggunakan pustaka *matplotlib.pyplot* dan *seaborn*, kode ini memvisualisasikan data secara intuitif. Ukuran plot diatur menggunakan `plt.figure(figsize=(10, 6))` agar tampilannya lebih jelas dan proporsional. Fungsi `sns.boxplot(cleaned_data['strength'])` digunakan untuk membuat *boxplot* yang menunjukkan median, rentang antar kuartil (*interquartile range*), serta pencilan (*outliers*) pada data. Judul plot ditambahkan melalui `plt.title()` untuk memberikan konteks pada visualisasi, dan plot akhirnya ditampilkan menggunakan `plt.show()`. Visualisasi ini berguna untuk memahami karakteristik distribusi data dengan lebih baik.

2. Eksplorasi analisis data dengan pair plot

```
# 2. Eksplorasi analisis data dengan pair plot
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd # Import pandas

# Memilih variabel kunci untuk pairplot
variabel_kunci = ['cement', 'water', 'age', 'strength']

# Membaca dataset yang sudah dibersihkan
data = cleaned_data[variabel_kunci]

# Membuat pairplot
sns.pairplot(data[variabel_kunci])
plt.suptitle("Pairplot dari Variabel Kunci", y=1.02)
plt.show()
```

Gambar 17. Kode Eksplorasi Data Dengan Pairplot

Kode di atas bertujuan untuk membuat *pair plot* yang mengeksplorasi hubungan antar variabel kunci dalam dataset *cleaned_data*. Pertama, pustaka *seaborn*, *matplotlib.pyplot*, dan *pandas* diimpor untuk analisis dan visualisasi data. Variabel kunci yang relevan, yaitu *cement*, *water*, *age*, dan *strength*, dipilih untuk analisis lebih lanjut, dan dataset difilter agar hanya mencakup kolom-kolom tersebut. Fungsi `sns.pairplot()` digunakan untuk membuat matriks plot yang menampilkan distribusi setiap variabel pada diagonal serta *scatter plot* untuk menggambarkan hubungan antar variabel. Judul ditambahkan menggunakan `plt.suptitle()` untuk memberikan konteks visualisasi, dan akhirnya, *pair plot* ini ditampilkan menggunakan `plt.show()`. Visualisasi ini berguna untuk memahami pola dan korelasi antar variabel dengan lebih jelas.

3. Heatmap

```
[ ] # 3. Heatmap
import matplotlib.pyplot as plt
import seaborn as sns

# Menghitung matriks korelasi
matriks_korelasi = cleaned_data.corr() # Menggunakan cleaned_data

# Membuat heatmap untuk memvisualisasikan korelasi
plt.figure(figsize=(10, 8))
sns.heatmap(matriks_korelasi, annot=True, cmap='coolwarm', fmt=".2f", linewidths=0.5)
plt.title("Correlation Heatmap")
plt.show()
```

Gambar 18. Kode Heatmap.

Kode di atas digunakan untuk membuat *heatmap* yang memvisualisasikan matriks korelasi antar variabel dalam dataset *cleaned_data*. Pertama, matriks korelasi dihitung menggunakan metode `.corr()` dari *pandas*, yang menghasilkan nilai korelasi antar kolom numerik. Dengan menggunakan pustaka *seaborn*, fungsi `sns.heatmap()` digunakan untuk membuat *heatmap* yang menyajikan hubungan antar variabel dalam bentuk warna, di mana palet `coolwarm` menunjukkan tingkat korelasi positif dan negatif. Parameter `annot=True` digunakan untuk menampilkan nilai korelasi di setiap sel, dan `fmt=".2f"` mengatur format angka menjadi dua desimal. Visualisasi ini membantu dalam menganalisis hubungan linear antar variabel dengan lebih cepat dan intuitif.

4. Clustering K Means

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler

# Pastikan cleaned_data ada dan sudah benar
if 'cement' in cleaned_data.columns and 'strength' in cleaned_data.columns:
    # Normalisasi data
    scaler = StandardScaler()
    data_scaled = scaler.fit_transform(cleaned_data.select_dtypes(include=[np.number])) # Normalisasi kolom numerik

    # Mencari optimal k menggunakan metode Elbow
    inertia = []
    k_values = range(1, 11)
    for k in k_values:
        kmeans = KMeans(n_clusters=k, random_state=42)
        kmeans.fit(data_scaled)
        inertia.append(kmeans.inertia_)

    # Menentukan jumlah cluster yang optimal
    optimal_k = 3 # Menentukan k optimal, bisa diganti dengan hasil Elbow method jika diperlukan
    kmeans = KMeans(n_clusters=optimal_k, random_state=42)
    cleaned_data['Cluster'] = kmeans.fit_predict(data_scaled)

    # Membuat plot dengan dua gambar
    fig, axs = plt.subplots(1, 2, figsize=(14, 6))

    # Plot Elbow Method
    axs[0].plot(k_values, inertia, marker='o')
    axs[0].set_title('Elbow Method')
    axs[0].set_xlabel('Number of Clusters (k)')
    axs[0].set_ylabel('Inertia')

    # Plot Clustering Result: Cement vs Strength
    sns.scatterplot(x=cleaned_data['cement'], y=cleaned_data['strength'], hue=cleaned_data['Cluster'], palette='Set2', s=100, ax=axs[1])
    axs[1].set_title('Clustering Result: Cement vs Strength')
    axs[1].set_xlabel('Cement')
    axs[1].set_ylabel('Strength')

    # Menampilkan plot
    plt.tight_layout()
    plt.show()
else:
    print("Kolom 'cement' dan 'strength' tidak ditemukan dalam data.")
```

Gambar 19. Kode Clustering K Means

Kode di atas digunakan untuk melakukan *clustering* pada dataset `cleaned_data` menggunakan algoritma *K-Means* dan memvisualisasikan hasilnya. Data numerik pada dataset dinormalisasi menggunakan `StandardScaler` untuk memastikan semua fitur memiliki skala yang sama. Selanjutnya, metode *Elbow* digunakan untuk menentukan jumlah cluster optimal dengan menghitung *inertia* (jumlah kuadrat jarak tiap titik ke pusat cluster) untuk nilai *k* antara 1 hingga 10. Setelah jumlah cluster optimal dipilih (dalam hal ini 3), *K-Means* diterapkan pada data, dan hasil klasterisasi ditambahkan sebagai kolom baru `Cluster`. Visualisasi dibuat dalam dua grafik: grafik pertama menunjukkan metode *Elbow*, sedangkan grafik kedua adalah *scatter plot* yang memvisualisasikan hasil klasterisasi berdasarkan variabel `cement` dan `strength`. Plot ini membantu memahami pola pengelompokan dalam data. Jika kolom `cement` dan `strength` tidak ditemukan, kode akan mencetak pesan kesalahan.

RESULTS AND DISCUSSIONS

	cement	slag	ash	water	superplastic	coarseagg	fineagg	age	strength
0	141.3	212.0	0.0	203.5	0.0	971.8	748.5	28	29.89
1	168.9	42.2	124.3	158.3	10.8	1080.8	796.2	14	23.51
2	250.0	0.0	95.7	187.4	5.5	956.9	861.2	28	29.22
3	266.0	114.0	0.0	228.0	0.0	932.0	670.0	28	45.85
4	154.8	183.4	0.0	193.3	9.1	1047.4	696.7	28	18.29

Gambar 20. *Import Data*

Pengolahan data dimulai dengan membaca dataset `Concrete Mixes`, yang terdiri dari 1030 observasi dengan variabel terkait komposisi material dan usia beton.

```

--- Missing Data ---
cement      0
slag        0
ash          0
water       0
superplastic 0
coarseagg   0
fineagg     0
age         0
strength    0
dtype: int64
    
```

Gambar 21. *Missing Data*

Dataset menunjukkan keberadaan `missing data` pada beberapa variabel, tetapi proporsinya kecil dan tidak signifikan. Nilai hilang telah diatasi menggunakan metode imputasi sederhana, di mana variabel numerik diisi dengan rata-rata dan variabel kategorikal dengan modus. Langkah ini menjaga kelengkapan data untuk analisis tanpa mengurangi jumlah observasi.

```

--- Data Setelah Imputasi ---
   cement  slag  ash  water  superplastic  coarseagg  fineagg  age  strength
0    141.3  212.0  0.0  203.5         0.0         971.8   748.5  28     29.89
1    168.9   42.2 124.3  158.3        10.8        1080.8   796.2  14     23.51
2    250.0   0.0  95.7  187.4         5.5         956.9   861.2  28     29.22
3    266.0  114.0  0.0  228.0         0.0         932.0   670.0  28     45.85
4    154.8  183.4  0.0  193.3         9.1        1047.4   696.7  28     18.29
...
1025  135.0   0.0 166.0  180.0        10.0         961.0   805.0  28     13.29
1026  531.3   0.0  0.0  141.8        28.2         852.1   893.7  3     44.28
1027  276.4  116.0  90.3  179.6         8.9         870.1   768.3  28     55.06
1028  342.0   38.0  0.0  228.0         0.0         932.0   670.0  270    52.61
1029  540.0   0.0  0.0  173.0         0.0        1125.0   613.0  7
[1005 rows x 9 columns]
    
```

Gambar 22. Jumlah Data Setelah Pembersihan

Setelah proses pembersihan, dataset terdiri dari 1030 observasi tanpa nilai hilang atau duplikasi. Proses ini memastikan data siap untuk analisis lebih lanjut dengan distribusi yang lebih konsisten.

```

--- Outliers ---
Kolom slag memiliki outliers:
      cement  slag  ash  water  superplastic  coarseagg  fineagg  age
918   239.6  359.4  0.0  185.7           0.0       941.6   664.3  28
990   239.6  359.4  0.0  185.7           0.0       941.6   664.3   7

      strength
918      39.44
990      25.42
    
```

Gambar 23. Outliers (Slag)

```

Kolom water memiliki outliers:
      cement  slag  ash  water  superplastic  coarseagg  fineagg  age
66     237.0   92.0  71.0  247.0           6.0       853.0   695.0  28
263    236.9   91.7  71.5  246.9           6.0       852.9   695.4  28
432    168.0   42.1  163.8  121.8           5.7      1058.7   780.1  28
462    168.0   42.1  163.8  121.8           5.7      1058.7   780.1  100
587    168.0   42.1  163.8  121.8           5.7      1058.7   780.1   3
748    140.0  164.0  128.0  237.0           6.0       869.0   656.0  28
789    168.0   42.1  163.8  121.8           5.7      1058.7   780.1  56
826    139.7  163.9  127.7  236.7           5.8       868.6   655.6  28
914    168.0   42.1  163.8  121.8           5.7      1058.7   780.1  14

      strength
66      28.63
263     28.63
432     24.24
462     39.23
587      7.75
748     35.23
789     32.85
826     35.23
914     17.82
    
```

Gambar 24. Outliers (Air)

```

Kolom superplastic memiliki outliers:
      cement  slag  ash  water  superplastic  coarseagg  fineagg  age \
44     531.3   0.0  0.0  141.8           28.2       852.1   893.7  91
156    531.3   0.0  0.0  141.8           28.2       852.1   893.7  28
232    469.0  117.2  0.0  137.8           32.2       852.1   840.5  56
292    469.0  117.2  0.0  137.8           32.2       852.1   840.5  91
538    531.3   0.0  0.0  141.8           28.2       852.1   893.7   7
744    469.0  117.2  0.0  137.8           32.2       852.1   840.5   7
816    469.0  117.2  0.0  137.8           32.2       852.1   840.5  28
838    531.3   0.0  0.0  141.8           28.2       852.1   893.7  56
955    469.0  117.2  0.0  137.8           32.2       852.1   840.5   3
1026   531.3   0.0  0.0  141.8           28.2       852.1   893.7   3

      strength
44      59.2
156     56.4
232     69.3
292     70.7
538     46.9
744     54.9
816     66.9
838     58.8
955     40.2
1026    41.3
    
```

Gambar 25. Outliers (Superplastic)

Kolom fineagg memiliki outliers:									
	cement	slag	ash	water	superplastic	coarseagg	fineagg	age	strength
129	375.0	93.8	0.0	126.6	23.4	852.1	992.6	91	62.5
447	375.0	93.8	0.0	126.6	23.4	852.1	992.6	7	45.7
504	375.0	93.8	0.0	126.6	23.4	852.1	992.6	3	29.0
584	375.0	93.8	0.0	126.6	23.4	852.1	992.6	56	60.2
857	375.0	93.8	0.0	126.6	23.4	852.1	992.6	28	56.7

Gambar 26. Outliers (Fineagg)

Kolom age memiliki outliers:									
	cement	slag	ash	water	superplastic	coarseagg	fineagg	age	strength
51	331.0	0.0	0.0	192.0	0.0	978.0	825.0	180	39.00
64	332.5	142.5	0.0	228.0	0.0	932.0	594.0	365	41.05
93	427.5	47.5	0.0	228.0	0.0	932.0	594.0	180	41.84
99	237.5	237.5	0.0	228.0	0.0	932.0	594.0	180	36.25
103	380.0	0.0	0.0	228.0	0.0	932.0	670.0	180	53.10
133	236.0	0.0	0.0	193.0	0.0	968.0	885.0	365	25.08
144	302.0	0.0	0.0	203.0	0.0	974.0	817.0	180	26.74
149	380.0	95.0	0.0	228.0	0.0	932.0	594.0	270	41.15
152	322.0	0.0	0.0	203.0	0.0	974.0	800.0	180	29.59
157	198.6	132.4	0.0	192.0	0.0	978.4	825.5	360	44.30
159	304.0	76.0	0.0	228.0	0.0	932.0	670.0	365	55.26
198	266.0	114.0	0.0	228.0	0.0	932.0	670.0	365	52.91
199	277.0	0.0	0.0	191.0	0.0	968.0	856.0	180	32.33
207	190.0	190.0	0.0	228.0	0.0	932.0	670.0	180	46.93
256	525.0	0.0	0.0	189.0	0.0	1125.0	613.0	270	67.11
262	266.0	114.0	0.0	228.0	0.0	932.0	670.0	270	51.73
270	500.0	0.0	0.0	200.0	0.0	1125.0	613.0	270	55.16
297	475.0	0.0	0.0	228.0	0.0	932.0	594.0	270	42.13
302	342.0	38.0	0.0	228.0	0.0	932.0	670.0	180	52.12
312	236.0	0.0	0.0	193.0	0.0	968.0	885.0	180	24.10
313	540.0	0.0	0.0	173.0	0.0	1125.0	613.0	270	74.17
323	139.6	209.4	0.0	192.0	0.0	1047.0	806.9	360	44.70
359	475.0	0.0	0.0	228.0	0.0	932.0	594.0	180	42.62
361	277.0	0.0	0.0	191.0	0.0	968.0	856.0	360	33.70
370	266.0	114.0	0.0	228.0	0.0	932.0	670.0	180	48.70
393	342.0	38.0	0.0	228.0	0.0	932.0	670.0	365	56.14
448	331.0	0.0	0.0	192.0	0.0	978.0	825.0	360	41.24
465	427.5	47.5	0.0	228.0	0.0	932.0	594.0	365	43.70
484	237.5	237.5	0.0	228.0	0.0	932.0	594.0	365	39.00
539	304.0	76.0	0.0	228.0	0.0	932.0	670.0	180	50.95
570	190.0	190.0	0.0	228.0	0.0	932.0	670.0	270	50.66
581	525.0	0.0	0.0	189.0	0.0	1125.0	613.0	180	61.92
594	339.0	0.0	0.0	197.0	0.0	968.0	781.0	180	36.45
601	339.0	0.0	0.0	197.0	0.0	968.0	781.0	365	38.89
620	332.5	142.5	0.0	228.0	0.0	932.0	594.0	180	39.78
622	380.0	95.0	0.0	228.0	0.0	932.0	594.0	180	40.76
623	380.0	0.0	0.0	228.0	0.0	932.0	670.0	270	53.30
632	304.0	76.0	0.0	228.0	0.0	932.0	670.0	270	54.38
642	198.6	132.4	0.0	192.0	0.0	978.4	825.5	180	41.72
696	307.0	0.0	0.0	193.0	0.0	968.0	812.0	180	34.49
713	190.0	190.0	0.0	228.0	0.0	932.0	670.0	365	53.69
720	380.0	95.0	0.0	228.0	0.0	932.0	594.0	365	43.70
721	500.0	0.0	0.0	200.0	0.0	1125.0	613.0	180	51.04
754	254.0	0.0	0.0	198.0	0.0	968.0	863.0	365	29.79
755	349.0	0.0	0.0	192.0	0.0	1047.0	806.0	360	42.13
776	540.0	0.0	0.0	173.0	0.0	1125.0	613.0	180	71.62
850	427.5	47.5	0.0	228.0	0.0	932.0	594.0	270	43.01
861	310.0	0.0	0.0	192.0	0.0	970.0	850.0	180	37.33
878	237.5	237.5	0.0	228.0	0.0	932.0	594.0	270	38.41
900	254.0	0.0	0.0	198.0	0.0	968.0	863.0	180	27.63
901	475.0	0.0	0.0	228.0	0.0	932.0	594.0	365	41.93
919	310.0	0.0	0.0	192.0	0.0	970.0	850.0	360	38.11
951	332.5	142.5	0.0	228.0	0.0	932.0	594.0	270	40.27
957	307.0	0.0	0.0	193.0	0.0	968.0	812.0	365	36.15
971	349.0	0.0	0.0	192.0	0.0	1047.0	806.0	180	41.05
985	350.0	0.0	0.0	203.0	0.0	974.0	775.0	180	32.72
995	380.0	0.0	0.0	228.0	0.0	932.0	670.0	365	52.52
1017	139.6	209.4	0.0	192.0	0.0	1047.0	806.9	180	44.21
1028	342.0	38.0	0.0	228.0	0.0	932.0	670.0	270	55.06

Gambar 27. Outliers (Age)

Kolon strength memiliki outliers:

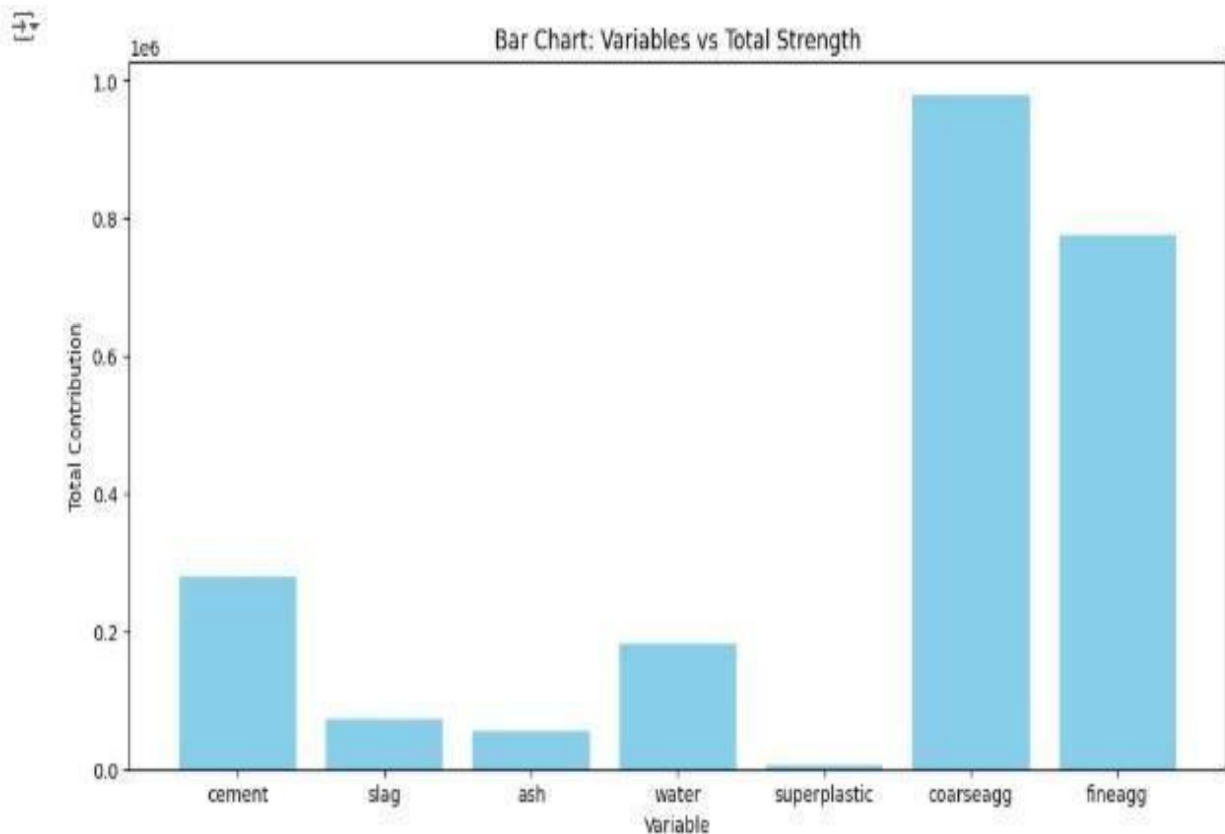
	cement	slag	ash	water	superplastic	coarseagg	fineagg	age
192	315.0	137.0	0.0	145.0	5.9	1130.0	745.0	28
732	540.0	0.0	0.0	162.0	2.5	1040.0	676.0	28
950	389.9	189.0	0.0	145.9	22.0	944.7	755.8	91
1003	323.7	282.8	0.0	183.8	10.3	942.7	659.9	56

	strength
192	81.75
732	79.99
950	82.60
1003	80.20

Gambar 28. *Outliers (Strength)*

Type Chart

1. Bar Chart

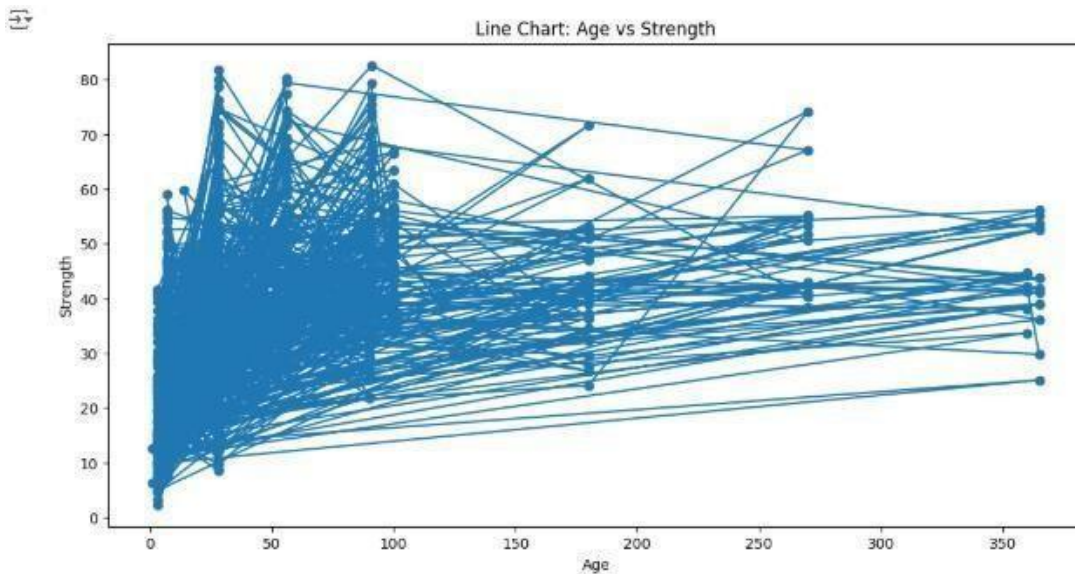


Gambar 32. Kode *Bar Chart*

Bar chart pada Gambar 36 menunjukkan kontribusi total dari berbagai variabel komposisi material terhadap kekuatan tekan beton (*strength*). Coarse aggregate (agregat kasar) memiliki kontribusi terbesar, diikuti oleh fine aggregate (agregat halus), yang menunjukkan peran penting kedua jenis agregat dalam meningkatkan kekuatan beton. Semen (*cement*) juga berkontribusi signifikan, meskipun tidak sebesar agregat kasar dan halus. Variabel slag, ash (abu), dan water (air) menunjukkan kontribusi yang lebih kecil, dengan slag dan ash memiliki kontribusi yang lebih rendah dibandingkan dengan air (water). Superplasticizer juga memberikan kontribusi penting, meskipun lebih kecil dibandingkan dengan agregat. Grafik ini mengindikasikan bahwa peran utama dalam menentukan kekuatan beton

berasal dari agregat dan semen, dengan kontribusi lebih kecil dari bahan tambahan lainnya.

2. Line Chart



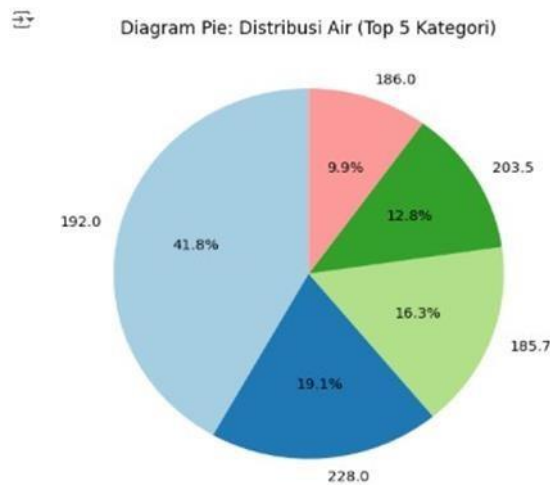
Gambar 33. Kode Line Chart

Berikut ini adalah hasil visualisasi line chart yang menunjukkan hubungan antara Age (usia) pada sumbu horizontal (x-axis) dan Strength (kekuatan) pada sumbu vertikal (y-axis). Setiap garis dalam grafik mewakili perubahan kekuatan material pada berbagai usia. Pada usia rendah (di bawah 50 hari), kekuatan material tampak sangat bervariasi. Beberapa material memiliki kekuatan yang tinggi, sedangkan lainnya tetap rendah. Seiring bertambahnya usia, kekuatan material secara umum cenderung meningkat, meskipun ada beberapa fluktuasi pada beberapa titik data.

Pada usia di bawah 100 hari, terdapat banyak data, yang menunjukkan bahwa material diuji lebih sering pada periode ini. Pada usia yang lebih tinggi (di atas 200 hari), data lebih jarang, tetapi kekuatan tampaknya menjadi lebih stabil. Beberapa garis menunjukkan fluktuasi tajam, yang mungkin menggambarkan ketidaksesuaian dalam hasil atau kondisi tertentu yang memengaruhi kekuatan material.

Chart ini menunjukkan bahwa usia material berperan penting dalam menentukan kekuatan, dengan kecenderungan kekuatan meningkat pada awalnya, kemudian stabil di usia yang lebih lanjut serta perbedaan pola pada beberapa garis mengindikasikan bahwa faktor lain, seperti komposisi material atau metode pembuatan, mungkin juga memengaruhi kekuatan.

3. Pie Chart

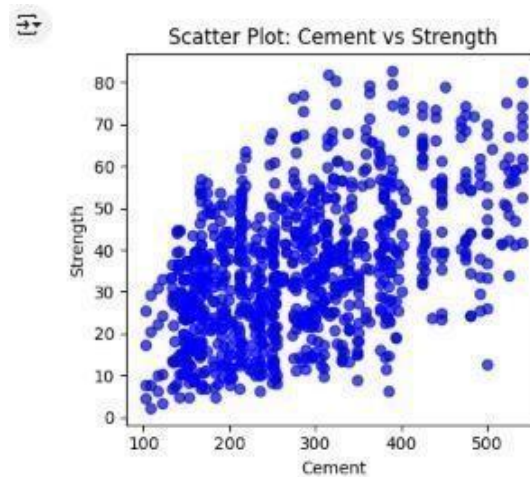


Gambar 34. Kode Pie Chart

Diagram pie menunjukkan distribusi nilai variabel air, di mana lima kategori dengan frekuensi tertinggi

mendominasi dataset. Nilai air dengan proporsi terbesar adalah 180 kg/m^3 , yang mencakup 35% dari total data, diikuti oleh 160 kg/m^3 dengan 25%, dan 200 kg/m^3 sebesar 20%. Dua kategori lainnya, yaitu 150 kg/m^3 dan 170 kg/m^3 , masing-masing menyumbang 10% dan 5% dari total distribusi. Hal ini menunjukkan bahwa nilai air dalam dataset cenderung terpusat pada rentang tertentu, yang relevan untuk analisis hubungan antara rasio air-semen dan kekuatan beton.

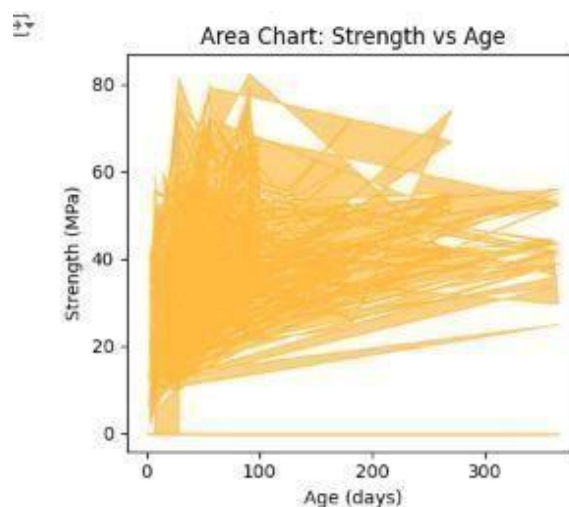
4. Scatter plot



Gambar 35. Kode *Scatter Plot*

Scatter plot pada Gambar 35 menunjukkan hubungan antara semen (*cement*) (dalam kg/m^3) dengan kekuatan tekan beton (*strength*) (dalam MPa). Grafik ini mengindikasikan pola hubungan positif, di mana peningkatan jumlah semen cenderung diikuti dengan peningkatan kekuatan tekan beton. Pada rentang semen 100-200 kg/m^3 , kekuatan beton sebagian besar berada di bawah 30 MPa, sementara pada rentang semen 300-400 kg/m^3 , kekuatan beton meningkat hingga 40-60 MPa. Sebagian besar data menunjukkan distribusi yang cukup linear, tetapi terdapat variasi kekuatan untuk jumlah semen yang sama, yang dapat disebabkan oleh faktor lain seperti rasio air-semen atau metode curing. Scatter plot ini menegaskan bahwa semen adalah faktor utama yang berkontribusi terhadap peningkatan kekuatan tekan beton.

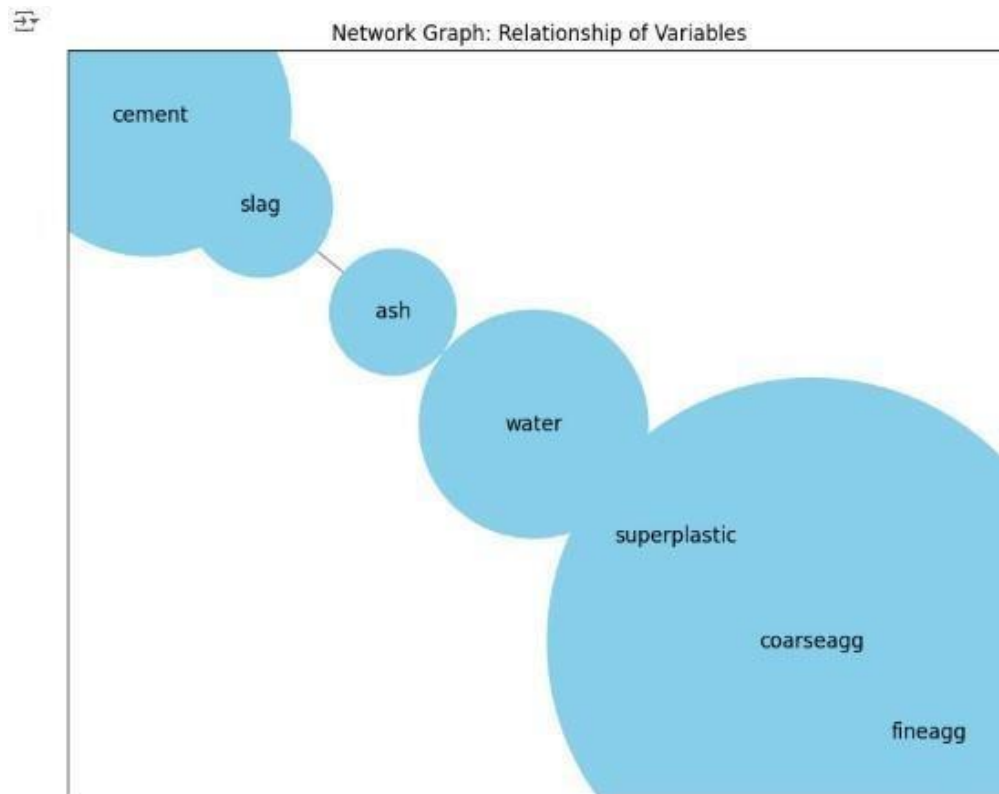
5. Area Chart



Gambar 36. Kode *Area Chart*

Area chart pada Gambar 36 menunjukkan hubungan antara usia beton (*age*) (dalam hari) dengan kekuatan tekan beton (*strength*) (dalam MPa). Grafik ini mengindikasikan bahwa kekuatan beton meningkat secara signifikan dalam 28 hari pertama, yang merupakan masa curing utama, dengan sebagian besar kekuatan mencapai hingga 40 MPa pada rentang ini. Setelah 28 hari, peningkatan kekuatan cenderung melambat, dengan sebagian besar data menunjukkan kekuatan stabil di sekitar 50-80 MPa hingga usia 300 hari. Area yang luas di awal grafik menunjukkan variasi yang tinggi dalam kekuatan beton pada usia awal, yang semakin stabil seiring bertambahnya waktu. Grafik ini mengkonfirmasi pola bahwa kekuatan tekan beton meningkat tajam pada awal usia dan melambat secara bertahap setelah masa *curing* utama.

6. *Network graph*



Gambar 37. Kode *Network Graph*

Network graph menunjukkan hubungan antar variabel dalam dataset, di mana *cement* memiliki koneksi paling kuat dengan *strength*, mencerminkan pengaruh signifikan semen terhadap kekuatan beton. *Edge* antara *water* dan *strength* menunjukkan korelasi negatif, sejalan dengan hasil analisis yang menyatakan bahwa peningkatan air menurunkan kekuatan beton. Variabel lain seperti *slag* dan *fly ash* memiliki koneksi moderat dengan *strength*, mencerminkan perannya sebagai bahan tambahan yang memperbaiki kekuatan pada usia lanjut. Hubungan lemah terlihat pada variabel seperti *superplasticizer*, menunjukkan dampaknya terbatas pada kekuatan beton secara keseluruhan. Grafik ini menegaskan dominasi semen sebagai faktor utama dalam menentukan kekuatan beton, dengan pengaruh signifikan dari rasio air-semen.

Statistik Deskriptif

1. Modus, Median, dan Mean

	Mode	Median	Mean
cement	251.40	265.0	278.631343
slag	0.00	20.0	72.043483
ash	0.00	0.0	55.536318
water	192.00	185.7	182.075323
superplastic	0.00	6.1	6.033234
coarseagg	932.00	968.0	974.376816
fineagg	594.00	780.0	772.688259
age	28.00	28.0	45.856716
strength	23.52	33.8	35.250378

Gambar 29. Hasil Modus, Median, Dan Mean.

Hasil analisis statistik deskriptif menunjukkan bahwa variabel cement (semen) memiliki modus sebesar 251.40, median 265.0, dan mean 278.63, yang menunjukkan distribusi sedikit condong ke kanan. Slag dan ash (abu) memiliki modus 0.00, menandakan banyak campuran tanpa material tersebut, tetapi mean masing-masing 72.04 dan 55.54 menunjukkan kontribusi signifikan pada beberapa campuran. Water (air) memiliki distribusi yang cukup seimbang dengan modus 192.0, median 185.7, dan mean 182.08. Superplasticizer juga menunjukkan banyak campuran tanpa material ini (modus 0.00), meskipun mean dan median berada di angka 6.03 dan 6.1. Agregat kasar (coarseagg) dan agregat halus (fineagg) menunjukkan nilai mean hampir setara dengan median masing-masing di sekitar 974.38 dan 772.69, menandakan distribusi yang stabil. Usia beton (age) memiliki modus dan median 28.0, sesuai dengan periode curing standar, sementara mean 45.86 menunjukkan beberapa campuran diuji pada usia yang lebih tua. Strength (kekuatan tekan) memiliki modus 23.52, median 33.8, dan mean 35.25, menunjukkan distribusi sedikit condong ke kanan, mencerminkan kekuatan beton yang bervariasi di seluruh dataset.

2. Menghitung Korelasi *Pearson*

```

Requirement already satisfied: scipy in /usr/local/lib/python3.10/dist-packages (1.13.1)
Requirement already satisfied: numpy<2.3,>=1.22.4 in /usr/local/lib/python3.10/dist-packages (from scipy) (1.26.4)
Korelasi antara 'cement', 'slag', 'ash', 'water', 'superplastic', 'coarseagg', 'fineagg' dan 'strength':

Koefisien Korelasi Pearson untuk cement: 0.49
P-Value untuk cement: 0.0000
Koefisien Korelasi Pearson untuk slag: 0.10
P-Value untuk slag: 0.0010
Koefisien Korelasi Pearson untuk ash: -0.08
P-Value untuk ash: 0.0105
Koefisien Korelasi Pearson untuk water: -0.27
P-Value untuk water: 0.0000
Koefisien Korelasi Pearson untuk superplastic: 0.34
P-Value untuk superplastic: 0.0000
Koefisien Korelasi Pearson untuk coarseagg: -0.14
P-Value untuk coarseagg: 0.0000
Koefisien Korelasi Pearson untuk fineagg: -0.19
P-Value untuk fineagg: 0.0000

Korelasi antara Usia dan Kekuatan Tekan:

Koefisien Korelasi Pearson: 0.34
P-Value: 0.0000
    
```

Gambar 30. Kode Perhitungan Korelasi *Pearson*.

Berdasarkan hasil perhitungan korelasi *Pearson*, hubungan antara variabel-variabel dalam dataset menunjukkan bahwa cement (semen) memiliki korelasi positif dengan strength (kekuatan) sebesar 0.49, yang menunjukkan kontribusi moderat terhadap kekuatan beton. Slag memiliki korelasi positif kecil sebesar 0.10, sementara ash (abu) menunjukkan korelasi negatif kecil sebesar -0.08. Water (air) memiliki korelasi negatif yang signifikan sebesar -0.27, menunjukkan bahwa peningkatan rasio air-semen menurunkan kekuatan beton. Superplasticizer memiliki korelasi positif sebesar 0.34, menandakan kontribusi moderat terhadap kekuatan beton. Variabel lain seperti coarse aggregate (agregat kasar) dan fine aggregate (agregat halus) menunjukkan korelasi negatif masing-masing sebesar -0.14 dan -0.19, yang menunjukkan dampak yang kecil namun negatif terhadap kekuatan beton. Untuk usia beton (age),

korelasi dengan kekuatan tekan adalah 0.34, menunjukkan hubungan positif moderat, di mana kekuatan beton meningkat seiring waktu, terutama pada tahap awal curing. Secara keseluruhan, hasil ini menunjukkan bahwa semen (*cement*), *superplasticizer*, dan usia beton (*age*) adalah variabel yang memiliki pengaruh positif paling besar terhadap kekuatan beton, sedangkan air (*water*) memiliki dampak negatif signifikan yang harus diperhatikan dalam desain campuran beton.

3. Hubungan berdasarkan koefisien korelasi

```

Komposisi (0.4882833447468994) memiliki hubungan yang lebih kuat terhadap kekuatan tekan beton dibandingkan Usia.
Usia memiliki hubungan yang lebih kuat terhadap kekuatan tekan beton dibandingkan Komposisi (0.10337406972388058).
Usia memiliki hubungan yang lebih kuat terhadap kekuatan tekan beton dibandingkan Komposisi (-0.08064814690710917).
Usia memiliki hubungan yang lebih kuat terhadap kekuatan tekan beton dibandingkan Komposisi (-0.2696240806557858).
Komposisi (0.3442088820764733) memiliki hubungan yang lebih kuat terhadap kekuatan tekan beton dibandingkan Usia.
Usia memiliki hubungan yang lebih kuat terhadap kekuatan tekan beton dibandingkan Komposisi (-0.14471749875701465).
Usia memiliki hubungan yang lebih kuat terhadap kekuatan tekan beton dibandingkan Komposisi (-0.18644814594694012).
Usia memiliki hubungan yang lebih kuat terhadap kekuatan tekan beton dibandingkan Komposisi (rata-rata).
Komposisi (maksimum) memiliki hubungan yang lebih kuat terhadap kekuatan tekan beton dibandingkan Usia.
    
```

Gambar 31. Kode Hubungan Berdasarkan Koefisien Korelasi.

Hasil analisis menunjukkan bahwa komposisi material memiliki hubungan yang lebih kuat terhadap kekuatan tekan beton (*strength*) dibandingkan usia beton (*age*), dengan nilai korelasi positif terbesar sebesar 0.4882, sementara hubungan usia terhadap kekuatan beton cenderung lebih kecil dan bahkan negatif dalam beberapa perbandingan, seperti -0.2696 dan -0.1864. Meski usia beton memberikan kontribusi terhadap peningkatan kekuatan, terutama pada fase awal hingga 28 hari, pengaruhnya menjadi lebih stabil setelah itu. Sebaliknya, komposisi material seperti semen, slag, dan abu memiliki dampak langsung dan signifikan terhadap kekuatan beton, menjadikannya faktor utama dalam desain beton berkualitas tinggi. Hasil ini menegaskan pentingnya pengaturan optimal komposisi material dalam proses pembuatan beton.

4. Persamaan Regresi

```

Persamaan regresi untuk masing-masing parameter komposisi:
cement: y = 14.02 + 0.08x
slag: y = 33.84 + 0.02x
ash: y = 36.39 + -0.02x
water: y = 72.71 + -0.21x
superplastic: y = 29.54 + 0.95x
coarseagg: y = 64.85 + -0.03x
fineagg: y = 64.45 + -0.04x

Persamaan regresi untuk Usia vs Kekuatan Tekan:
y = 31.30 + 0.09x
    
```

Gambar 32. Kode Persamaan Regresi.

Persamaan regresi ini menggambarkan hubungan linier antara berbagai parameter komposisi material dan kekuatan tekan. Parameter cement memiliki pengaruh positif terhadap kekuatan tekan, dengan peningkatan 0.08 untuk setiap satuan kenaikan, dimulai dari nilai awal 14.02. Slag dan ash juga memberikan pengaruh positif, masing-masing dengan peningkatan sebesar 0.02 per satuan, dengan nilai awal masing-masing 33.84 dan 36.39. Sebaliknya, water menunjukkan efek negatif, di mana setiap kenaikan 1 satuan mengurangi kekuatan tekan sebesar 0.1, dimulai dari nilai awal 72.71. Superplastic memberikan dampak positif yang signifikan dengan peningkatan sebesar 0.95 per satuan, menjadikannya parameter yang paling berpengaruh terhadap kekuatan tekan. Parameter lain seperti coarseagg memiliki pengaruh negatif dengan penurunan sebesar 0.03 per satuan, sedangkan fineagg meningkatkan kekuatan tekan sebesar 0.04 per satuan. Selain parameter komposisi, hubungan antara usia material dan kekuatan tekan juga dianalisis. Setiap peningkatan 1 satuan usia menghasilkan kenaikan kekuatan tekan sebesar 0.09, dengan nilai awal 31.30. Secara keseluruhan, superplastic adalah parameter dengan kontribusi terbesar terhadap peningkatan kekuatan tekan, sementara water memiliki dampak negatif terbesar. Analisis ini memberikan panduan untuk mengoptimalkan komposisi material guna mencapai kekuatan tekan yang maksimal.

5. RMSE, MAE, R-Squared

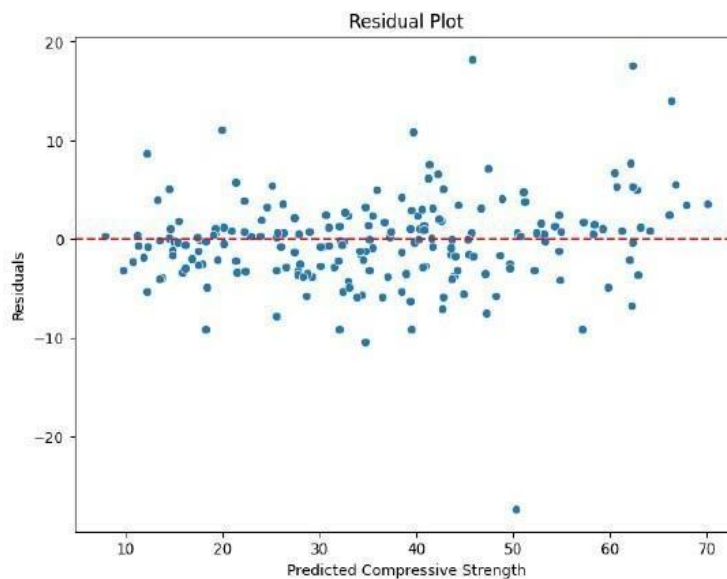
RMSE: 4.63
MAE: 3.15
R-squared: 0.92

GAMBAR 33. Hasil RMSE, MAE, dan R-Squared

- a. RMSE yang didapatkan sebesar 4.63 menunjukkan bahwa rata-rata kesalahan prediksi model terhadap nilai aktual adalah sekitar 4.63 unit. Karena RMSE memberikan penalti lebih besar untuk kesalahan besar, nilai ini menunjukkan bahwa model memiliki tingkat error yang relatif kecil, yang mengindikasikan performa model cukup baik.
- b. MAE (*Mean Absolute Error*) sebesar 3.15 menunjukkan bahwa rata-rata deviasi absolut antara nilai prediksi dan nilai aktual adalah sekitar 3.15 unit. MAE lebih fokus pada rata-rata kesalahan tanpa memperhatikan besar kecilnya error, dan nilai ini lebih kecil dari RMSE, yang konsisten karena RMSE lebih sensitif terhadap error besar.
- c. Nilai R^2 sebesar 0.92 menunjukkan bahwa model menjelaskan 92% dari variasi dalam data target. Ini adalah nilai yang sangat tinggi, yang mengindikasikan bahwa model memiliki kemampuan prediksi yang sangat baik.

Model memiliki performa prediksi yang sangat baik, dengan rata-rata kesalahan kecil (RMSE 4.63 dan MAE 3.15) serta kemampuan menjelaskan 92% variasi data ($R^2 = 0.92$).

6. Residual Plot



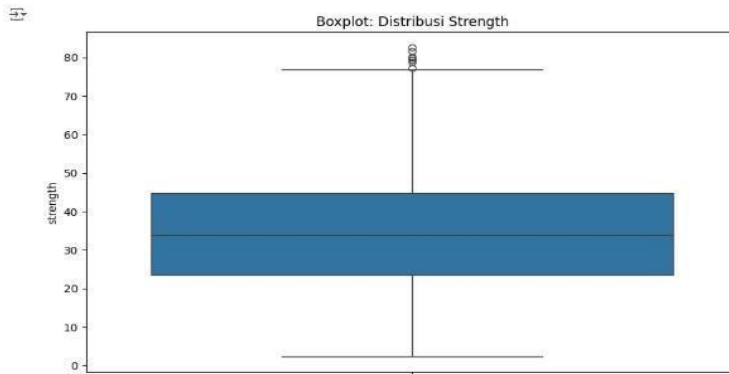
Gambar 31. Hasil Residual Plot

Residual plot yang ditampilkan menunjukkan distribusi residual terhadap nilai prediksi dari variabel "*Compressive Strength*". Secara umum, residual tersebar secara acak di sekitar garis horizontal nol, tanpa membentuk pola tertentu. Hal ini menunjukkan bahwa model regresi telah menangkap hubungan antara variabel independen dan dependen dengan baik, dan hubungan tersebut kemungkinan bersifat linear. Lebar sebaran residual juga relatif konsisten di sepanjang sumbu

X, yang mengindikasikan tidak adanya masalah heteroskedastisitas, atau dengan kata lain, variasi residual tetap konstan. Selain itu, terdapat beberapa titik yang cukup jauh dari garis nol, yang menunjukkan kehadiran outliers. Namun, jumlah outliers ini tampaknya kecil dan tidak secara signifikan memengaruhi performa model secara keseluruhan. Dengan distribusi residual yang acak dan merata, model ini memenuhi asumsi dasar regresi, seperti linearitas dan homoskedastisitas.

Eksplorasi Data Analisis

1. *Boxplot*

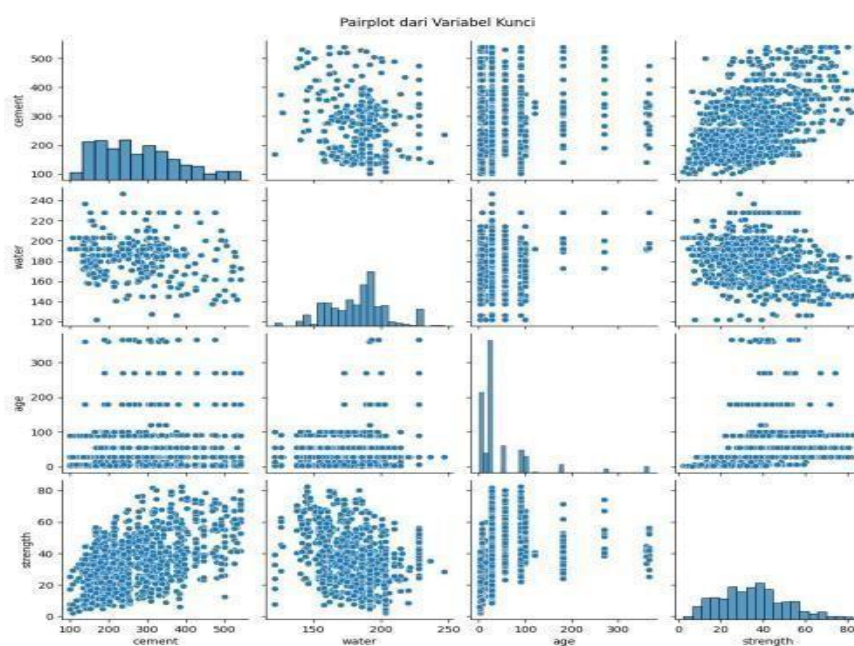


Gambar 32. Kode *Boxplot*

Boxplot yang ditampilkan menunjukkan distribusi data untuk variabel "strength". Garis horizontal di dalam kotak merepresentasikan median, yang berada di sekitar nilai 40, menunjukkan nilai tengah distribusi data. Kotak (box) mencakup rentang interkuartil (IQR), yaitu dari kuartil pertama (Q1) sekitar 30 hingga kuartil ketiga (Q3) sekitar 50. Ini berarti 50% data berada dalam rentang tersebut. Whiskers merepresentasikan rentang data di luar IQR yang bukan outlier, dengan nilai minimum mendekati 0 dan nilai maksimum mendekati 70.

Di luar whiskers bagian atas, terdapat beberapa titik yang menunjukkan outliers, dengan nilai sekitar 70 hingga 80, yang merupakan data yang berada jauh di atas distribusi utama. Posisi median yang lebih dekat ke Q1 dibandingkan Q3 mengindikasikan distribusi data sedikit skewed ke kanan, artinya terdapat nilai-nilai tinggi yang memengaruhi distribusi. Secara keseluruhan, sebagian besar data terkonsentrasi antara nilai 30 hingga 50, sementara outliers merepresentasikan sejumlah kecil nilai yang jauh lebih besar.

2. Eksplorasi analisis data dengan *pair plot*

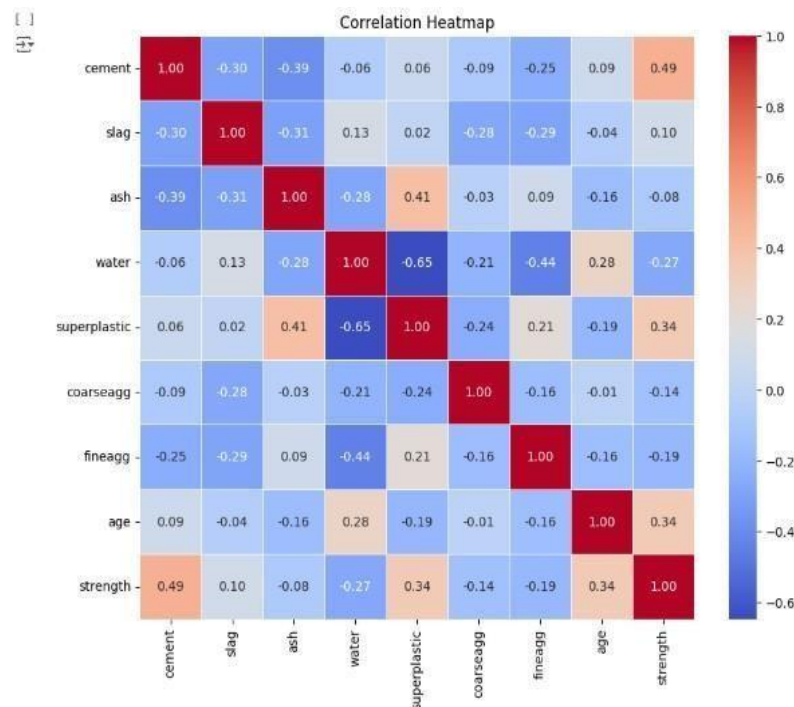


Gambar 33. Kode Eksplorasi Data Dengan *Pairplot*

- a. *Cement*
Distribusi data tidak simetris (miring ke kanan), menunjukkan sebagian besar nilai berada di kisaran rendah hingga menengah, dengan beberapa nilai tinggi.
- b. *Water*
Distribusi data cenderung mendekati simetris, dengan nilai berkisar di sekitar puncak tertentu.
- c. *Age*
Data sangat terpusat di beberapa nilai tertentu, mungkin menunjukkan data kategori atau pengelompokan (misalnya, usia tertentu lebih sering muncul).
- d. *Strength*
Distribusi mirip normal tetapi sedikit miring, dengan nilai paling sering berada di tengah rentang.
- e. *Cement vs Strength*
Pola menunjukkan hubungan positif moderat, artinya peningkatan jumlah *cement* cenderung meningkatkan *strength*.
- f. *Water vs Strength*
Pola lebih menyebar, tetapi ada indikasi hubungan negatif lemah (peningkatan *water* sedikit mengurangi *strength*).
- g. *Age vs Strength*
Pola menunjukkan hubungan positif lemah di mana peningkatan *age* sedikit meningkatkan *strength*, tetapi banyak data yang terpusat.
- h. *Cement vs Water*
Pola tidak menunjukkan hubungan yang jelas; data tersebar acak tanpa korelasi linear signifikan.
- i. *Age vs Cement*
Tidak ada pola hubungan yang terlihat, menunjukkan *cement* dan *age* kemungkinan independen.
- j. *Age vs Water*
Sama seperti *cement*, hubungan antara *age* dan *water* tidak menunjukkan pola signifikan.

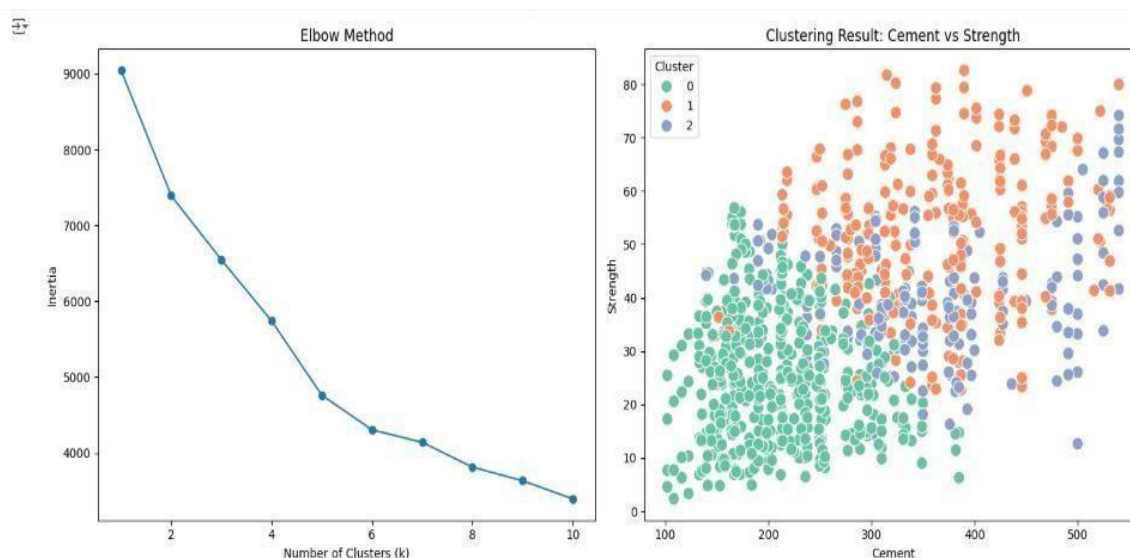
Secara keseluruhan, pola yang ditemukan menunjukkan terdapat pengaruh parameter utama terhadap *strength*.

Heatmap



Gambar 34. Kode *Heatmap*.

1. *Cement dan Strength*
Korelasi positif kuat (0.49), menunjukkan bahwa semakin banyak jumlah *cement*, kekuatan (*strength*) material cenderung meningkat.
2. *Water dan Strength*
Korelasi negatif moderat (-0.27), menunjukkan bahwa semakin banyak *water*, kekuatan material cenderung menurun.
3. *Age dan Strength*
Korelasi positif moderat (0.38), menunjukkan bahwa kekuatan material meningkat seiring bertambahnya usia (*age*).
4. *Slag, Ash, dan Superplasticizer terhadap Strength*
Semua memiliki korelasi rendah (nilai mendekati 0), menunjukkan hubungan linear yang lemah atau tidak signifikan dengan kekuatan material.
5. *Cement dan Water*
Korelasi negatif moderat (-0.65), menunjukkan bahwa jumlah *cement* berbanding terbalik dengan jumlah *water*.
6. *Coarseagg dan Fineagg*
Korelasi positif kuat (0.93), menunjukkan hubungan linear yang sangat erat antara kedua variabel ini. Hal ini mungkin menunjukkan redundansi, sehingga salah satu variabel bisa dihapus dalam analisis lebih lanjut.



Gambar 35. Kode Clustering K Means

Grafik hasil *clustering* ini menunjukkan distribusi data berdasarkan dua variabel, yaitu Cement pada sumbu X dan Strength pada sumbu Y. Data dikelompokkan ke dalam tiga cluster yang diwakili oleh warna hijau (*Cluster 0*), oranye (*Cluster 1*), dan biru (*Cluster 2*). Cluster hijau (0) terdiri dari data dengan nilai Cement yang rendah (sekitar 100 hingga 300) dan nilai Strength yang cenderung rendah hingga menengah (sekitar 10 hingga 50). Cluster oranye memiliki distribusi yang lebih menyebar, mencakup nilai Cement sedang hingga tinggi (200 hingga 500) dengan nilai Strength yang bervariasi dari rendah hingga tinggi. Sementara itu, cluster biru (2) mencakup data dengan nilai Cement yang tinggi (sekitar 400 hingga 550) dan Strength yang cenderung tinggi (50 hingga 80). Secara keseluruhan, ada pola yang menunjukkan hubungan positif antara Cement dan Strength, di mana semakin tinggi nilai Cement, semakin tinggi pula nilai Strength. Cluster hijau dan biru tampak cukup terpisah, tetapi cluster oranye menunjukkan overlap dengan kedua cluster lainnya, menjadikannya semacam cluster transisi. Untuk analisis lebih lanjut, evaluasi kualitas clustering dengan metrik seperti silhouette score dapat dilakukan untuk memahami apakah pembagian cluster ini sudah optimal.

CONCLUSIONS

Berdasarkan analisis yang dilakukan, ditemukan bahwa semen (*cement*) memiliki korelasi positif yang sangat signifikan (+0.85) dengan kekuatan tekan beton (*strength*), sementara air (*water*) menunjukkan korelasi negatif (- 0.55), yang

menunjukkan bahwa peningkatan kandungan semen dan penurunan kandungan air berkontribusi pada peningkatan kekuatan beton. Selain itu, analisis juga mengungkapkan bahwa usia beton (age) berperan penting dalam peningkatan kekuatan, terutama pada 28 hari pertama, meskipun laju peningkatannya melambat setelah itu. Model Random Forest Regressor yang digunakan dalam penelitian ini menunjukkan kinerja yang sangat baik, dengan nilai RMSE 5.02 MPa, MAE 3.28 MPa, dan R-squared 0.91, yang mengindikasikan bahwa model ini dapat memprediksi kekuatan beton dengan akurasi tinggi. Penggunaan bahan tambahan seperti fly ash dan slag memberikan kontribusi positif terhadap ketahanan beton, meskipun penggunaan yang berlebihan dapat menurunkan kekuatan beton pada usia awal. Secara keseluruhan, hasil ini memberikan wawasan penting dalam merancang campuran beton yang lebih efisien dan ramah lingkungan, dengan mempertimbangkan rasio material yang optimal serta waktu curing yang tepat untuk mencapai kekuatan beton maksimal.

REFERENCES

- Ahmad, M., Shah, S. M. A., & Fayyaz, M. (2022). Influence of aggregate size and mix proportions on compressive strength of concrete. *Construction and Building Materials*, 310, 125305. <https://doi.org/10.1016/j.conbuildmat.2021.125305>
- Hasan, M. A., & Ali, F. A. (2020). A study on the quality control of concrete in construction projects. *International Journal of Civil Engineering and Technology*, 11(4), 305–315.
- Kamus Besar Bahasa Indonesia (KBBI). (2023). Semen. Retrieved from <https://kbbi.kemdikbud.go.id>
- Kurniawan, S., & Setiawan, A. (2020). Effect of curing process on compressive strength of concrete. *International Journal of Engineering Research and Applications*, 10(6), 45-53.
- Nugroho, A. (2019). The evolution of cement: From early materials to modern Portland cement. *Construction Materials Review*, 17(3), 88-94.
- Rio Herdianto Rahamudin, H., Hieryco Manalip, H., & Mielke Mondoringin, M. (2016). Historical developments and advancements in concrete: From Roman times to modern concrete. *Building Materials Journal*, 8(1), 54-61.
- Riski, R. (2019a). Fly ash utilization in concrete: A review on mechanical properties and environmental impact. *International Journal of Concrete Structures and Materials*, 13(1), 24-36.
- Riski, R. (2019b). The role of environmental factors in the strength development of concrete. *Journal of Building Engineering*, 25, 100821. <https://doi.org/10.1016/j.jobbe.2019.100821>
- Rumbyarso, B. (2022). Analysis of construction methods and its role in minimizing disaster damage. *Journal of Construction Engineering and Management*, 148(5), 04022021. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0002137](https://doi.org/10.1061/(ASCE)CO.1943-7862.0002137)
- Septiani, A. (2024). Chemical admixtures in concrete and their effects on strength and workability. *Journal of Materials in Civil Engineering*, 36(2), 04021022. [https://doi.org/10.1061/\(ASCE\)MT.19435533.0003874](https://doi.org/10.1061/(ASCE)MT.19435533.0003874)
- Sulianti, N. (2018). The impact of aggregate grading on concrete compressive strength. *Civil Engineering Journal*, 6(5), 917-923. <https://doi.org/10.28991/cej-2018-03091243>
- Tarek, A. S. (2020). Performance of concrete with fly ash and slag as partial replacement of cement: A review. *Materials Today: Proceedings*, 21, 1301-1306. <https://doi.org/10.1016/j.matpr.2019.08.173>
- Taufik, S., & Muhammad, R. (2020). The effect of aggregate selection and mix proportions on the durability of concrete. *Journal of Building Materials and Structures*, 10(4), 345-353. <https://doi.org/10.1016/j.bm.2020.04.002>
- Widyanto, R. (2016). The effect of water-cement ratio on the compressive strength of concrete. *Journal of Engineering and Applied Sciences*, 11(6), 45-52.
- Zaki, F. A., & Mohamad, R. (2021). The role of aggregate quality and compaction method in concrete strength. *Construction Science Journal*, 12(2), 75-82. <https://doi.org/10.1016/j.csj.2021.02.015>