



**TEAM DIAGRAM**

# Analisis Kejadian Banjir Jakarta

*Naoby Prawira  
Haura Adzkia Delfina  
Syifa Salsabila*

*Seleksi Data Science Academy  
COMPFEST 2023*

*Dataset Kejadian Banjir Provinsi  
DKI Jakarta*

# Team Members



**Naoby Prawira**  
S1 Informatika  
Telkom University



**Haura Adzkia Delfina**  
S1 Data Sains  
Telkom University



**Syifa Salsabila**  
S1 Data Sains  
Telkom University

## Latar Belakang

Banyak faktor yang berdampak terhadap peristiwa banjir yang tidak kunjung selesai di Provinsi DKI Jakarta, Indonesia. Dari sekian banyak faktor tersebut, curah hujan merupakan salah satu yang memiliki dampak signifikan terhadap peristiwa tersebut. Ketinggian genangan air pada saat banjir akan berdampak langsung kepada aspek kehidupan lainnya. Dengan adanya analisis terhadap data kejadian banjir dapat membantu untuk memperkirakan dampak yang ditimbulkan oleh curah hujan dalam jumlah tertentu.



# Rumusan Masalah

- 
1. Bagaimana pola perubahan rata-rata ketinggian genangan air saat banjir di setiap bulan dari tahun ke tahun?
  2. Bagaimana pola perubahan curah hujan bulanan di tiap tahunnya?
  3. Bagaimana hubungan antara curah hujan dengan tinggi genangan air saat banjir?

# Rumusan Hipotesis

- 
1. Rata-rata ketinggian genangan air pada tiap bulannya fluktuatif didasari oleh banyak faktor, khususnya curah hujan.
  2. Curah hujan bergantung pada musim yang sedang berlangsung di Indonesia
  3. Terdapat hubungan positif antara curah hujan dan rata-rata ketinggian genangan air saat terjadi banjir.

# Tujuan



1. Menganalisis pola perubahan ketinggian banjir bulanan dari tahun ke tahun.
2. Menyelidiki hubungan antara curah hujan dan tinggi muka air banjir.
3. Mengamati dampak tinggi muka air terhadap konsekuensi banjir.
4. Mengidentifikasi pola perubahan jumlah RW terdampak banjir dari tahun ke tahun.

## Data Assessment



Dataset curah hujan dan kejadian banjir diperoleh dari data Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) serta Badan Penanggulangan Bencana Daerah (BPBD) DKI Jakarta. Dalam dataset yang dirangkum ini, terdapat volume curah hujan disertai dengan dampak yang ditimbulkan. Dataset berisi data rekapitulasi kejadian banjir tiap bulannya, mulai dari jiwa terdampak, rata-rata ketinggian air, dan informasi lainnya yang digabungkan dengan dataset yang berisi data curah hujan bulanan. Dataset yang dirangkum memiliki 60 *record* dengan 17 atribut.

# **Data Preprocessing**

## **Reduksi Data**



Pada reduksi data ini, data yang digunakan akan dipilih sesuai kebutuhan penelitian dengan menghapus kolom yang tidak berhubungan atau memiliki value yang sama di setiap record.



## **Replace Missing Values**

Melakukan input terhadap missing values dengan median di atribut terkait.

## **Transformasi Data**

Pada proses ini dilakukan konversi data terhadap format nilai yang tidak relevan agar memenuhi homogenitas ragam dan sebaran data menjadi normal.



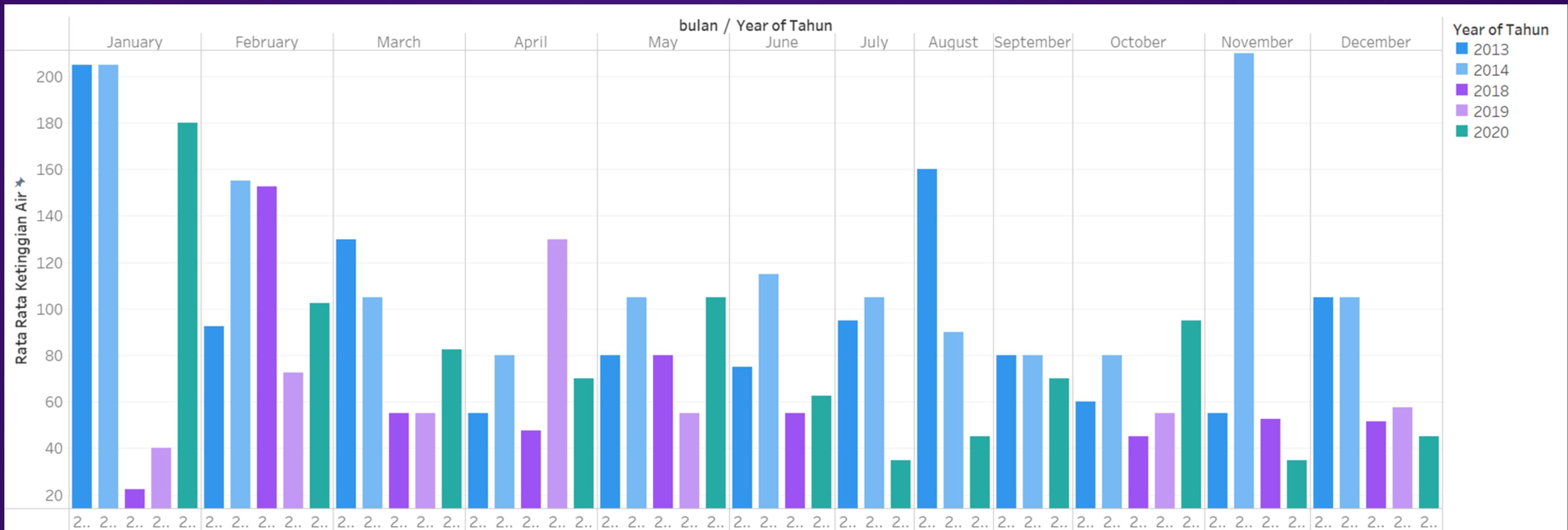
## **Normalisasi Data**

Normalisasi data dilakukan agar atribut data dalam skala yang sama saat diproses.



# Exploratory Data Analysis (EDA)

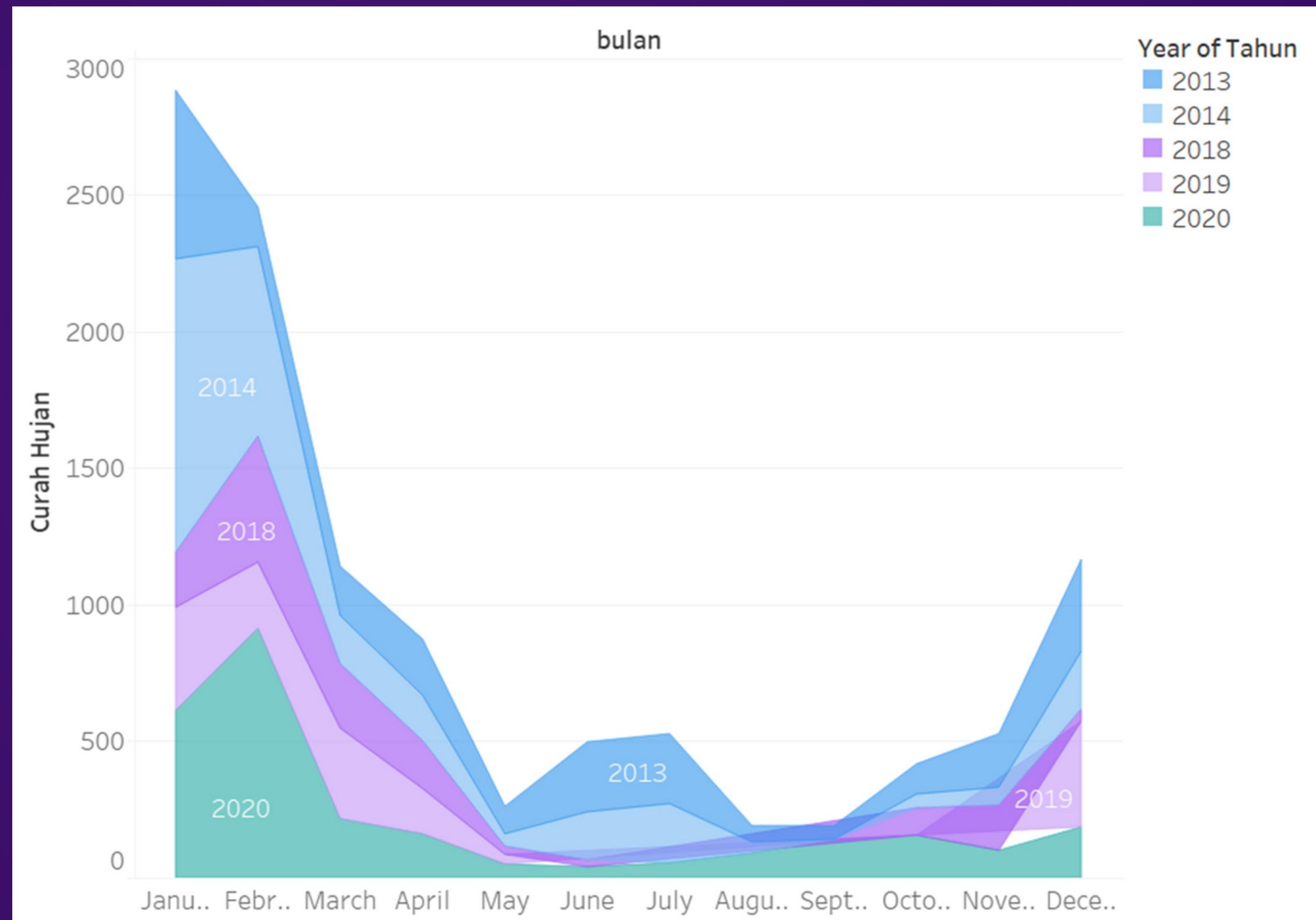
## Pola Tinggi Genangan Air



Rata Rata Ketinggian Air for each Tahun Year broken down by bulan Month. Color shows details about Tahun Year.

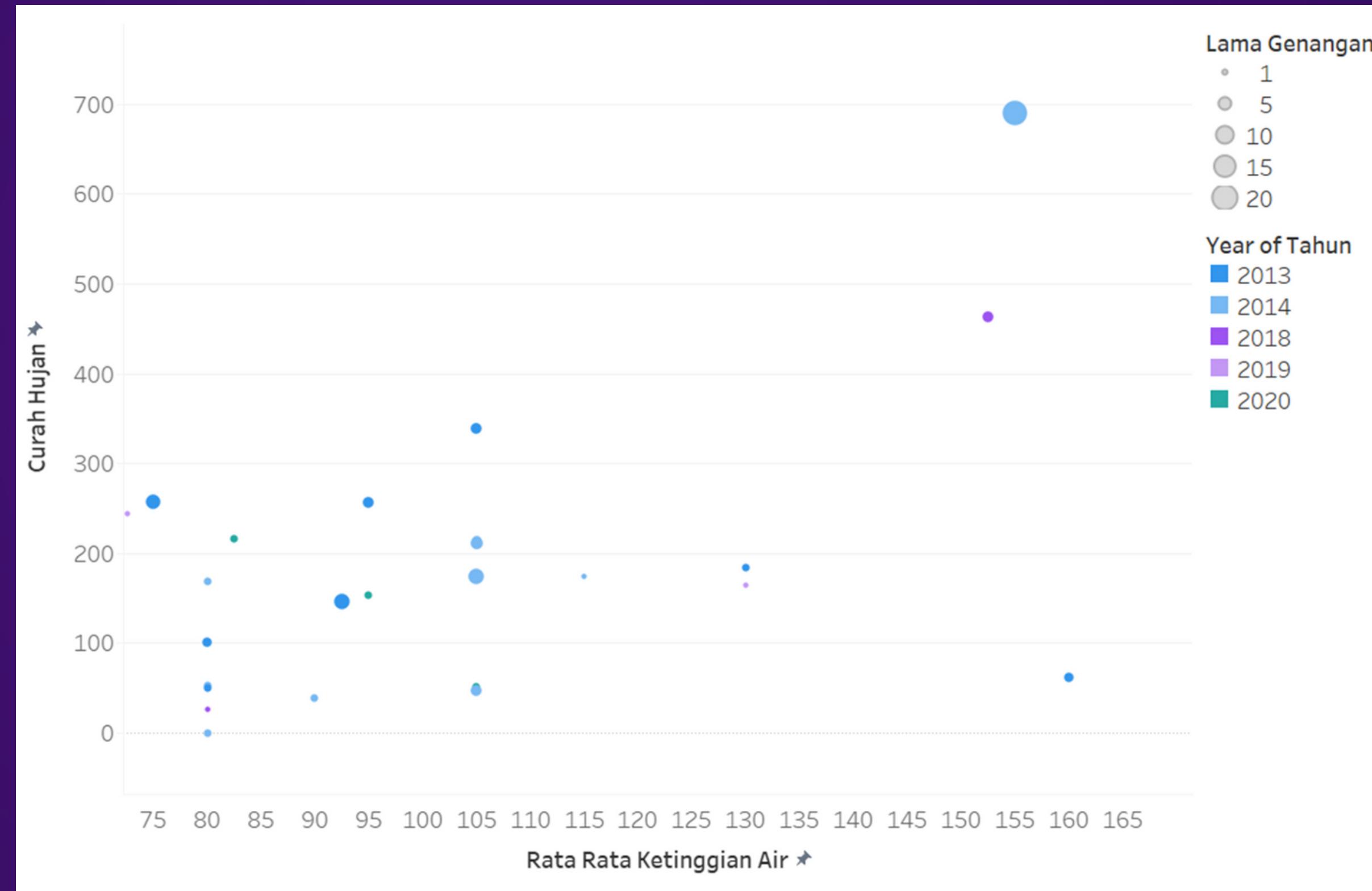
# Exploratory Data Analysis (EDA)

## Pola Curah Hujan



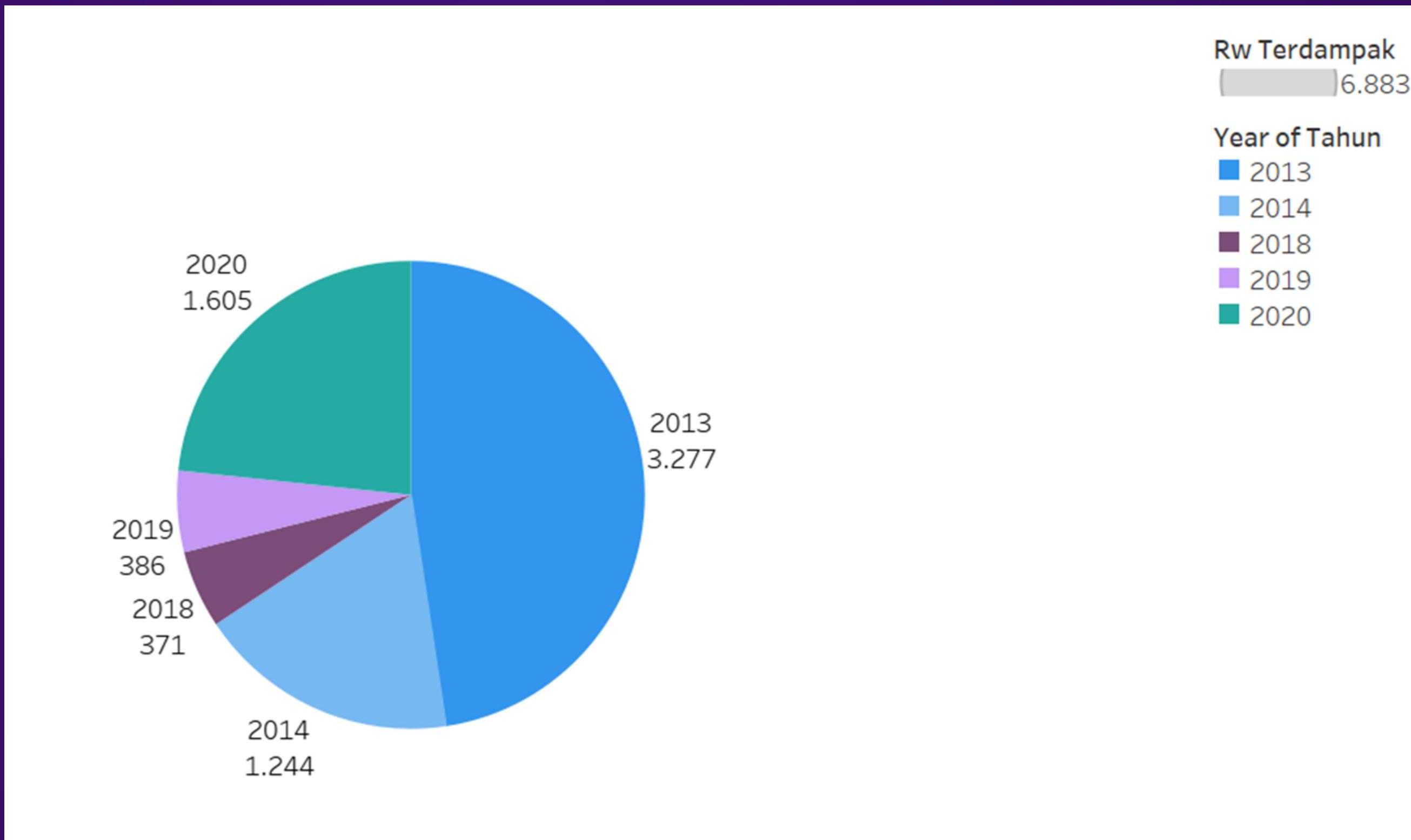
# Exploratory Data Analysis (EDA)

## Hubungan Antara Curah Hujan dengan Tinggi Genangan Air



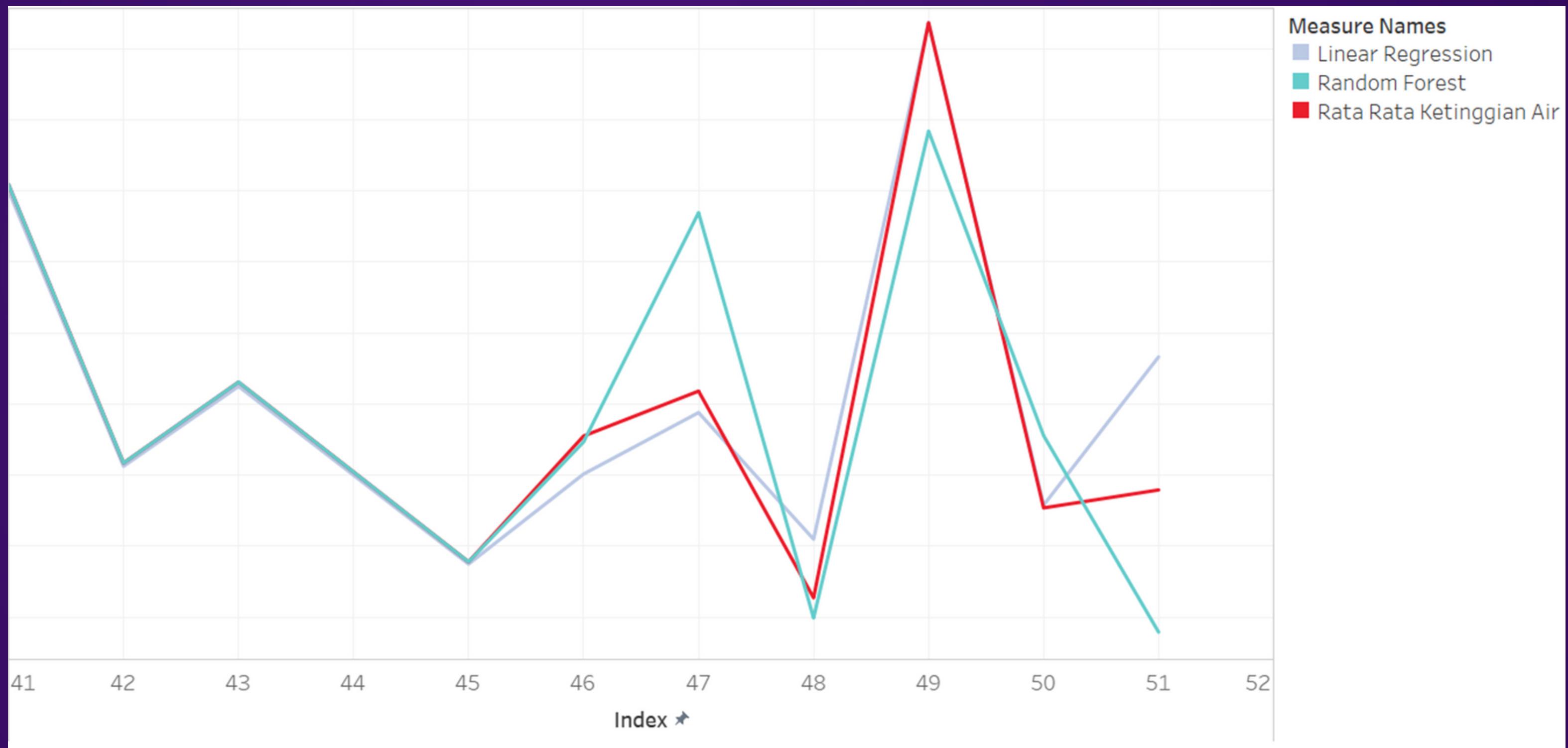
# Exploratory Data Analysis (EDA)

## Jumlah RW Terdampak

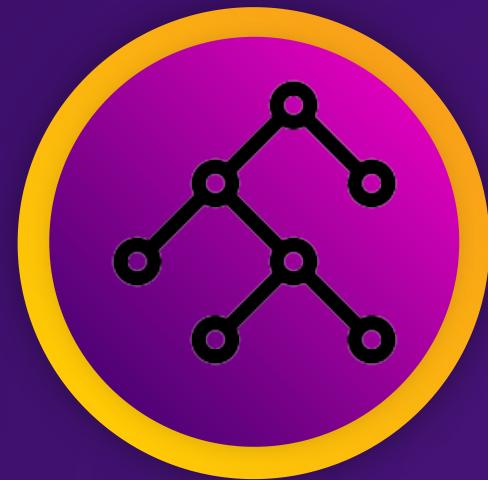


# Exploratory Data Analysis (EDA)

## Perbandingan Hasil Prediksi dengan Data Aktual



# Prediction Models



## Random Forest Regressor

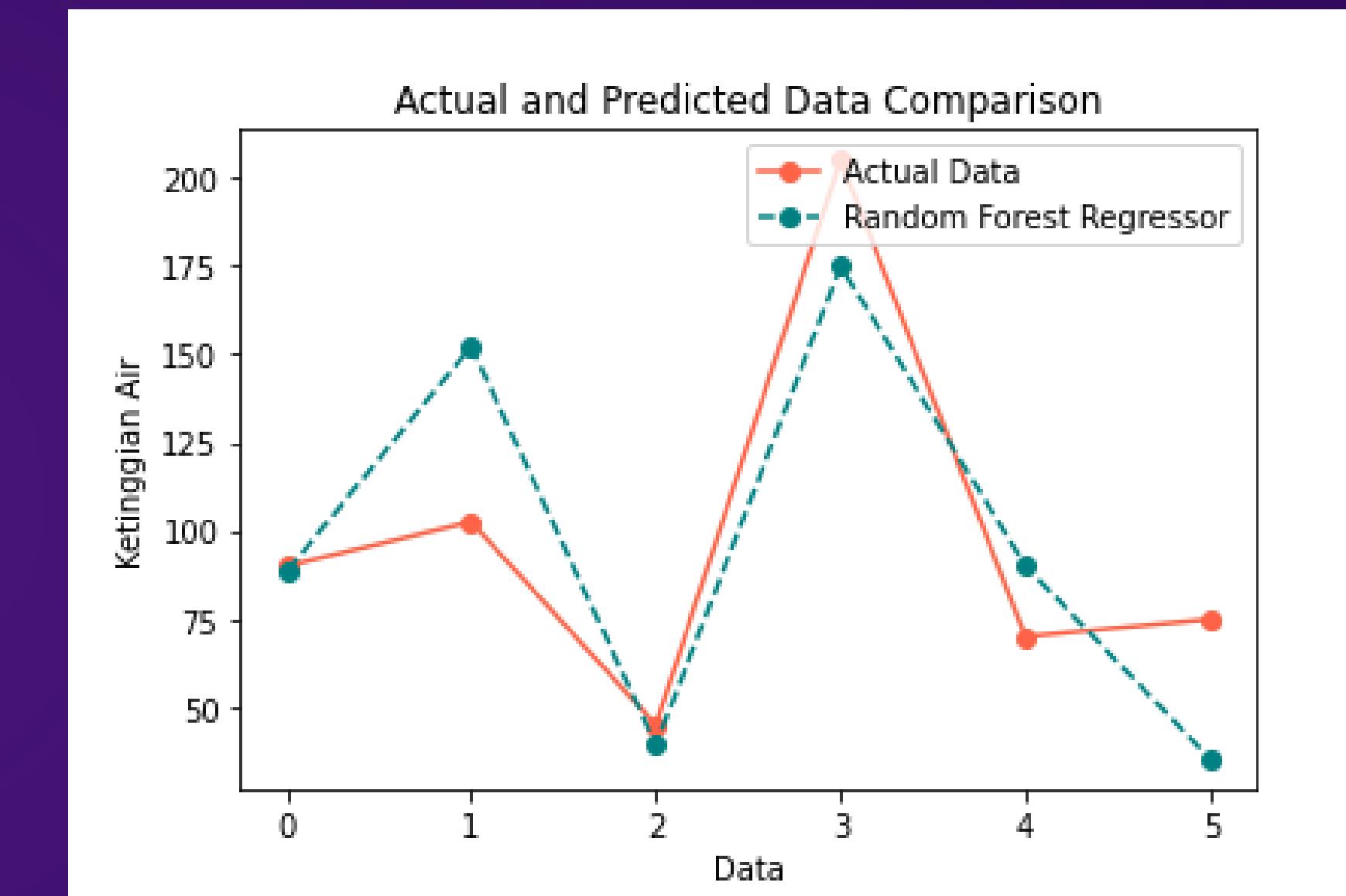
Model algoritma ini bekerja dengan menggunakan gabungan dari banyak *decision tree* untuk melakukan prediksi. Dengan penggabungan tersebut, model dapat memberikan prediksi yang akurat dan stabil.



## Linear Regression

Model algoritma ini bekerja dengan melihat pola hubungan antara variabel prediktor dan variabel dependen. Penggunaan ini berdasar dengan adanya asumsi hubungan linier antara data atribut dengan target prediksi.

# Random Forest Regressor

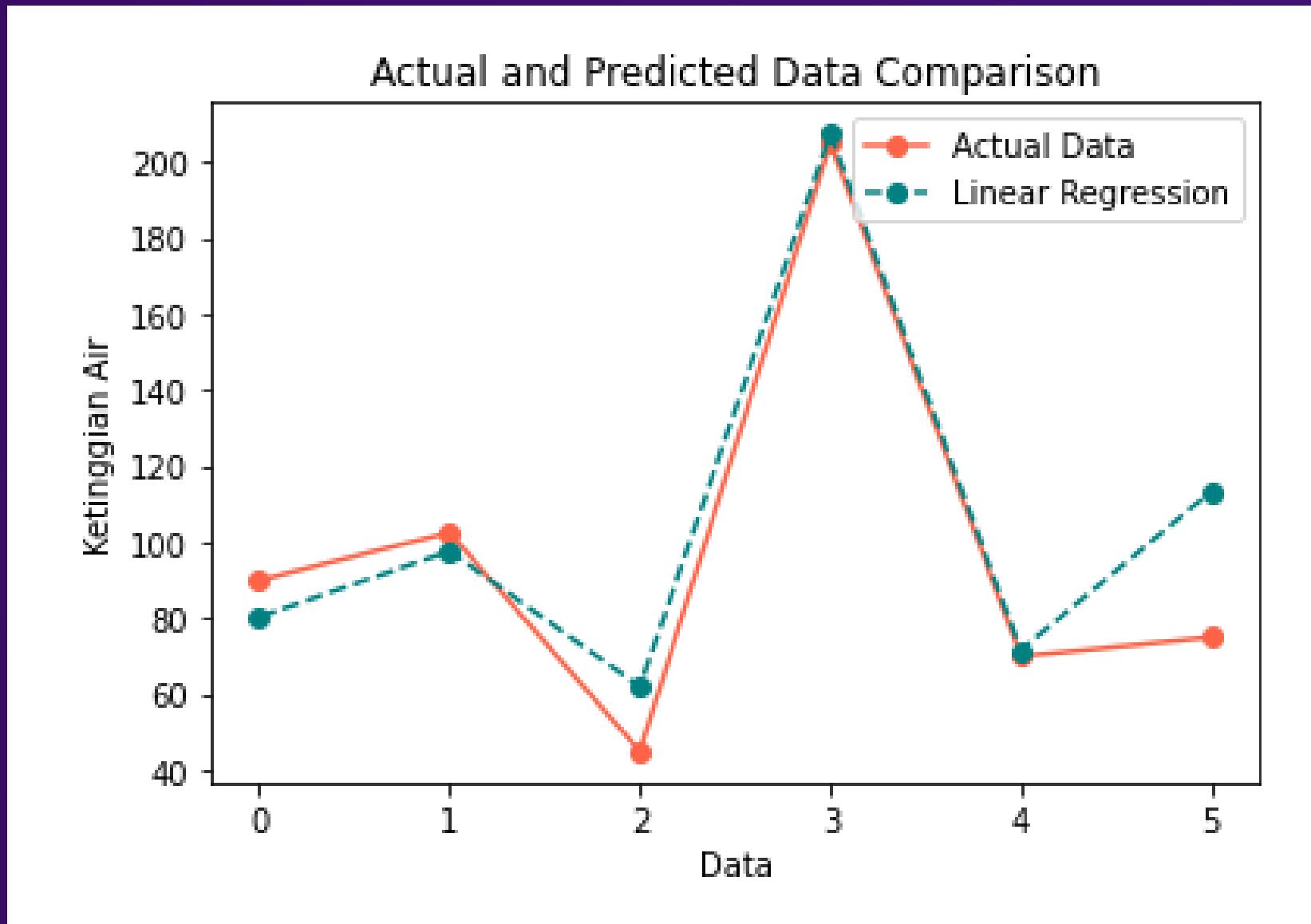


MSE : 896.2783375000002

MAE : 24.4575

R2 Score : 0.6564932314857941

# Linear Regression



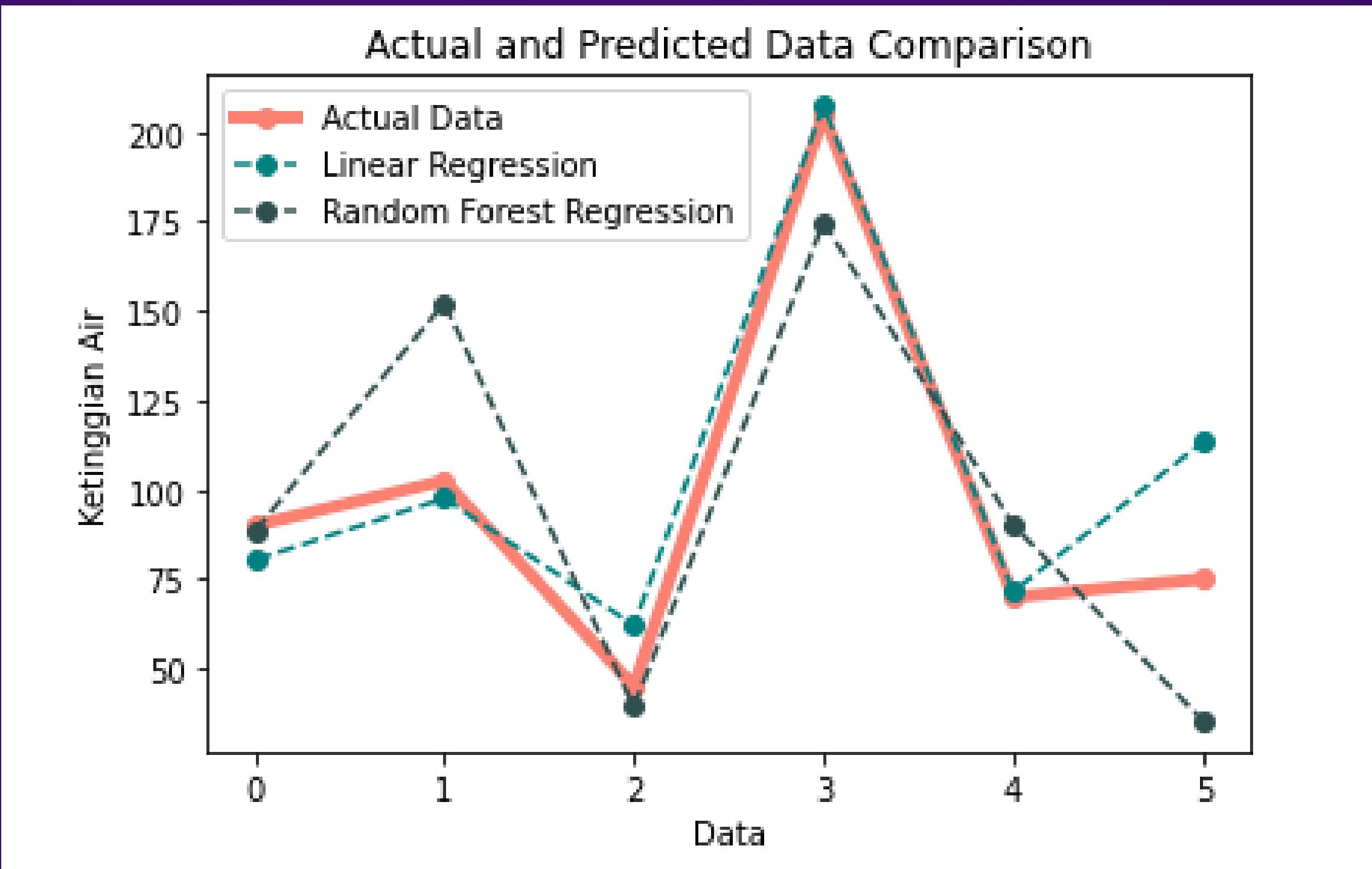
MSE: 313.6074308105866

MAE: 12.29304323765101

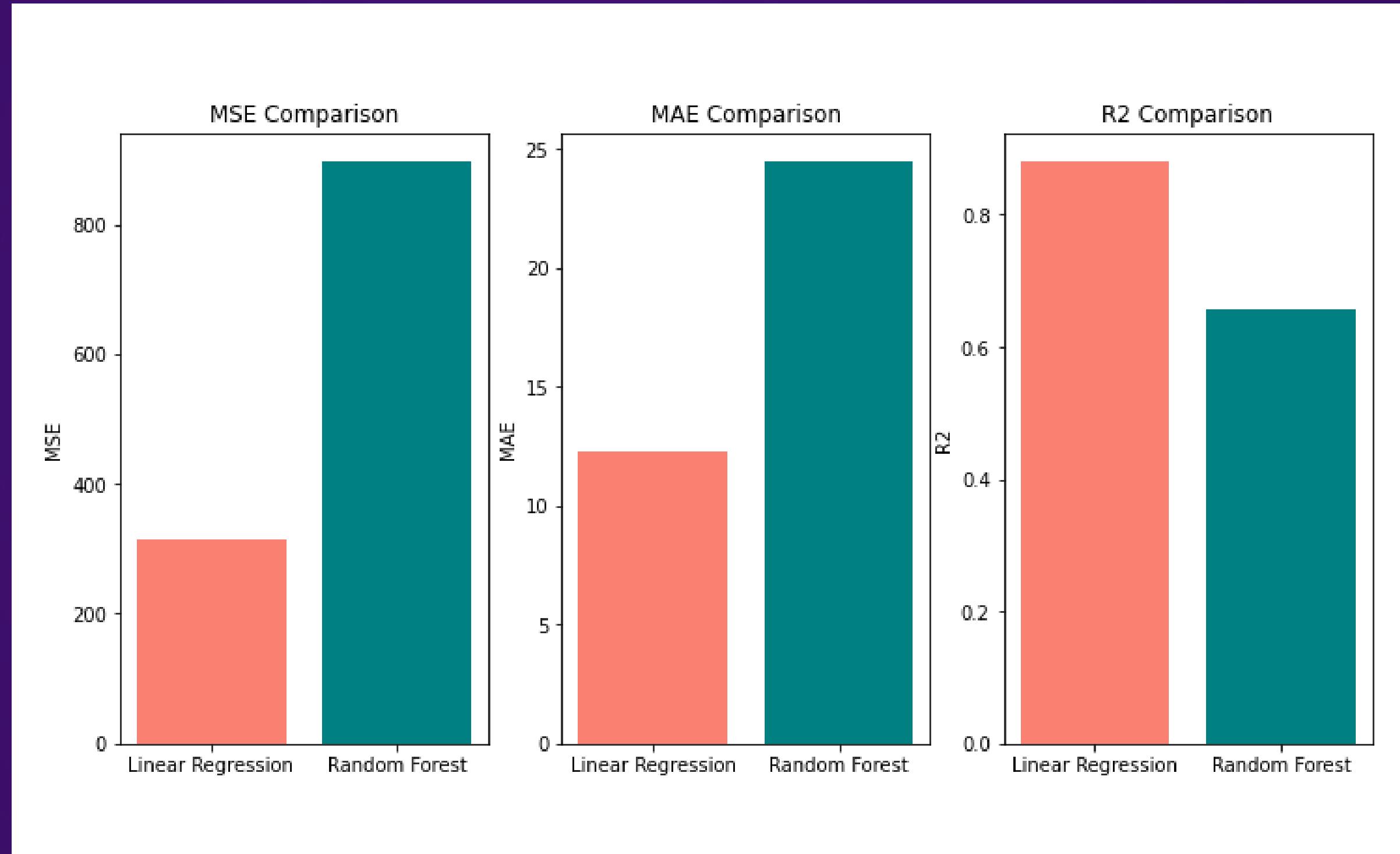
R2 Score: 0.8798071194710906



# Models Performance Comparison



# Models Performance Comparison



## KESIMPULAN

Berdasarkan hasil analisis dan prediksi, terdapat beberapa poin yang dapat disimpulkan, yaitu:

- Terdapat hubungan linier positif antara atribut fitur terhadap atribut target (genangan air) pada dataset rekapitulasi kejadian banjir DKI Jakarta
- Data yang tersedia tidak kompatibel (terlalu sedikit dan tidak balance) untuk dimodelkan ke dalam model machine learning, sehingga besar kemungkinan terjadinya overfitting/underfitting.
- Model prediksi terbaik untuk kasus dataset ini adalah dengan menggunakan Regresi Linear dengan performa yang sedikit lebih baik dari model Random Forest.

## Rekomendasi

Setelah melakukan penelitian, terdapat beberapa rekomendasi yang dapat diberikan, yaitu:

- Untuk pihak yang bertanggung jawab terkait pengumpulan data, dapat menetapkan standar pencatatan data yang baik dan terbuka sehingga data yang dikumpulkan dapat diolah oleh masyarakat umum dengan harapan dapat membantu menyelesaikan permasalahan di masyarakat.
- Untuk peneliti selanjutnya, dapat mengumpulkan data yang lebih lengkap dengan permintaan ke instansi terkait, dikarenakan data yang tersebar secara daring tidak tersaji secara lengkap dan tercatat dengan baik.

A circular inset photograph showing two people in business attire shaking hands. The person on the left is wearing a dark suit and a white shirt. The person on the right is wearing a light-colored shirt. They are both wearing dark trousers. The background is blurred, suggesting an indoor office environment.

Terima  
Kasih