

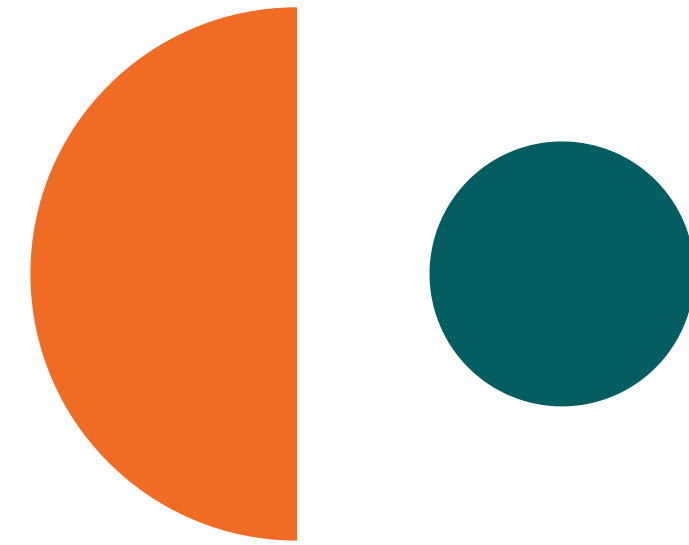
# 50 STARUPS

KELOMPOK 4

# DATASET

Dataset yang digunakan dalam analisis ini adalah 50\_Startups, yang berisi informasi tentang investasi perusahaan rintisan dan keuntungan yang diperoleh. Variabel yang tersedia dalam dataset ini antara lain:

- R&D Spend: Investasi dalam penelitian dan pengembangan.
  - Administration: Biaya administrasi perusahaan.
  - Marketing Spend: Biaya pemasaran.
  - State: Lokasi perusahaan (New York, California, Florida).
  - Profit: Keuntungan yang diperoleh.
- 



# HASIL ANALISIS

## Tabel perbandingan Metrik Model

=== TABEL PERBANDINGAN METRIK MODEL ===

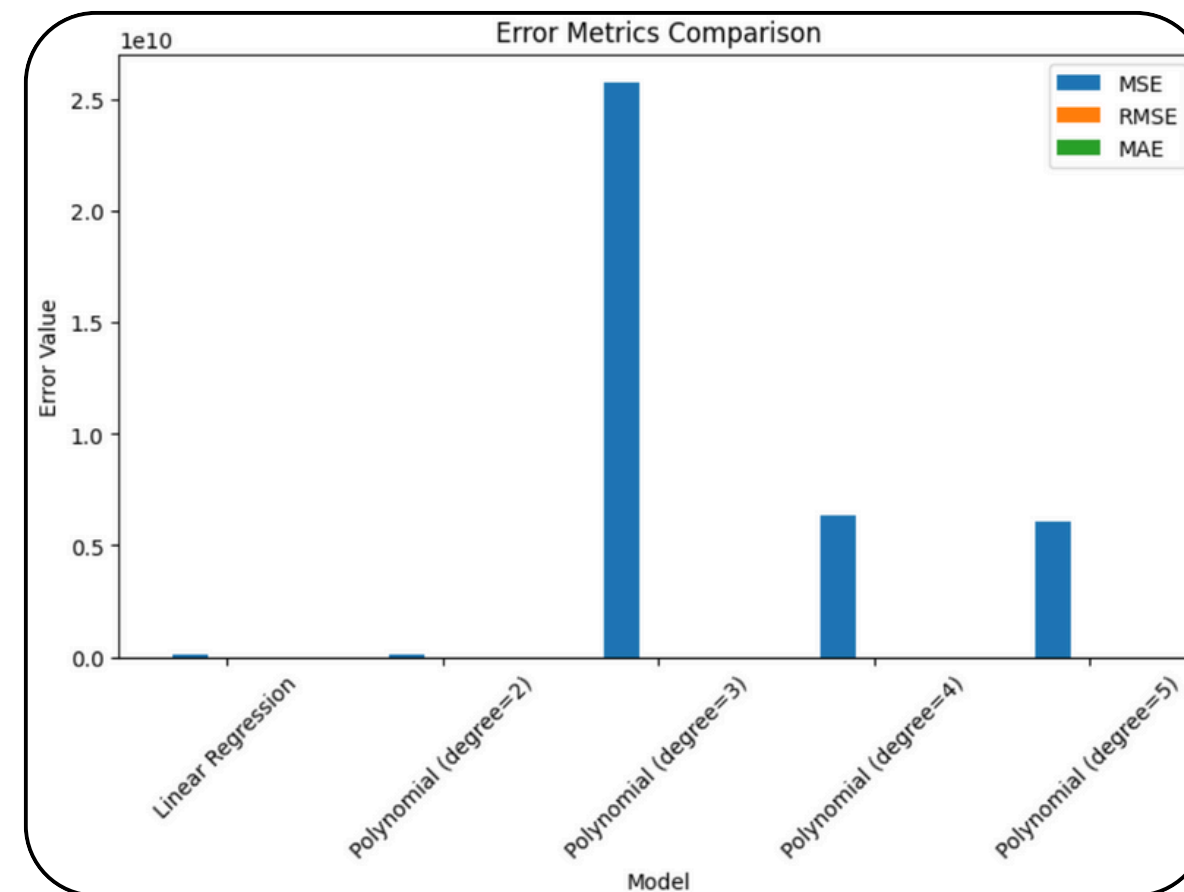
	Model	MSE	RMSE	MAE	MAPE	R2
	Linear Regression	8.482696e+07	9210.154995	7395.433532	0.089299	0.939711
	Polynomial (degree=2)	9.792206e+07	9895.557550	9139.212385	0.085270	0.930404
	Polynomial (degree=3)	2.573562e+10	160423.250579	80920.020787	1.392137	-17.291117
	Polynomial (degree=4)	6.337058e+09	79605.643319	41791.591546	0.524016	-3.503947
	Polynomial (degree=5)	6.063615e+09	77869.220074	44847.204752	0.680648	-3.309603

Regresi linear dan polinomial derajat 2 memiliki performa terbaik dengan  $R^2$  sekitar 0,93. Model polinomial dengan derajat  $\geq 3$  mengalami overfitting, ditunjukkan oleh  $R^2$  negatif dan peningkatan drastis pada MSE, RMSE, serta MAE.



# HASIL ANALISIS

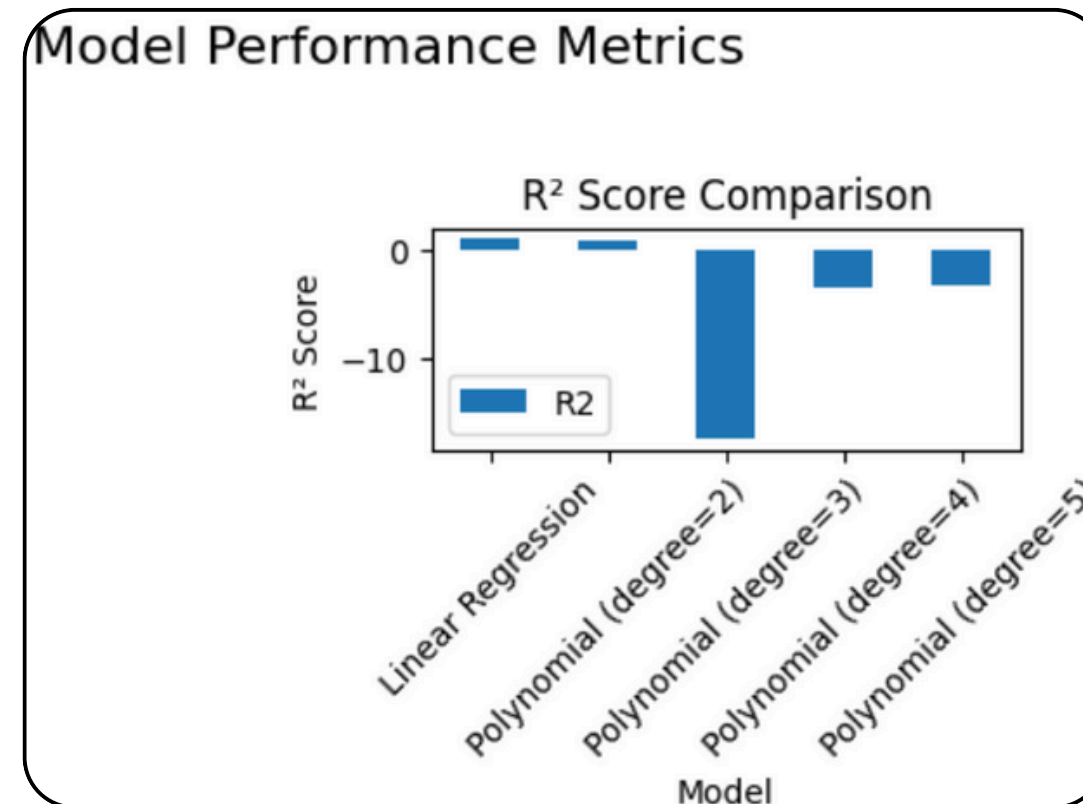
## Visualisasi Perbandingan Metrik Error



Grafik menunjukkan bahwa regresi polinomial derajat  $\geq 3$  mengalami peningkatan drastis pada MSE, RMSE, dan MAE, mengindikasikan overfitting. Sementara itu, regresi linear dan polinomial derajat 2 memiliki error yang jauh lebih kecil, menunjukkan performa yang lebih stabil dan akurat.

# HASIL ANALISIS

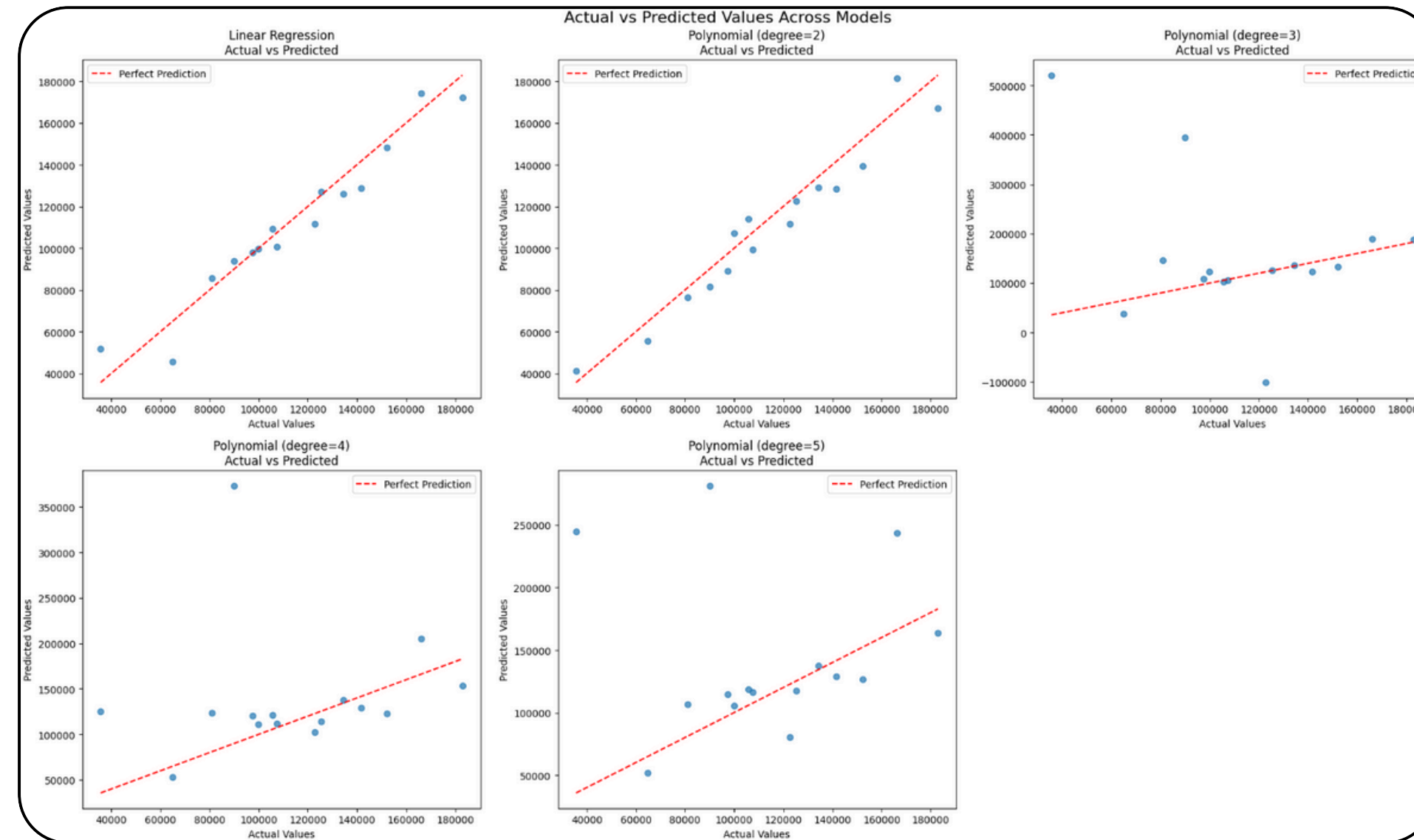
## Visualisasi R<sup>2</sup> Score



Grafik menunjukkan bahwa regresi linear dan polinomial derajat 2 memiliki nilai  $R^2$  mendekati 1, menandakan performa yang baik. Sebaliknya, model polinomial dengan derajat  $\geq 3$  memiliki nilai  $R^2$  negatif, menunjukkan overfitting parah dan kegagalan dalam generalisasi data.

# HASIL ANALISIS

## Visualisasi Actual vs Predicted



Grafik menunjukkan bahwa regresi linear dan polinomial derajat 2 memiliki prediksi yang paling mendekati garis ideal (garis merah putus-putus). Sebaliknya, model polinomial dengan derajat  $\geq 3$  menunjukkan penyebaran yang semakin jauh dari garis ideal, menandakan prediksi yang buruk dan overfitting yang signifikan.

# HASIL ANALISIS

## Identifikasi Model Terbaik

```
=== MODEL TERBAIK ===  
Model Terbaik: Linear Regression  
Alasan:  
- Memiliki  $R^2$  Tertinggi: 0.9397  
- MSE Terendah: 84826955.0353  
- RMSE Terendah: 9210.1550
```

Regresi linear dipilih sebagai model terbaik karena memiliki  $R^2$  tertinggi (0,9397), menunjukkan kecocokan terbaik dengan data. Selain itu, model ini memiliki nilai MSE dan RMSE terendah, menandakan kesalahan prediksi yang lebih kecil dibandingkan model lainnya.

# HASIL ANALISIS

## Analisis variabel yang paling berpengaruh

Variabel yang paling berpengaruh terhadap profit berdasarkan analisis korelasi:  
R&D Spend: 0.9729  
Marketing Spend: 0.7478  
Administration: 0.2007

Berdasarkan analisis korelasi diatas, variabel yang paling berpengaruh terhadap profit adalah R&D Spend dengan korelasi sebesar 0.9729, menunjukkan hubungan yang sangat kuat. Marketing Spend juga berpengaruh dengan korelasi 0.7478, namun tidak sekuat R&D Spend. Sementara itu, Administration memiliki pengaruh paling lemah dengan korelasi 0.2007, menandakan kontribusi yang kecil terhadap profit.





# KESIMPULAN



Hasil analisis menunjukkan bahwa Linear Regression adalah model terbaik dengan  $R^2$  Score 0.9397, menjelaskan 93.97% variasi data dengan akurat, sementara model polinomial mengalami overfitting. Oleh karena itu, regresi linear lebih efektif karena stabil dan tidak rentan terhadap kesalahan penyesuaian data. Untuk memastikan performa tetap optimal, disarankan melakukan validasi dengan data baru, menerapkan regularisasi jika ingin model lebih kompleks, serta menggunakan uji validasi silang.

**THANK  
YOU**

