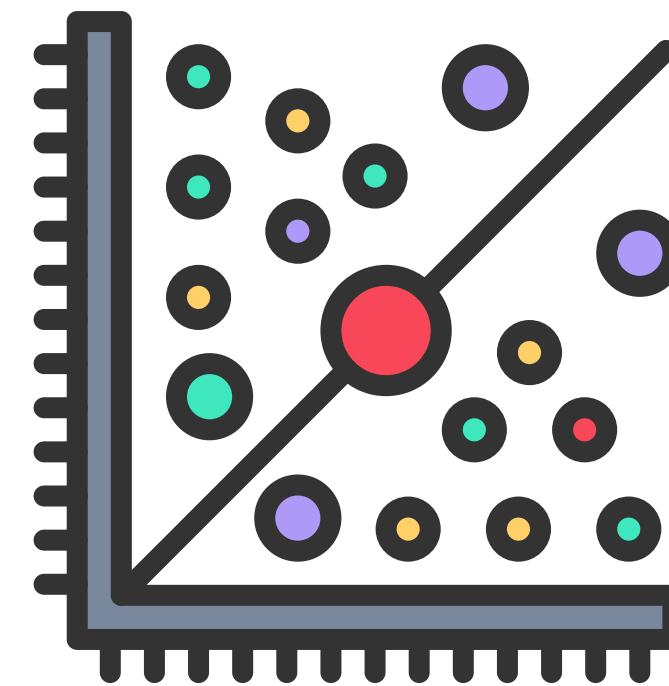


Project 03

Machine Learning : Quantity Sold E-Commerce Products Prediction

Outline :

- Monthly Sales Tren Analysis
- Best-Selling Products
- Quantity Sold E-Commerce Products Prediction

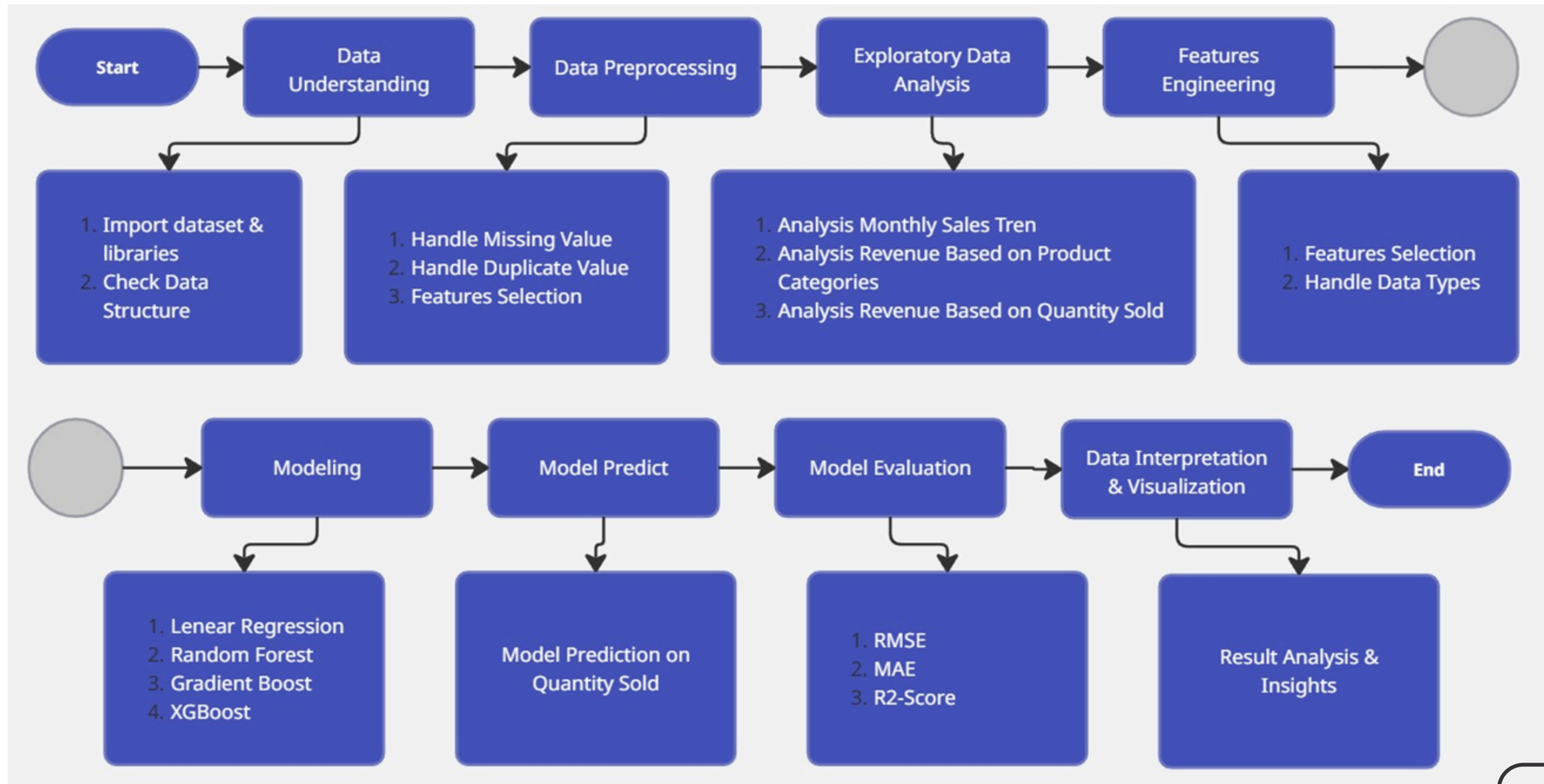


Overview

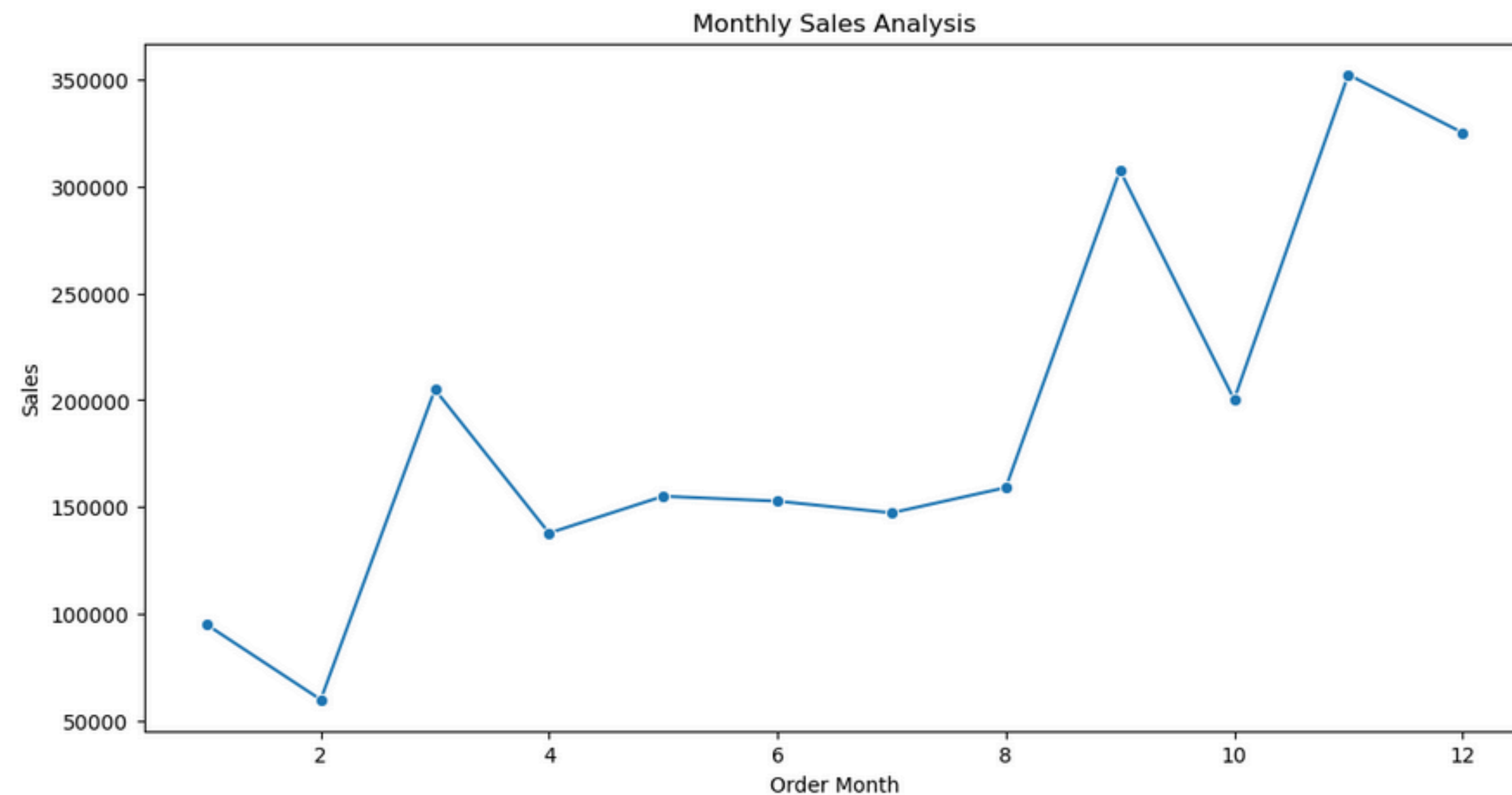
- Dalam e-commerce, memahami tren penjualan musiman dan produk terlaris sangat penting untuk strategi bisnis. Namun, perusahaan sering kesulitan dalam memprediksi penjualan, menentukan faktor yang memengaruhi penjualan (diskon, kategori produk, lokasi), serta mengoptimalkan stok dan pemasaran.
- Data penjualan yang kompleks membutuhkan analisis mendalam untuk mengidentifikasi pola pembelian, menentukan produk dengan permintaan tinggi, dan menyusun strategi harga yang lebih efektif. Oleh karena itu, proyek ini bertujuan untuk menganalisis tren penjualan, mengidentifikasi produk terlaris, serta membangun model prediksi yang akurat.
- Metode yang digunakan meliputi eksplorasi dan preprocessing data, feature engineering, serta penerapan model ML seperti Linear Regression, Random Forest, Gradient Boosting, dan XGBoost. Hasilnya divisualisasikan untuk memberikan insight yang dapat membantu optimasi strategi bisnis berbasis data.



Flow Chart



Monthly Sales Analysis



	Order month	Sales
0	1	94924.8356
1	2	59751.2514
2	3	205005.4888
3	4	137762.1286
4	5	155028.8117
5	6	152718.6793
6	7	147238.0970
7	8	159044.0630
8	9	307649.9457
9	10	200322.9847

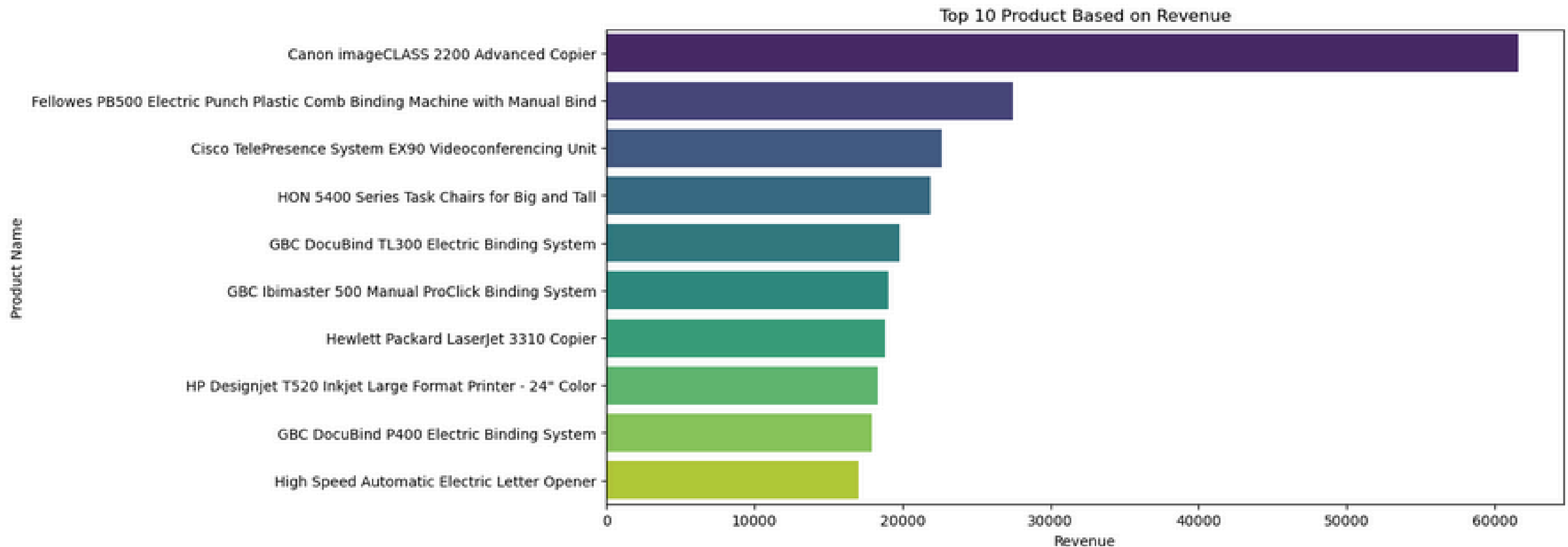
Monthly Sales Analysis

Insight Tren Penjualan :

- Lowest Sales: Februari (~59,751) → Kemungkinan efek pasca-liburan.
- Highest Sales: November (~307,649) → Bisa dipengaruhi oleh event belanja besar (Year End Sales).
- Lonjakan signifikan terjadi di Maret, lalu tren stabil hingga Agustus.
- Peningkatan mulai September, mencapai puncak di November, lalu sedikit turun di Desember.
- Perusahaan perlu strategi promosi lebih agresif di awal tahun dan optimalisasi stok di bulan puncak penjualan.

Best Selling Products

Top 10 Products Based on Revenue



Best Selling Products

	Product Name	Sales
404	Canon imageCLASS 2200 Advanced Copier	61599.824
650	Fellowes PB500 Electric Punch Plastic Comb Bin...	27453.384
444	Cisco TelePresence System EX90 Videoconferenci...	22638.480
786	HON 5400 Series Task Chairs for Big and Tall	21870.576
686	GBC DocuBind TL300 Electric Binding System	19823.479
688	GBC Ibimaster 500 Manual ProClick Binding System	19024.500
805	Hewlett Packard LaserJet 3310 Copier	18839.686
787	HP Designjet T520 Inkjet Large Format Printer ...	18374.895
683	GBC DocuBind P400 Electric Binding System	17965.068
813	High Speed Automatic Electric Letter Opener	17030.312

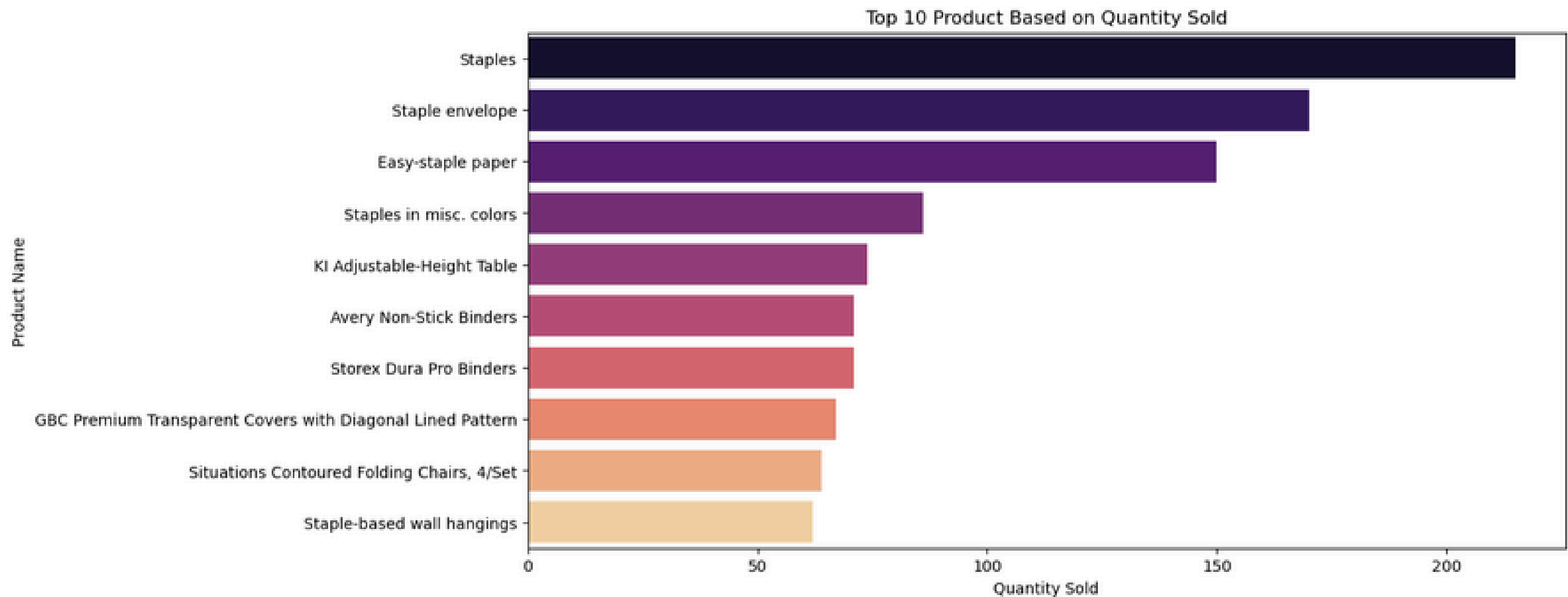


Insights :

- Highest Sales: Canon imageCLASS 2200 Advanced Copier (~61,599)
- Lowest Sales: High Speed Automatic Electric Letter Opener (~17,030).
- Mesin fotokopi dan sistem binding mendominasi penjualan yang merupakan kebutuhan tinggi untuk peralatan kantor.

Best Selling Products

Top 10 Products Based on Quantity Sold



Best Selling Products

	Product Name	Quantity
1499	Staples	215
1493	Staple envelope	170
537	Easy-staple paper	150
1500	Staples in misc. colors	86
942	KI Adjustable-Height Table	74
259	Avery Non-Stick Binders	71
1511	Storex Dura Pro Binders	71
699	GBC Premium Transparent Covers with Diagonal L...	67
1450	Situations Contoured Folding Chairs, 4/Set	64
1497	Staple-based wall hangings	62

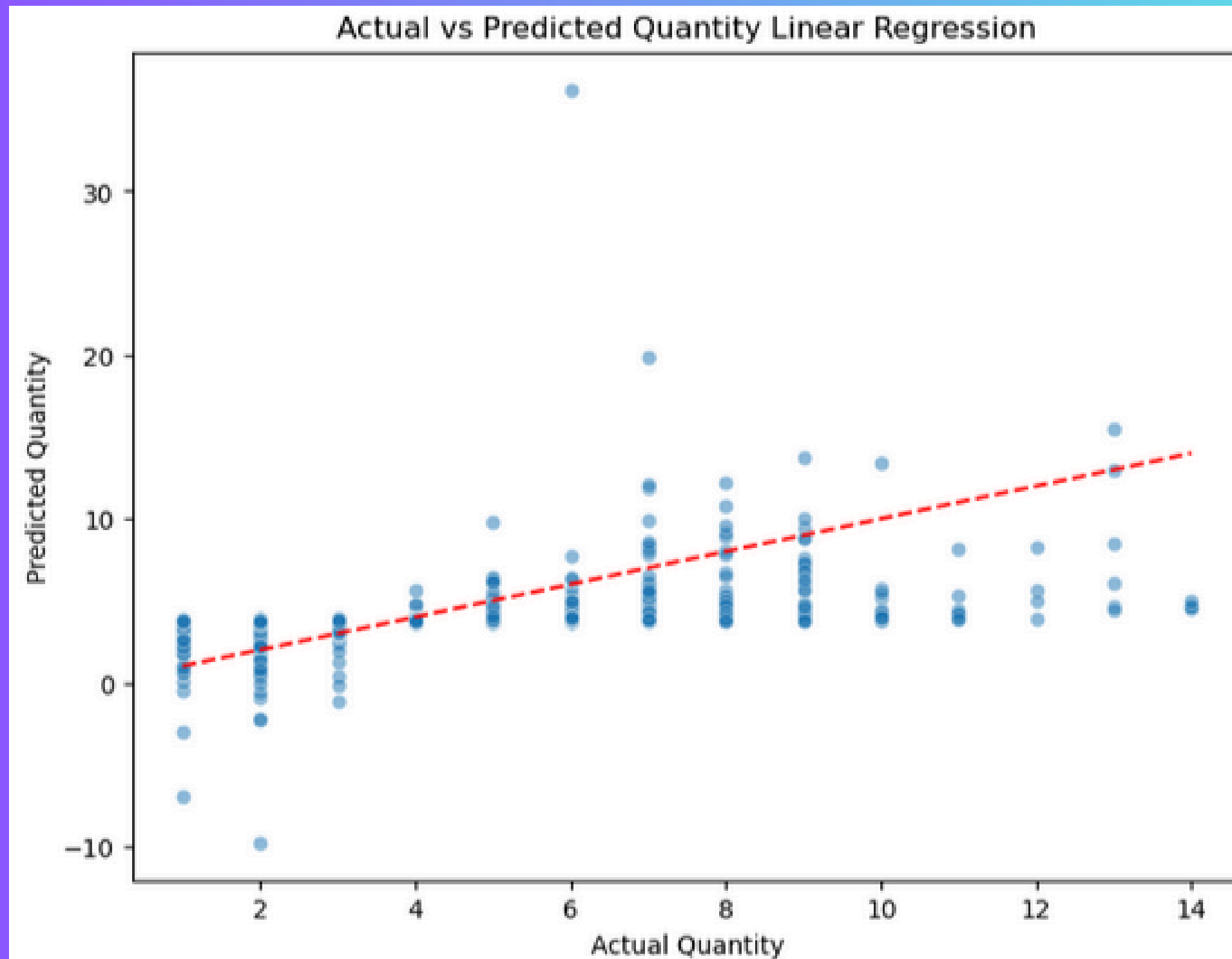
Insights :

- Highest Quantity Sold: Staples (215 unit)
- Lowest Quantity Sold: Staple-based wall hangings (62 unit)
- Produk dengan jumlah terjual tinggi umumnya merupakan kebutuhan kantor dasar (staples, kertas, amplop).

Feature vs Quantity Sold

Feature	Coefficient	Interpretasi
Discount	0.060222	Diskon memiliki pengaruh positif paling besar terhadap penjualan. Semakin besar diskon, semakin banyak barang terjual.
Order Month	0.012535	Bulan pemesanan juga berpengaruh, artinya ada tren musiman dalam penjualan.
Order Year	0.009257	Tahun pemesanan sedikit berpengaruh, bisa jadi ada pertumbuhan atau perubahan tren dari tahun ke tahun.
Sales	0.003850	Total penjualan sedikit mempengaruhi jumlah barang yang terjual, tapi tidak terlalu signifikan.
Category	-0.003340	Kategori produk punya pengaruh negatif, artinya kategori tertentu bisa mengurangi jumlah barang terjual.
Price	-0.014489	Harga memiliki pengaruh negatif terbesar, artinya semakin mahal barang, semakin sedikit jumlahnya yang terjual (logis karena efek permintaan vs harga).

Linear Regression

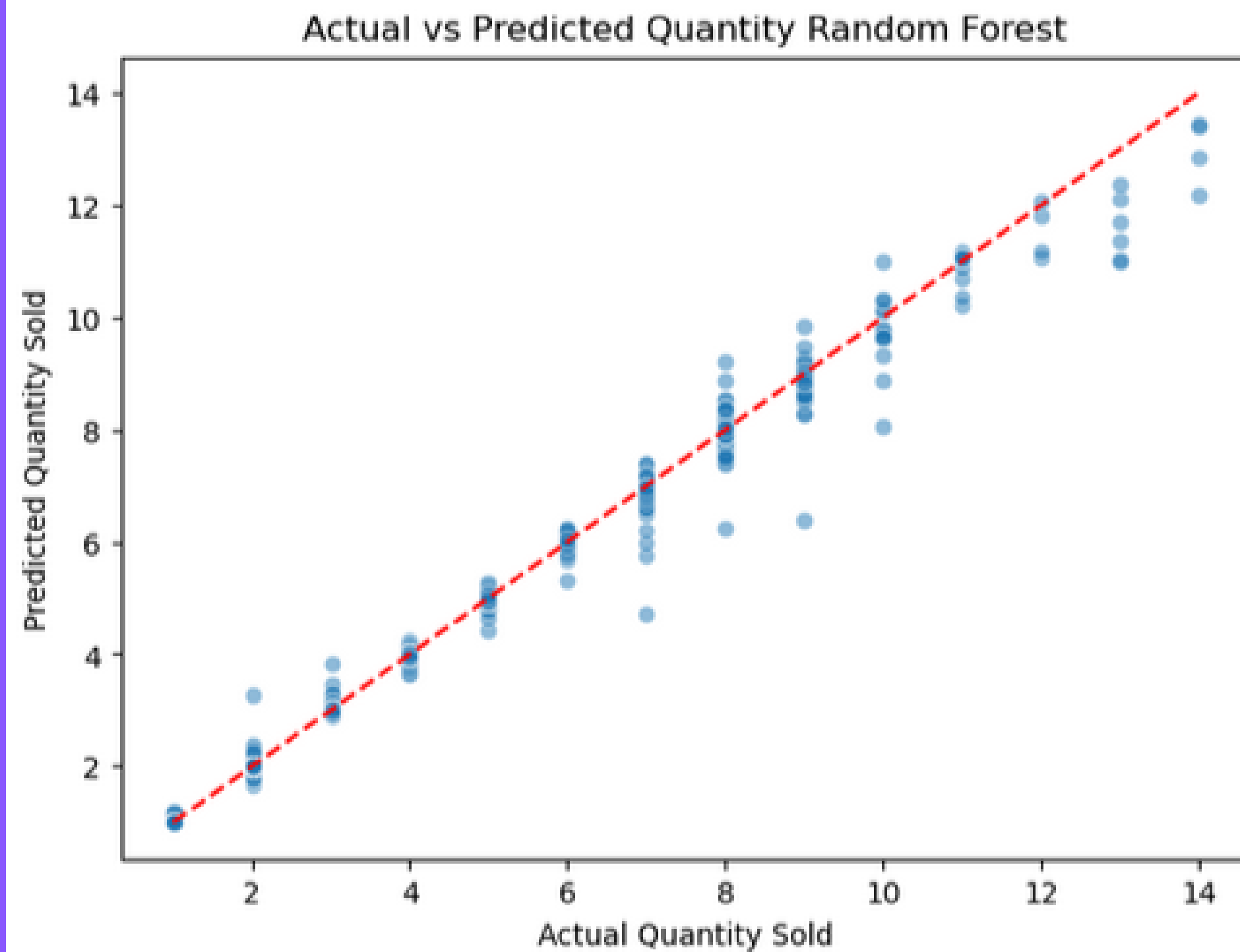


Insights :

- Sebaran titik yang cukup jauh dari garis regresi menunjukkan error yang cukup besar.
- Model ini cenderung kurang akurat untuk memprediksi kuantitas dalam nilai yang lebih tinggi.



Random Forest



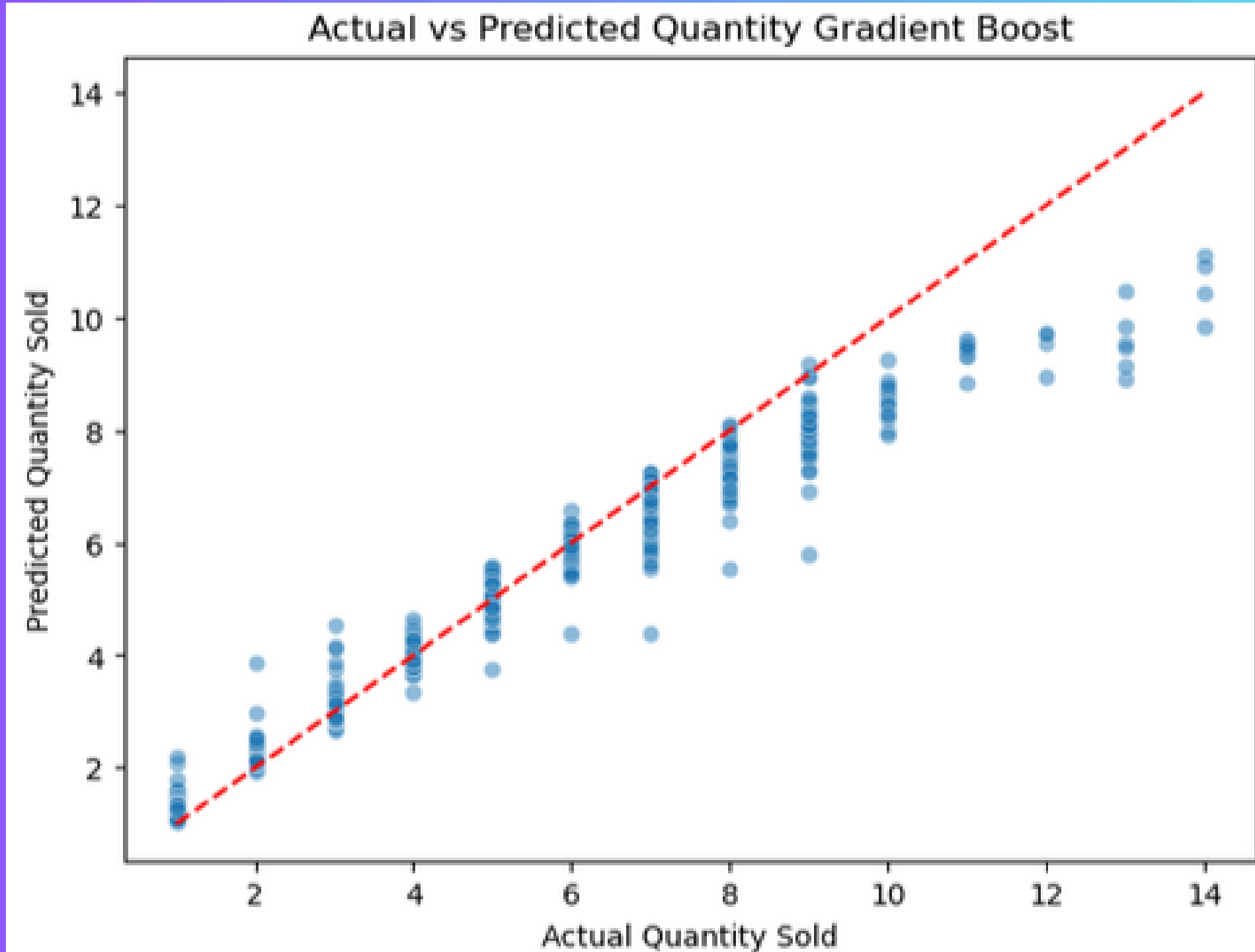
Insights :

- Sebaran titik lebih rapat di sekitar garis dibandingkan model Linear Regression, menunjukkan error lebih kecil.
- Model ini lebih baik dalam menangkap pola dibandingkan model sebelumnya, membuatnya lebih cocok untuk prediksi kuantitas penjualan.



Gradient Boost

Supervised Learning

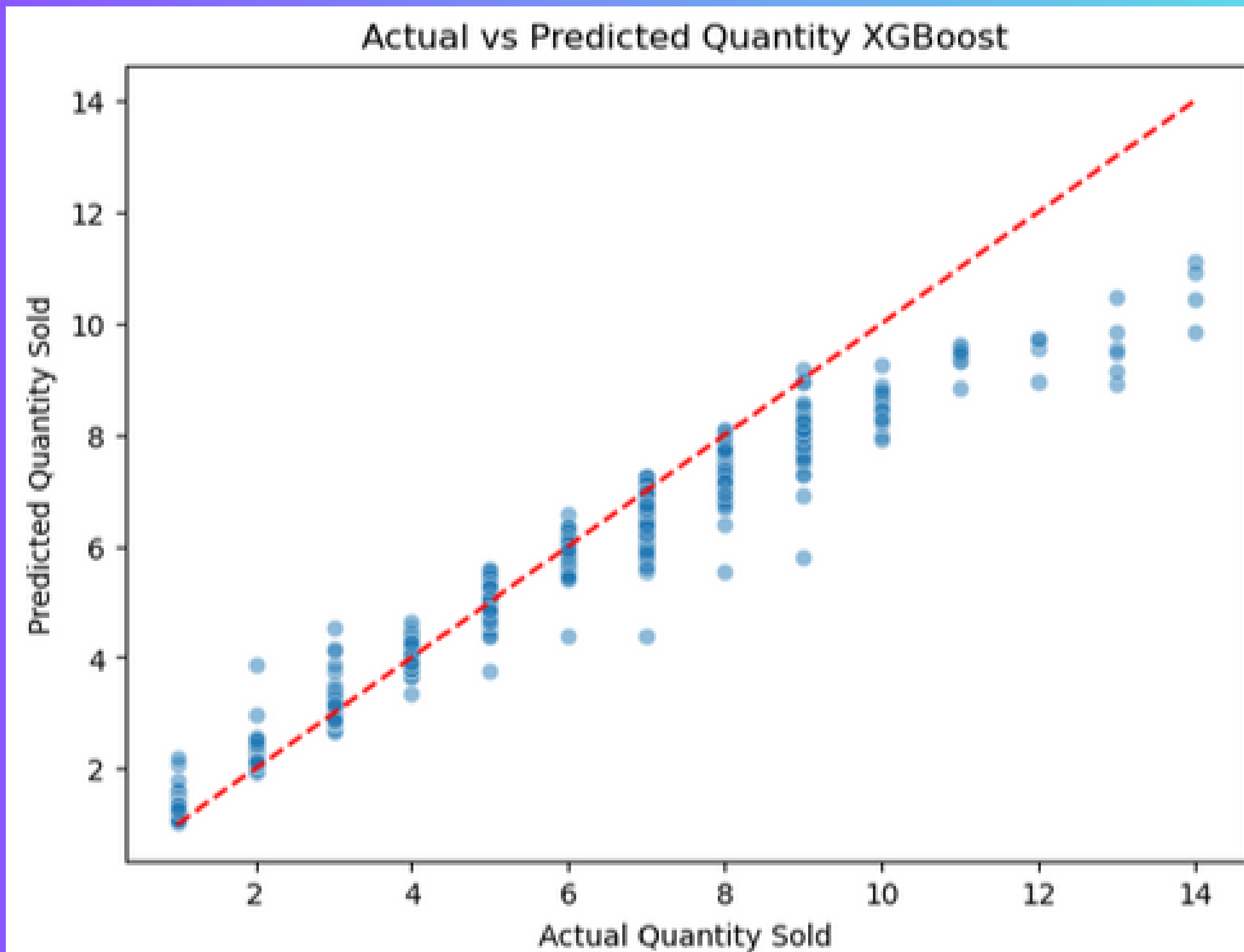


Insights :

- Model Gradient Boost menunjukkan hubungan yang cukup baik antara kuantitas aktual dan prediksi.
- Dibandingkan dengan Random Forest, model ini cenderung lebih menyebar, menunjukkan bahwa ada sedikit lebih banyak error.



Extreme Gradient Boost



Insights :

- Model XGBoost menunjukkan pola prediksi yang sangat baik dengan sebaran titik yang mengikuti garis ideal (garis merah).
- Hasil prediksi lebih akurat dibandingkan Linear Regression dan Gradient Boost, dengan lebih sedikit penyimpangan.
- Secara keseluruhan, XGBoost merupakan salah satu model terbaik untuk prediksi kuantitas penjualan dalam dataset ini.



Conclusion

Model	MAE	RMSE	R ² Score
Linear Regression	1.50	2.10	0.13
Random Forest	0.05	0.18	0.99
Gradient Boost	0.27	0.47	0.96
XGBoost	0.10	0.21	0.99



Kesimpulan:

XGBoost dan Random Forest menjadi model terbaik untuk prediksi jumlah penjualan karena memberikan hasil yang sangat akurat dengan error yang kecil!

Analisis Hasil Model:

- Random Forest dan XGBoost memiliki performa terbaik, dengan MAE & RMSE yang rendah serta R² mendekati 1.
- Linear Regression kurang cocok, karena R² hanya 0.13, menunjukkan model ini tidak bisa menangkap pola data dengan baik.
- Gradient Boosting cukup baik, tapi masih kalah dengan Random Forest & XGBoost dalam hal error dan akurasi.



[Link Github Project 03 : E-Commerce Quantity Sold Prediction](#)