

## Penjelasan Aplikasi:

Pada penelitian ini, dilakukan analisis terhadap data penjualan dan pengembalian barang menggunakan algoritma Random Forest. Data yang digunakan adalah data penjualan (SALES) dan pengembalian (RETUR) yang diambil dari sebuah CSV file. Langkah pertama dalam analisis ini adalah mengonversi kolom 'TGL\_INVC' menjadi format tanggal yang dapat diolah dan mengekstraksi komponen tanggal seperti hari, bulan, dan tahun. Data kemudian difilter berdasarkan jenis transaksi (SALES atau RETUR) yang dipilih melalui sidebar pada aplikasi Streamlit.

Setelah memisahkan data penjualan dan pengembalian, dilakukan pengelompokan dan penghitungan jumlah kuantitas barang (QTYSALES) untuk setiap produk. Pada data penjualan, produk dengan jumlah penjualan tertinggi diidentifikasi, sedangkan pada data pengembalian, produk dengan jumlah pengembalian tertinggi diidentifikasi dan ditampilkan dengan tanda minus untuk membedakan dari penjualan.

Data selanjutnya diolah untuk mendapatkan informasi tanggal dan bulan dengan penjualan atau pengembalian terbanyak. Informasi ini kemudian ditampilkan dalam aplikasi Streamlit untuk memberikan wawasan tambahan mengenai tren penjualan dan pengembalian berdasarkan waktu.

Untuk membangun model prediksi, digunakan algoritma Random Forest. Data yang digunakan untuk pelatihan dan pengujian model terdiri dari fitur-fitur seperti hari, bulan, tahun, nama barang, dan nama outlet. Sebelum dilakukan pelatihan model, kolom 'NAMABARANG' dan 'NAMAOUTLET' dikonversi menjadi format numerik menggunakan LabelEncoder. Model kemudian dilatih dengan membagi data menjadi data pelatihan dan data uji dengan rasio 80:20.

Hasil dari model Random Forest menunjukkan tingkat akurasi sebesar 93.91% untuk prediksi data penjualan dan 90.48% untuk data pengembalian. Selain akurasi, metrik evaluasi lain seperti recall, precision, dan F1 score juga dihitung. Recall untuk penjualan dan pengembalian ditampilkan dalam bentuk persentase, memberikan gambaran seberapa baik model dalam mendeteksi transaksi penjualan dan pengembalian yang sebenarnya terjadi.

Diagram batang menggunakan Plotly Express dibuat untuk menggambarkan jumlah barang yang terjual atau dikembalikan. Selain itu, grafik garis juga digunakan untuk menampilkan metrik evaluasi model dalam bentuk persentase, memberikan visualisasi yang jelas mengenai kinerja model. Hasil analisis menunjukkan bahwa model memiliki performa yang baik dalam memprediksi transaksi penjualan dan pengembalian, dengan semua metrik utama menunjukkan nilai yang memuaskan.

Seluruh proses analisis dan hasilnya disajikan dalam aplikasi Streamlit, memungkinkan interaksi langsung dengan data dan visualisasi yang dinamis. Penelitian ini memberikan wawasan penting mengenai pola penjualan dan pengembalian barang serta menunjukkan bagaimana algoritma Random Forest dapat digunakan secara efektif untuk analisis dan prediksi dalam konteks data penjualan.

### Penjelasan Activity Diagram:

1. Mulai: Proses dimulai dari tahap awal.
2. Baca data CSV: Data diambil dari file CSV yang berisi informasi penjualan dan retur.
3. Konversi TGL\_INVC ke format tanggal: Mengubah format tanggal pada kolom 'TGL\_INVC' agar dapat diolah lebih lanjut.
4. Ekstraksi komponen tanggal dari TGL\_INVC: Memisahkan tanggal, bulan, dan tahun dari kolom 'TGL\_INVC'.
5. Hapus kolom TGL\_INVC: Menghapus kolom 'TGL\_INVC' setelah ekstraksi tanggal selesai.
6. Sidebar: Pilih antara SALES atau RETUR: Pengguna memilih untuk melihat data penjualan (SALES) atau data retur (RETUR) melalui menu sidebar.
7. Filter data berdasarkan pilihan: Data difilter berdasarkan pilihan pengguna (SALES atau RETUR).
8. Hitung jumlah QTYSALES atau QTYRETUR untuk setiap barang: Menghitung jumlah barang yang terjual (SALES) atau dikembalikan (RETUR) untuk setiap produk.
9. Tampilkan produk dengan jumlah SALES terbanyak dan RETUR terbanyak: Menampilkan produk dengan jumlah penjualan terbanyak (SALES) atau jumlah retur terbanyak (RETUR).
10. Split data untuk melatih model Random Forest: Data dibagi menjadi data pelatihan dan data uji untuk melatih model Random Forest.
11. Latih model Random Forest: Model Random Forest dilatih menggunakan data pelatihan.
12. Lakukan prediksi pada data uji: Model yang telah dilatih digunakan untuk melakukan prediksi menggunakan data uji.
13. Mengukur akurasi model: Menghitung akurasi dari model yang telah dilatih menggunakan data uji.
14. Menghitung metrik evaluasi: Recall, Presisi, F1 Score: Menghitung metrik evaluasi lainnya seperti recall, presisi, dan F1 score untuk mengevaluasi performa model.
15. Tampilkan hasil berdasarkan pilihan: Menampilkan hasil analisis berdasarkan pilihan pengguna (SALES atau RETUR).
16. Tampilkan data tabel SALES dan RETUR: Menampilkan data dalam bentuk tabel untuk SALES atau RETUR sesuai dengan pilihan pengguna.
17. Buat diagram batang dengan Plotly: Membuat diagram batang untuk visualisasi data penjualan atau retur menggunakan Plotly.
18. Evaluasi akurasi model: Melakukan evaluasi akhir terhadap akurasi dan performa model yang telah dilatih.
19. Selesai: Proses analisis data penjualan dan retur selesai.