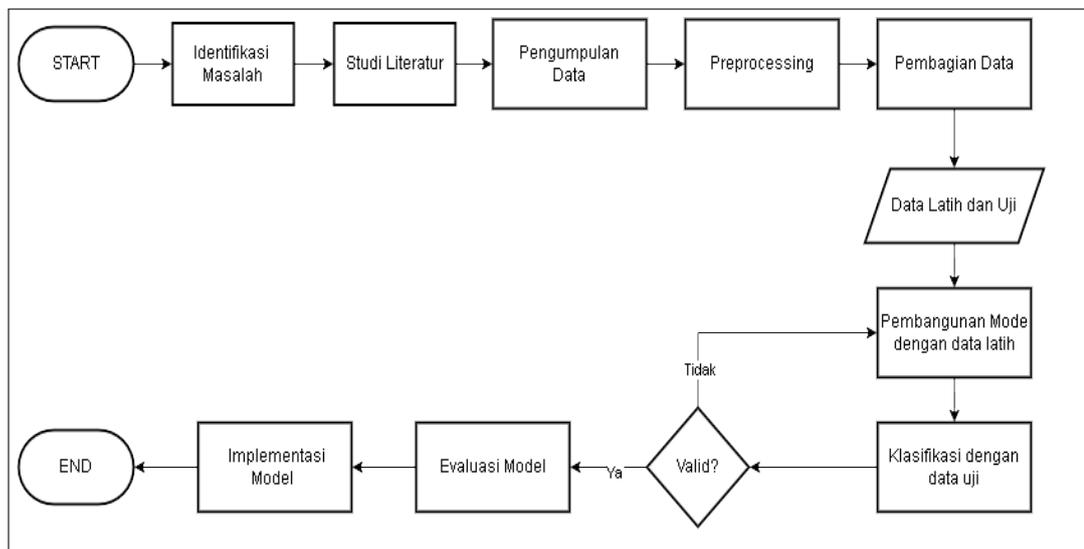


BAB 3 METODOLOGI PENELITIAN

Pada penelitian ini, terdapat beberapa tahapan metodologi yang dilalui untuk melaksanakan penelitian. Tahapan-tahapan tersebut meliputi identifikasi masalah, studi literatur, pengumpulan data, *preprocessing*, pembagian data, klasifikasi, pembangunan serta pelatihan model, menganalisis *output*, dan evaluasi, seperti yang digambarkan ada Gambar 3.1.



Gambar 3.1. *Flowchart* metodologi penelitian

Setiap tahapan dalam metodologi ini dirancang untuk memastikan bahwa penelitian dilakukan secara sistematis dan hasil yang diperoleh dapat diandalkan.

3.0.1 Identifikasi Masalah

Pendidikan Montessori yang berkembang di Indonesia meningkatkan permintaan terhadap alat peraga Montessori. CV Kamoemontessori menghadapi tantangan dalam memprediksi permintaan pasar yang fluktuatif, mengakibatkan masalah manajemen stok. Prediksi yang tidak akurat berpotensi mengakibatkan kelebihan atau kekurangan produk, merugikan perusahaan dan menurunkan kepuasan pelanggan. Identifikasi masalah ini menetapkan dasar untuk penelitian ini dengan tujuan mengembangkan sistem prediksi penjualan yang akurat.

3.0.2 Studi Literatur

Tahap studi literatur dilakukan untuk memahami informasi dari beberapa sumber seperti jurnal, artikel, atau buku yang berhubungan dengan topik prediksi penjualan dan pengelolaan stok menggunakan algoritma *random forest*. Pada tahap ini, akan dilakukan aktivitas seperti menelusuri, membaca, memahami, dan mendemonstrasikan teori-teori dari algoritma yang akan digunakan serta penelitian terkait yang sudah ada sebelumnya.

3.0.3 Pengumpulan Data

Tahap pengumpulan data dilakukan secara berkelanjutan untuk mengakomodasi sifat dinamis dari data penjualan. Dataset mencakup data historis penjualan CV Kamoemontessori dari bulan Januari 2022 hingga Desember 2023 serta data terkini yang akan terus diperbarui secara real-time atau pada interval yang sudah ditentukan. Data ini mencakup informasi penting seperti nomor produk, nama produk, jumlah stok awal, stok terjual per bulan, dan data transaksi lainnya yang relevan.

Dataset yang dikumpulkan akan digunakan untuk membangun model prediksi yang dapat menganalisis dan mempelajari tren penjualan terkini. Data penjualan yang dikumpulkan ini akan menjadi dasar dalam proses pelatihan dan pengujian model prediksi untuk memastikan bahwa model mampu memberikan hasil prediksi yang akurat dan dapat diandalkan.

Proses pengumpulan data ini penting untuk memastikan bahwa model prediksi memiliki data yang cukup dan beragam untuk melakukan analisis yang tepat. Dengan menggunakan data historis dan terkini, model dapat mengidentifikasi pola dan tren yang mungkin tidak terlihat hanya dari data dalam periode waktu yang singkat. Ini memungkinkan model untuk memberikan prediksi yang lebih akurat mengenai penjualan produk di masa depan.

3.0.4 Preprocessing

Tahap preprocessing merupakan tahap untuk melakukan pengecekan kembali dataset yang didapat dari tahap sebelumnya. Tahap *preprocessing* dilakukan secara berkelanjutan dan otomatis untuk menjamin integritas dan relevansi informasi yang digunakan oleh model prediksi. Pada tahap ini akan dilakukan penghapusan data yang tidak relevan, pengisian nilai yang hilang,

normalisasi skala data untuk memastikan konsistensi, dan encoding variabel kategorikal. Pendekatan ini penting karena sifat data yang dinamis dan bertambah, yang membutuhkan model prediktif untuk beradaptasi dengan tren penjualan terkini.

Berbeda dengan penelitian prediksi banjir, yang menggunakan *set* data historis dari periode 2015 hingga 2018. Dalam penelitian tersebut, *preprocessing* dilakukan sekali di awal studi. Metode *preprocessing* mencakup *Exploratory Data Analysis* (EDA) yang mendalam untuk memahami distribusi dan karakteristik data, seperti intensitas curah hujan dan tinggi muka air sungai. Proses ini juga melibatkan transformasi label kategori dari bentuk *string* ke numerik dan pembagian data menjadi *set* pelatihan dan pengujian dengan rasio yang telah ditetapkan. Dalam konteks tersebut, *preprocessing* lebih bersifat statis dan tidak memerlukan pembaruan berkelanjutan karena sifat data yang tidak berubah selama periode penelitian.

Perbedaan utama dalam tahapan *preprocessing* antara kedua penelitian ini terletak pada dinamika dan volume data. Penelitian ini harus mengakomodasi perubahan dan penambahan data secara real-time, yang memerlukan otomatisasi dan fleksibilitas dalam *preprocessing*, sementara penelitian banjir fokus pada analisis data historis yang tetap, di mana proses *preprocessing* yang lebih manual dan analitis menjadi lebih relevan.

3.0.5 Pembagian Data

Data dibagi menjadi dua set, yaitu *data train* dan *data test*. Pembagian data dilakukan dengan mempertimbangkan keseimbangan dan representasi data yang akan dipakai dalam pembangunan model.

3.0.6 Pembangunan Model

Setelah tahap *preprocessing* selesai dilakukan, langkah selanjutnya adalah pembangunan model. Tahap ini terbagi menjadi dua bagian, yaitu pembuatan model dan pengujian model. Pada penelitian ini, pembuatan model dilakukan menggunakan algoritma Random Forest yang kemudian dilatih dengan dataset yang telah disiapkan.

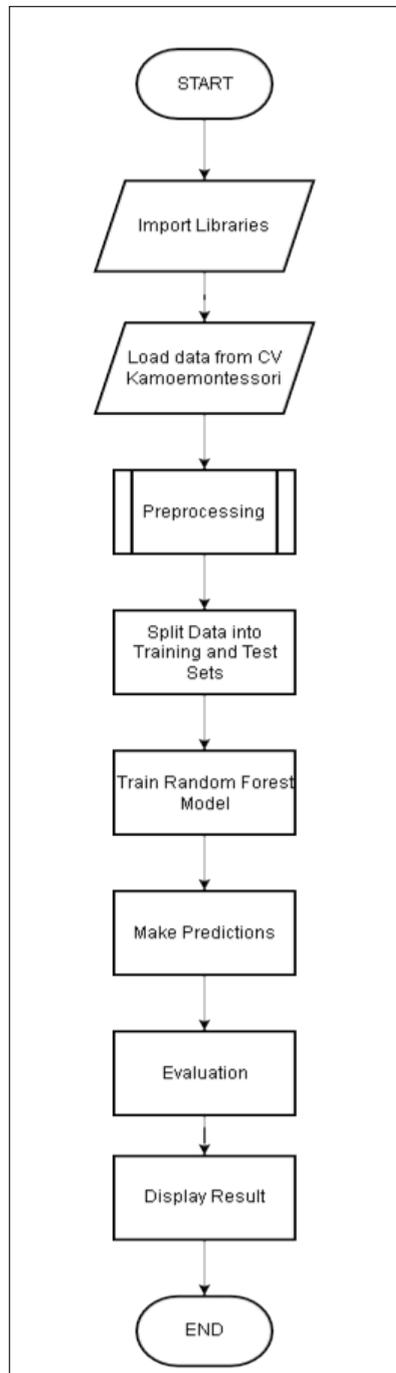
3.0.7 Evaluasi Model

Tahap evaluasi model bertujuan untuk mengukur apakah hasil yang diperoleh sesuai dengan yang diharapkan dan seberapa baik model tersebut dalam memprediksi penjualan alat peraga Montessori. Evaluasi ini sangat penting untuk memastikan akurasi model sebelum digunakan lebih lanjut dalam aplikasi praktis.

3.0.8 Flowchart

Flowchart adalah diagram yang menggambarkan langkah-langkah dan keputusan dalam proses pengembangan sistem. Pada penelitian ini, *flowchart* digunakan untuk menggambarkan alur kerja dalam pembuatan sistem klasifikasi serta tahap-tahap penggunaan dalam aplikasi. Diagram ini memudahkan pemahaman tentang proses yang dilakukan dan memastikan setiap langkah telah direncanakan dengan baik.

A Flowchart Perancangan Sistem Prediksi



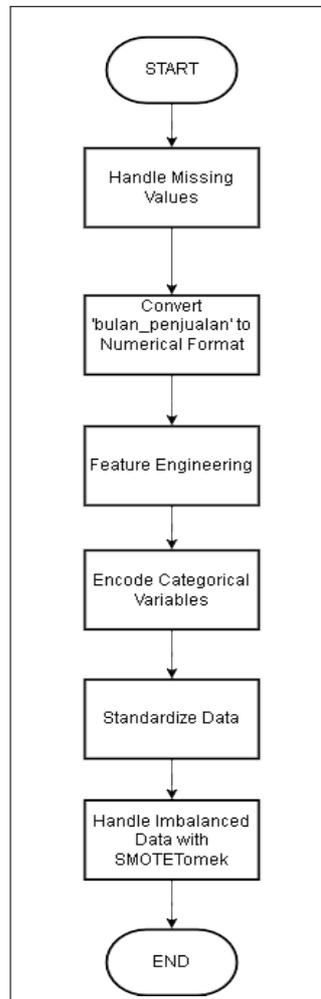
Gambar 3.2. *Flowchart* perancangan sistem prediksi

Pada Gambar 3.2 *flowchart* yang baru menggambarkan langkah-langkah utama dalam pengembangan sistem prediksi penjualan alat peraga Montessori.

Langkah pertama adalah mengimpor *libraries* yang dibutuhkan untuk pemrosesan data, pengambilan data, pembangunan model, dan evaluasi model. Libraries yang digunakan antara lain pandas, numpy, gspread, oauth2client, sklearn, imblearn, matplotlib, dan seaborn. Libraries ini menyediakan alat dan fungsi penting yang digunakan dalam berbagai tahapan analisis data dan pembelajaran mesin. Setelah mengimpor *libraries*, langkah selanjutnya adalah memuat data penjualan dari CV Kamoemontessori. Data ini diambil dari Google Sheets menggunakan kredensial layanan akun untuk autentikasi dan akses. Data tersebut kemudian diubah menjadi DataFrame pandas untuk memudahkan pemrosesan lebih lanjut. Tahap berikutnya adalah *preprocessing*, yang mencakup pembersihan dan persiapan data. Pada tahap ini, atribut yang tidak diperlukan dihapus, nilai yang hilang ditangani, data dikategorikan, dan tipe data diubah sesuai kebutuhan. Setelah *preprocessing*, data dibagi menjadi *set* pelatihan dan *set* pengujian untuk memisahkan data yang akan digunakan untuk melatih model dan data yang akan digunakan untuk menguji model.

Setelah pembagian data, langkah berikutnya adalah melatih model menggunakan algoritma *Random Forest*. Model dilatih menggunakan data pelatihan dan hasilnya diuji menggunakan data pengujian. Prediksi dibuat berdasarkan model yang telah dilatih. Selanjutnya adalah evaluasi, dimana hasil prediksi dievaluasi untuk mengukur akurasi model. Evaluasi ini melibatkan penggunaan metrik seperti *classification report* dan *confusion matrix* untuk memahami kinerja model dalam memprediksi peningkatan penjualan. Terakhir, hasil prediksi ditampilkan dalam bentuk tabel yang menunjukkan produk yang diprediksi mengalami peningkatan penjualan. Proses ini berakhir dengan evaluasi akhir dan interpretasi hasil prediksi untuk pengambilan keputusan lebih lanjut.

B Flowchart Preprocessing

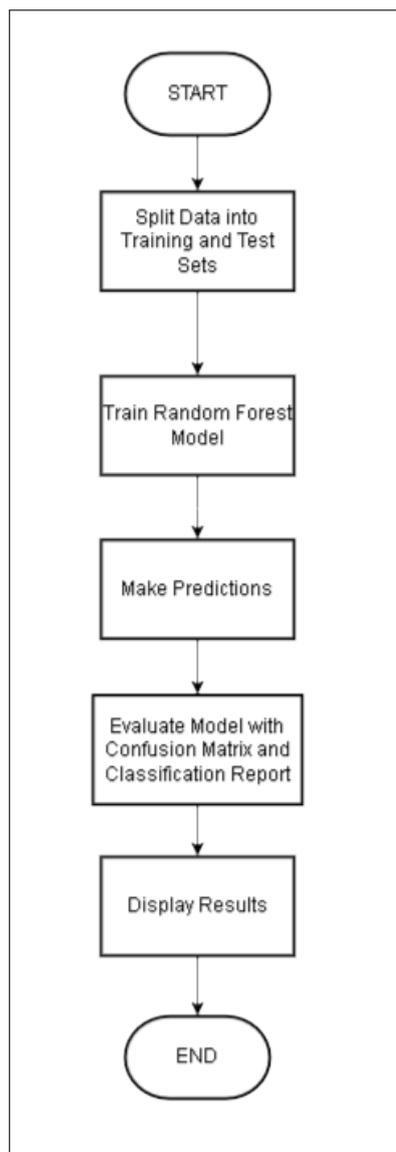


Gambar 3.3. *Flowchart preprocessing*

Pada Gambar 3.3 terdapat beberapa langkah yang dilakukan untuk mempersiapkan data sebelum dilakukan pemodelan. Langkah-langkah tersebut meliputi menangani nilai yang hilang pada dataset, mengonversi kolom 'bulan_penjualan' dari format teks ke format numerik, dan melakukan feature engineering. Langkah feature engineering ini mencakup pembuatan fitur baru seperti 'sales_diff', 'sales_rolling_mean', 'sales_rolling_std', dan 'yearly_sales' untuk meningkatkan kinerja model. Setelah itu, variabel kategorikal seperti 'kode_produk' dan 'nama_produk' diubah menjadi variabel dummy menggunakan teknik one-hot encoding. Data kemudian distandarisasi untuk memastikan setiap fitur memiliki skala yang sama. Akhirnya, data yang tidak seimbang ditangani

menggunakan metode SMOTETomek untuk mengatasi masalah ketidakseimbangan kelas. Seluruh proses ini bertujuan untuk membersihkan dan mempersiapkan data sehingga siap untuk digunakan dalam pemodelan *machine learning*.

C Flowchart Classification



Gambar 3.4. *Flowchart classification*

Pada Gambar 3.4 menggambarkan langkah-langkah yang dilakukan untuk membangun dan mengevaluasi model klasifikasi menggunakan algoritma *Random Forest*. Langkah pertama adalah membagi data yang telah diproses ke dalam set

pelatihan dan set pengujian. Model *Random Forest* kemudian dilatih menggunakan data pelatihan. Setelah model dilatih, prediksi dilakukan pada data pengujian. Langkah selanjutnya adalah mengevaluasi kinerja model menggunakan *confusion matrix* dan *classification report* untuk mengukur *accuracy*, *precision*, *recall*, dan *f1-score*. Akhirnya, hasil evaluasi dan prediksi ditampilkan untuk dianalisis lebih lanjut. Proses ini membantu dalam menentukan efektivitas model dalam memprediksi peningkatan penjualan produk.

3.0.9 Penulisan Laporan

Tahapan ini akan dilakukan penyusunan laporan penelitian tentang sistem prediksi penjualan alat peraga montessori anak menggunakan algoritma *Random Forest* (studi kasus: CV Kamoemontessori)