**Memprediksi Kualitas Gerakan Renang Menggunakan Data Akselerometer**

**Pendahuluan**

Proyek ini bertujuan untuk memprediksi kualitas gerakan renang menggunakan data akselerometer yang dikumpulkan dari berbagai bagian tubuh perenang (misalnya, pergelangan tangan, pinggul, kaki) selama sesi latihan. Tujuannya adalah untuk mengklasifikasikan kualitas gerakan ke dalam beberapa kategori yang telah ditentukan (diwakili oleh variabel "classe"), yang mungkin mencerminkan teknik yang benar, sedikit penyimpangan, atau kesalahan signifikan. Klasifikasi ini akan dicapai dengan menerapkan teknik pembelajaran mesin pada data akselerometer.

**Pengumpulan dan Persiapan Data**

Kumpulan data terdiri dari pengukuran sensor akselerometer yang direkam saat perenang melakukan berbagai gaya renang (misalnya, gaya bebas, gaya dada, gaya punggung, gaya kupu-kupu) dengan kualitas gerakan yang berbeda-beda. Data pelatihan dimuat dan dipersiapkan melalui serangkaian langkah:

* **Penghapusan Kolom dengan Nilai Hilang Berlebihan:** Kolom-kolom yang memiliki proporsi nilai yang hilang terlalu tinggi akan diidentifikasi dan dihapus karena kemungkinan tidak memberikan informasi yang andal untuk pelatihan model. Ambang batas proporsi nilai hilang untuk penghapusan akan ditentukan (misalnya, di atas 50%).
* **Penghapusan Pengidentifikasi yang Tidak Perlu:** Kolom-kolom seperti ID pengguna, timestamp mentah, atau pengidentifikasi lain yang tidak secara langsung terkait dengan data sensor gerakan akan dihapus untuk menghindari pengenalan pola yang tidak relevan oleh model.
* **Penghapusan Fitur dengan Varians Rendah atau Nol:** Fitur-fitur yang memiliki varians yang sangat rendah atau nol tidak memberikan informasi diskriminatif dan dapat dihapus untuk menyederhanakan model dan mengurangi potensi *noise*.
* **Penanganan Nilai Hilang yang Tersisa:** Untuk nilai hilang yang tidak dihapus melalui langkah sebelumnya, strategi imputasi yang sesuai akan diterapkan. Metode seperti mengisi dengan nilai mean, median, atau modus dari kolom yang bersangkutan dapat digunakan, tergantung pada distribusi data.
* **Penanganan Fitur Kategorikal (jika ada):** Jika ada fitur kategorikal yang relevan setelah langkah-langkah sebelumnya (misalnya, gaya renang), fitur tersebut akan diubah menjadi format numerik yang dapat dipahami oleh algoritma pembelajaran mesin (misalnya, menggunakan *one-hot encoding*).

Tujuan dari pemrosesan awal ini adalah untuk memastikan bahwa model dilatih hanya pada fitur-fitur yang relevan, informatif, dan bersih.

**Analisis Data Eksplorasi**

Sebelum melatih model, analisis data eksplorasi akan dilakukan untuk memahami karakteristik dataset:

* **Analisis Korelasi:** Matriks korelasi akan dihitung untuk mengidentifikasi pasangan fitur yang berkorelasi tinggi. Jika ditemukan fitur-fitur yang sangat berkorelasi, salah satunya mungkin perlu dihilangkan untuk mengurangi redundansi dan potensi masalah *multicollinearity* dalam model.
* **Distribusi Kelas:** Distribusi variabel target "classe" (kualitas gerakan renang) akan diperiksa untuk mengidentifikasi apakah ada ketidakseimbangan kelas. Ketidakseimbangan kelas dapat mempengaruhi kinerja model, dan teknik khusus (misalnya, *oversampling*, *undersampling*, atau penggunaan bobot kelas) mungkin diperlukan untuk mengatasinya.
* **Visualisasi Data:** Visualisasi seperti *scatter plot* atau *boxplot* dapat digunakan untuk memahami hubungan antara fitur-fitur sensor yang berbeda dan variabel target "classe". Ini dapat membantu dalam mengidentifikasi fitur-fitur yang paling diskriminatif.

**Pelatihan Model**

Untuk memprediksi kualitas gerakan renang, model klasifikasi akan dipilih. Mengingat keberhasilan model Random Forest dalam proyek angkat besi dan kemampuannya untuk menangani data berdimensi tinggi serta hubungan non-linear, **Random Forest** akan menjadi kandidat utama. Namun, model lain seperti *Support Vector Machines (SVM)*, *K-Nearest Neighbors (KNN)*, atau algoritma *boosting* (misalnya, *Gradient Boosting*) juga dapat dieksplorasi dan dibandingkan.

Langkah-langkah pelatihan model akan meliputi:

1. **Pembagian Data:** Dataset yang telah diproses akan dibagi menjadi set pelatihan dan set validasi (atau pengujian). Proporsi pembagian yang umum adalah 70% untuk pelatihan dan 30% untuk validasi (atau dapat menggunakan teknik validasi silang).
2. **Pelatihan Model:** Model yang dipilih (misalnya, Random Forest) akan dilatih menggunakan set pelatihan. Jika menggunakan Random Forest, jumlah pohon keputusan (*n\_estimators*) akan ditentukan (misalnya, 100 atau nilai lain yang dioptimalkan).
3. **Validasi Model:** Setelah pelatihan, kinerja model akan dievaluasi menggunakan set validasi.

**Evaluasi Model**

Kinerja model akan dievaluasi menggunakan metrik klasifikasi yang relevan:

* **Akurasi (Accuracy):** Proporsi prediksi yang benar.
* **Presisi (Precision):** Untuk setiap kelas, proporsi gerakan yang diprediksi sebagai kualitas tersebut yang sebenarnya memiliki kualitas tersebut.
* **Recall (Sensitivity):** Untuk setiap kelas, proporsi gerakan yang sebenarnya memiliki kualitas tersebut yang berhasil diprediksi dengan benar.
* **F1-score:** Rata-rata harmonik presisi dan recall, berguna ketika ada ketidakseimbangan kelas.
* **Matriks Konfusi (Confusion Matrix):** Tabel yang menunjukkan jumlah prediksi benar dan salah untuk setiap kombinasi kelas aktual dan prediksi. Ini memberikan wawasan tentang jenis kesalahan yang dibuat oleh model.

Hasil evaluasi pada set validasi akan memberikan indikasi seberapa baik model dapat melakukan generalisasi pada data gerakan renang yang belum pernah dilihat.

**Prediksi pada Data Uji**

Setelah model yang memuaskan dikembangkan dan divalidasi, model tersebut akan diterapkan pada set data uji yang berisi 20 kasus gerakan renang baru yang kualitasnya tidak diketahui. Model akan memprediksi kualitas gerakan ("classe") untuk setiap kasus uji. Hasil prediksi ini akan disimpan dalam format yang sesuai untuk pelaporan.

**Kesimpulan**

Proyek ini bertujuan untuk berhasil mengembangkan model pembelajaran mesin yang mampu mengklasifikasikan kualitas gerakan renang berdasarkan data akselerometer. Dengan memilih model yang tepat (kemungkinan Random Forest sebagai titik awal) dan melakukan pemrosesan awal data serta evaluasi yang cermat, diharapkan model dapat mencapai kinerja yang baik dalam mengidentifikasi kualitas gerakan renang.

Perbaikan di masa depan dapat melibatkan:

* **Penyetelan Hyperparameter:** Mengoptimalkan *hyperparameter* model yang dipilih (misalnya, jumlah pohon dalam Random Forest, kedalaman maksimum pohon) menggunakan teknik seperti *grid search* atau *random search* untuk meningkatkan akurasi.
* **Eksplorasi Model Lain:** Membandingkan kinerja model Random Forest dengan algoritma klasifikasi lain seperti *Gradient Boosting*, *SVM*, atau *Neural Networks* (jika data dan sumber daya memungkinkan).
* **Penambahan Fitur:** Mengeksplorasi pembuatan fitur baru dari data akselerometer (misalnya, statistik jendela waktu, transformasi domain frekuensi) yang mungkin lebih informatif untuk memprediksi kualitas gerakan.
* **Penggunaan Teknik Pembelajaran Mendalam:** Jika dataset cukup besar, model pembelajaran mendalam seperti *Convolutional Neural Networks (CNNs)* atau *Recurrent Neural Networks (RNNs)* dapat dieksplorasi untuk menangkap pola temporal dalam data akselerometer.