The Sagitto Workflow

Sagitto functions as a **Data Science as a Service (DSaaS)** platform. This means the process is collaborative: **you provide the data, and Sagitto's data scientists build and manage the models for you.**

Here is the step-by-step process of how a machine learning model is built using the Sagitto system.

1. Data Collection (Your Role)

The foundation of a Sagitto model is high-quality physicochemical data, typically gathered using **Near-Infrared (NIR) Spectroscopy** or sometimes Machine Vision.

- **Scan Samples:** You use a spectrometer (handheld or benchtop) to scan your physical samples (e.g., fruit, textiles, chemicals). This generates the "spectral data" (the raw input for the AI).
- **Establish Ground Truth:** For the model to learn, it needs to know what it is looking at. You must analyze a subset of your samples using traditional methods (e.g., sending them to a wet chemistry lab) to get accurate reference values (like sugar content, moisture levels, or chemical composition).
- **Anonymization:** You can anonymize sample IDs and target names before sharing, ensuring data privacy while still allowing the model to detect patterns.

2. Data Ingestion & Pre-processing (Sagitto's Role)

Once you upload your spectral data and reference values to the Sagitto cloud platform, our automated pipeline takes over.

- **Outlier Detection:** Sagitto's algorithms scan the data to identify "outliers"—scans that are statistically abnormal due to user error or contaminated samples. These are flagged or removed to prevent them from confusing the model.
- **Preprocessing:** The raw spectral data is cleaned and transformed. This often involves mathematical techniques (like derivatives or scatter correction) to reduce noise and highlight the signal relevant to the chemical properties you are trying to measure.

3. Model Training & Validation (Sagitto's Role)

Our data scientists use the cleaned data to train machine learning algorithms (typically multivariate regression or classification models).

- **Algorithm Selection:** We select the best algorithm suited for spectroscopy data to map the relationship between the NIR spectrum (input) and the chemical property (output).
- **Group-Wise Cross-Validation:** Instead of a simple random split, Sagitto often uses "group-wise" cross-validation. If you scanned the same physical sample multiple times, this ensures all scans of that single sample stay together (either all in the training set or all in the test set). This prevents "data leakage" and ensures the model is truly being tested on unseen samples.

- **Metric Evaluation:** The model is evaluated using key metrics:
  - **R^2 (Coefficient of Determination):** How well the model explains the variance in the data.
  - **RMSE (Root Mean Square Error):** The standard deviation of the prediction errors.
  - **MAE (Mean Absolute Error):** The average size of the errors.

4. Benchmarking (Joint Role)

Before full deployment, the model is tested to prove its accuracy.

- **Blind Testing:** You may send Sagitto a set of spectral data *without* the reference values. Sagitto runs these through the new model and sends you the predictions.
- **Comparison:** You compare their predictions against your actual lab results to verify the model works in the real world. Sagitto offers this as a specific "Benchmarking Service" to compare their AI models against standard factory calibration models.

5. Deployment & Access (The Product)

Once the model is validated, it is deployed to the cloud.

- **Mobile App:** You can access the model's predictions via the Sagitto iOS or Android apps. When you scan a new sample with a connected spectrometer, the phone sends the data to the cloud, the model runs the prediction, and the result (e.g., "Dry Matter: 14%") appears on your screen in seconds.
- **Sagitto Website:** You can also access predictions as PDF reports that you can download from Sagitto's website
- **API Integration:** For enterprise users, Sagitto provides an API. This allows your internal factory systems to send spectral data programmatically and receive predictions instantly to automate production lines.

6. Continuous Improvement

The model is not static. As you scan more unique samples over time (e.g., different crop harvests or new product formulations), Sagitto retrains the model. This "active learning" loop ensures the model becomes more robust and accurate the more it is used.

Summary of Workflow

Step	Action	Owner
1	**Scan** physical samples with spectrometer	**You**
2	**Analyze** samples in lab for reference data	**You**
3	**Upload** data to cloud	**You**
4	**Clean** data & train ML algorithms	**Sagitto**
5	**Validate** model accuracy	**Both**
6	**Predict** new samples via App/API	**You**