Biotechnology

Scientific ML evaluation checklist

From the field note “Your model scored 0.89. Will it rank next quarter’s compounds?”.

The short version

- A random-split score mostly measures how well a model recognizes close relatives of its training data. Decide which question you need answered first.
- Split by scaffold or sequence cluster, by time and by assay batch, so the test set looks like the data the model will meet next.
- Report every score with its data and model versions, its uncertainty and a simple baseline, so anyone can rerun it.
- Keep scientists in the acceptance loop. They define similarity, endpoints and what counts as a useful result.

The checklist

- [ ] Write the decision the model supports and who acts on it.
- [ ] Have scientists approve similarity, grouping and endpoints.
- [ ] Hold out whole series, time windows or assay batches.
- [ ] Keep replicates together and block prohibited overlap.
- [ ] Choose a metric that matches the shortlist decision.
- [ ] Report intervals, per-segment results and simple baselines.
- [ ] Record data, split, code and model versions for every score.
- [ ] Re-evaluate on new data before anyone relies on the next prediction.

Want a second pair of eyes on your answers?

Book a free 30-minute discovery call at www.seriustech.com/contact or email contact@seriustech.com. We reply within 1 business day.

Read the full note with examples and sources at www.seriustech.com/blog/biotech-ml-model-evaluation. © SeriusTech.