

# Race Modeling Lexicon: Keyword Recall Sheet

A compact table of key concepts, decisions, and technical terms from the race modeling project. Use this to anchor daily work in consistent principles and avoid drift.

#	Keyword	One-Line Summary
1	Leakage	Including information in training that wouldn't be known at prediction time.
2	RPR	Racing Post Rating — expert-informed, pre-race signal that must be archived pre-off.
3	OR	Official Rating from BHA — stable and published prior to race day.
4	Race-Relative	Modeling a horse within its field, not in isolation.
5	Softmax	Normalizes predicted probabilities per race — ensures valid probability distribution.
6	Transformer	Neural architecture ideal for modeling sets of interacting entities like race fields.
7	Embeddings	Learnable vector representations of categorical IDs (trainer, jockey, course, etc).
8	Trainer/Jockey Form	Recent performance metrics — 14-day stats with strong predictive power.
9	GoingStick	Quantitative measure of ground firmness — highly granular going descriptor.
10	Commentary	Narrative features like spotlight and comment — source of expert judgment, used in NLP.
11	Feature Engineering	Crafting and transforming features to inject domain intelligence into models.
12	JSON Flattening	Converts nested raw data into structured tabular rows for modeling.
13	GroupShuffleSplit	Cross-validation method that respects groupings like race_id.
14	Per-Race Inference	Prediction occurs per race, not per horse — avoids cross-field contamination.
15	Simulated ROI	Evaluation metric that mimics betting outcomes — realism over raw accuracy.
16	Pipeline	The end-to-end structure connecting preprocessing, modeling, and evaluation.
17	NLP Embedding	Converts text (e.g., spotlight) into vectors using sentence transformers.
18	Backfill Risk	Using historical fields that were updated after the race introduces leakage.
19	Course Profile	Encoded characteristics of racecourses — useful for modeling course bias and context.
20	Capability Mapping	Self-reflective summary of technical and domain contributions to the project.

This file should be included in each daily working folder to reinforce conceptual grounding and modeling intent.