

TrackTempo Transformer Evaluation Pipeline

Date: 2025-03-30

1. Evaluation Pipeline Script

- Load predictions made by the transformer model.
- Merge predictions into the model_enriched DataFrame.
- Filter for evaluable rows:

```
eval_df = model_enriched[  
    model_enriched["decimal_odds"].notna() &  
    model_enriched["position"].notna()  
]
```

- Compute metrics:

Accuracy@1

ROI simulation

Model vs. SP (value margin, overlays)

Calibration or rank performance

2. Evaluation Documentation / PDF

- Create summary PDF after evaluation for reporting:

Total races evaluated

Top-1 accuracy

ROI from simulated betting

% where model beat SP (beat_sp_flag)

Distribution of model vs. market odds

- Can include plots (calibration, overlays, histograms)

- Optional: Export evaluation summary to CSV

3. Notebook or CLI Step

- After transformer inference (predictions):

```
from evaluation_tools import evaluate_model_predictions
```

```
eval_df = evaluate_model_predictions(model_enriched, transformer_outputs)
```

- Or run CLI command:

```
python run_evaluation.py --preds model_preds.csv --data model_enriched.pkl
```

4. Additional Notes

- Evaluation should only be done on runners with known position and SP.
- Model-ready data should NEVER include post-race fields during training.
- Future: add logging, plots, or visual dashboard output.
- You can simulate different betting strategies based on value flags.