Squirrels, Soccer, & Safe Predictions:

A Machine Learning Odyssey Through Nuts, Noise, and Neural Nets

Project Overview

- Two datasets: football matches and squirrel sightings
- Goal: build predictive models while avoiding data leakage
- Compare architectures: RNN, MLP, Logistic Regression, ShallowNet
- Explore metaphorical and literal squirrel behavior

Datasets:

- **Soccer Dataset (via API):**
  - Match outcomes (Home / Draw / Away)
  - Pre-match features only (leakage-safe)
- **Squirrel Dataset (via CSV):**
  - Sightings in Central Park
  - Features: fur color, activity, location, time of day

Models Used:

- RNN (LSTM) — for sequential match data
- MLP — for flat features (both datasets)
- Logistic Regression — baseline
- ShallowNet — custom regression model for squirrel behavior

Soccer Model Results

Model	Accuracy	Macro-F1	Train Time (s)
RNN (LSTM, safe features)	0.408	0.254	31.85
MLP (safe features)	0.482	0.400	33.00
Kernel Regressor (LR)	0.466	0.380	0.50

Draws were quite a challenge. It’s hard enough to predict a winner, let alone a tie.

Squirrel Model Observations: *They were nuts!*

- Predicting squirrel activity or location based on features
- Building NaN-safe pipeline with scaling and shallow neural net
- Comparing baseline vs enhanced (Swish, dropout, batchnorm)

Data leakage and model corrections were a significant part!

The Data Leakage Match

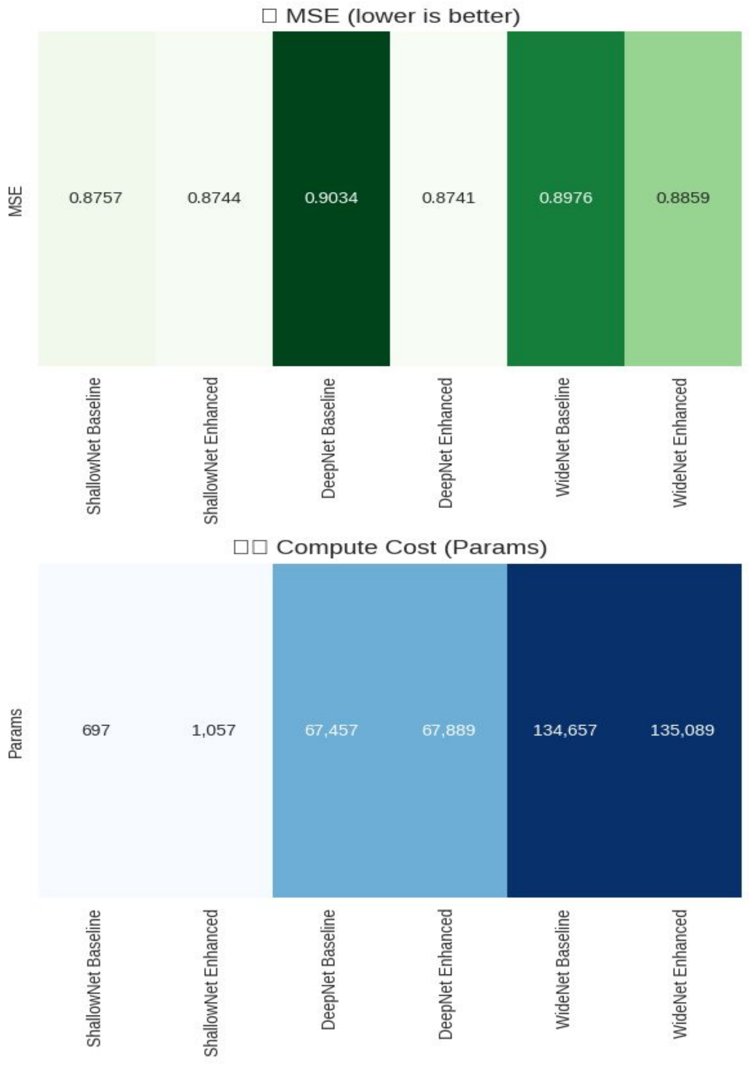
Squirrels: time-of-day and location vs. Soccer: post-match stats affected accuracy

Our fixes included concepts like using only pre-event features, averages of ‘goal differential’ BEFORE the actual match being predicted, and removing potential *NaN* conflicts.

The result? A fair fight. A tough battle, but the models ended up working a lot better.

Simplicity Wins When Shaped Well:

- Complex models (RNNs, deep nets) didn’t outperform simpler ones
- MLP and Logistic Regression delivered stronger, more balanced results
- Careful feature shaping, scaling, and leakage control made the difference
- Simplicity + disciplined approach >>> complexity + noise



Two datasets, one philosophy:
clean data, fair models

And machine learning isn’t just math –
it’s an awful lot of madness!