TrueSkill ELO Uncertainty Analysis
(Model Performance Confidence Assessment)

ELO Rating vs. Uncertainty

Model Ranking Uncertainty