**Project Title: Football (Soccer) Throw-Ins Prediction & +EV Detection**

**Freelancer Name: Tadrous Adel William**

**1) Executive Summary**

I will build a practical system that predicts football (soccer) **throw-ins**.
It will use historic and live match data to predict:

- Throw-ins for each team
- Total match throw-ins
- Probability ranges (distribution)
- **Fair odds** and **+EV** (positive expected value) compared to bookmaker lines

You will get clear numbers, confidence levels, and simple "bet/no bet" signals.

**2) Introduction and Background**

**About me:** I work with Python, SQL, and machine learning. I've built end-to-end data projects: data pipelines, feature engineering, model training, evaluation, and simple dashboards/APIs. I'm comfortable with Pandas, scikit-learn, and model reporting, and I focus on clear, honest metrics and documentation.

**) Proposed Solution & Added Value**

**What I will build**

1. **Data pipeline:** Pull historic and live data from sources like API-Football / Opta / StatsBomb (based on your licenses). Clean, standardize, and store the data.
2. **Features:** Pre-match (lineups, formations, rest days, home/away, team form), and live (possession %, territory, press intensity proxies, fouls, score/time, weather, referee).
3. **Models:** Start with strong count models (Poisson/Negative Binomial) and test tree-based ML (e.g., gradient boosting). Keep the best model by validation results.
4. **Outputs:** Mean predictions, probability ranges, confidence intervals, **fair odds**, and **+EV flags** for totals and team throw-ins.
5. **Interface:** Lightweight dashboard (and/or simple API) to view matches, predictions, fair odds, and recommended edges in real time.

6. **Backtesting & reporting:** Rolling backtests, calibration checks, and hit-rate tracking to prove value before going live.

**Added value**

- Context-aware pricing, not just averages
- Clear uncertainty (confidence intervals)
- Transparent backtesting so you know when the edge is real
- Fast, simple interface for action

**4) Timeline and Deliverables**

**Milestone 1 – Data & Baseline (Week 1–2)**

- Connect to chosen data source(s)
- Clean historical dataset for selected leagues
- Baseline model + short report (initial accuracy and gaps)

**Milestone 2 – Feature-Rich Models (Week 3–4)**

- Add contextual features (lineups, possession, referee, weather)
- Compare models, select the best
- Backtesting report: accuracy, calibration, stability

**Milestone 3 – Odds & +EV Engine (Week 5)**

- Convert distributions to fair odds
- Compare with bookmaker lines, flag edges
- Optional stake sizing (e.g., simple Kelly fraction)

**Milestone 4 – Productization (Week 6)**

- Simple web dashboard (login) with live refresh
- Documentation + handover guide
- Monitoring basics: data freshness and drift alerts

**Deliverables:** cleaned datasets, trained model(s), config files, dashboard access, reports, and documentation.

**5) Resources**

- **Data:** API-Football (quick start) and/or your Opta/StatsBomb access
- **Tools:** Python (Pandas, scikit-learn), basic database (Postgres/SQLite), simple web dashboard
- **Access needed:** API keys, target leagues, bookmaker markets/lines you care about
- **Hosting:** your VPS/cloud (I can containerize for easy run), or local to start

**6) Technical Specifications**

- **Targets:** home throw-ins, away throw-ins, total throw-ins; probability ranges (e.g., 25–29, 30–34, etc.).
- **Evaluation:** MAE/RMSE for counts; calibration and coverage for intervals; backtests by league/season to avoid data leakage.
- **Distribution to odds:** convert predicted distribution into win probabilities for each line/market; calculate fair odds and edge %.
- **Live updating:** configurable refresh (e.g., every 30–60 seconds) during matches.
- **Exports:** CSV/Excel for predictions and edge lists; simple dashboard view for quick decision-making.

**7) Terms and Conditions**

- **Scope:** throw-ins predictive model, odds/edge engine, dashboard, docs, and basic monitoring.
- **Data rights:** you provide/approve data sources and ensure license compliance.
- **Acceptance:** we agree success metrics (e.g., calibration quality and backtest edge after fees).
- **Ownership:** you own the delivered project and artifacts (excluding standard open-source tools).
- **Support:** post-handover support can be added as a separate plan.

**8) Conclusion and Call to Action**

I'm ready to start. Please confirm:

1. Which data provider(s) we will use first,
2. The leagues/competitions to include, and

3. Example bookmaker lines you want to target.

Once I have this, I will deliver the baseline quickly and iterate to the full live system.