Anvil o1

Predicting Federal Contract Winners

Milo Labs, Inc.

G. McCain M. Sadanand A. Boanoh

1 The Problem

The U.S. federal government is the largest buyer on Earth, awarding over $700B in contracts annually. For companies competing for that spend, the central question is deceptively simple:

> *"If I bid on this contract, will I win?"*

Preparing a federal proposal is expensive—often tens of thousands of dollars for routine bids and sometimes millions for major defense programs. Firms invest heavily in capture management, compliance documentation, and pricing strategies without clear visibility into whether they are a frontrunner or a long shot.

Historically, answering "will I win?" relied on intuition, relationships, and educated guesses. Incumbents had an advantage because they knew their own win rates; everyone else was flying blind.

Anvil o1 replaces guesswork with prediction: machine learning trained on millions of historical contract awards to forecast who will win future solicitations.

2 The Core Insight: Creating Supervised Labels

Every federal contract has two distinct moments, recorded in separate government systems:

1. **Solicitation:** the request for bids posted on `SAM.gov`
2. **Award:** the selection of a winner recorded in `FPDS`

These systems do not reliably link to one another. There is often no official field that maps an award back to its original solicitation. This creates a powerful opportunity:

> If you can match each award to its original solicitation, you create a labeled dataset where the input is *what the government asked for* and the label is *who won*.

That is supervised learning—and that is the core of Anvil o1.

3 Why Start with the Defense Logistics Agency (DLA)?

DLA is an ideal starting point because it offers:

3.1 Volume

DLA acts as the Pentagon's supply chain manager, purchasing everything from fuel and clothing to spare parts. High volume yields a large training corpus.

3.2 Standardization

Many DLA purchases are commodity items with National Stock Numbers (NSNs), and solicitation formats are relatively consistent—more modelable than bespoke, subjective procurements.

3.3 Repeatability

The same NSNs are purchased repeatedly; repetition is gold for learning vendor-win patterns.

3.4 Clear Outcomes

DLA contracts are often fixed-price awards with a single clear winner, simplifying the prediction target.

3.5 Provable ROI

DLA contractors skew small/medium businesses where bid costs matter; prioritization intelligence can be immediately valuable.

4 The Data Pipeline

4.1 Step 1: Collect the awards (FPDS)

We ingest the complete DLA contract archive including contract actions, vendors (DUNS/UEI), dollar values, PSC/NAICS codes, contracting offices, dates, and set-aside designations.

4.2 Step 2: Collect the solicitations (SAM.gov)

We pull solicitations containing descriptions, quantities, delivery requirements, set-asides, approved source constraints, attachments, and deadlines.

4.3 Step 3: Link solicitations ↔ awards

Because there is no shared key, linking is probabilistic. We infer connections using:

- **NSN matching** (when present)
- **Timing** (often 30–90 days from solicitation to award)
- **Amount correlation**
- **Contracting office alignment**
- **Textual similarity**

We prioritize precision over recall: better a smaller set of clean links than a larger noisy set.

4.4 Output: High-confidence linked pairs

The result is approximately **98,000** high-confidence linked solicitation–award pairs used for supervised training.

5 What the Training Data Looks Like

Each training example is a record where the `vendor` field is the prediction target:

```
{
  "text": "25--BOX,AMMUNITION STOW Proposed procurement for NSN 2541015263462 ...",
  "psc": "25",
  "naics": "336390",
  "set_aside": "",
  "vendor": "OSHKOSH DEFENSE",
  "amount": 250000.0
}
```

Field	**Description**
`text`	Raw solicitation text: description, quantities, delivery terms
`psc`	Product Service Code (e.g., "25" vehicular equipment components)
`naics`	Industry classification
`set_aside`	Socioeconomic restriction (SB, SDVOSB, 8(a), HUBZone, etc.)
`vendor`	**Winner** (prediction target)
`amount`	Dollar value of the award

6 Performance: Why 56% Top-10 Accuracy is Remarkable

6.1 Random baseline

DLA awarded contracts to approximately **227,000** unique vendors in the training period. A naive random baseline for Top-10 inclusion is:

$$P(\text{winner in top 10} \mid \text{random}) \approx \frac{10}{227{,}000} \approx 0.0044\%$$

6.2 Observed result

Anvil o1 achieves **56.2%** Top-10 accuracy, meaning the true winner appears in the model's top 10 predictions more than half the time.

6.3 Lift framing

The relevant question is not "is 18% Top-1 good?" but "how much better is the model than guessing without signal?"

7 Feature Engineering: What Drives Predictions

The model uses 200+ features, including:

1. **Product category signals:** PSC and NAICS sharply segment vendor specialization.
2. **Set-aside constraints:** eligibility restrictions act like hard filters.
3. **Vendor history:** wins in-category, recency, typical deal size, and office patterns.
4. **Text-derived features:** NSN patterns, delivery location, timelines, quantity signals.
5. **Agency behavior:** contracting-office/vendor affinity learned from history.

8 Model Architecture: Ranking with Gradient Boosting

Anvil o1 uses a gradient boosting ensemble (e.g., LightGBM/XGBoost) optimized for ranking (LambdaRank/LambdaMART).

8.1 Why ranking vs. classification

Predicting a single winner from $\sim 227{,}000$ vendor classes is unwieldy. Ranking learns a scoring function over candidate vendors and is evaluated by whether the winner appears in the top K.

8.2 Candidate generation

We avoid scoring all vendors by first narrowing the pool using:

- PSC-category participation (vendors with prior wins in-category)
- Set-aside eligibility
- Active status (recent activity)

This often yields 500–5,000 candidates per solicitation for final ranking.

9 What We Discovered

9.1 Incumbency dominates

The strongest predictor is whether a vendor has won the same NSN before.

9.2 Set-asides increase predictability

Set-asides constrain the candidate pool, making outcomes more predictable.

9.3 Category predictability varies

PSC	Category	Top-10 Accuracy
59	Electrical Components	68%
53	Hardware & Abrasives	61%
16	Aircraft Components	54%
84	Clothing & Textiles	49%

9.4 Pricing limits the ceiling

Many DLA buys use LPTA-style evaluation; bid price is not observable prior to award, introducing irreducible uncertainty.

9.5 New entrants are hard

First-time winners in a category have no history, so they are intrinsically difficult to predict.

10 Use Cases

10.1 Contractors: Bid/No-Bid intelligence

If a firm is ranked near the top, it is a strong signal to invest proposal resources; if ranked far down, it may not be worth the effort.

10.2 Contractors: Competitive positioning

The model identifies likely competitors and dominant incumbents, enabling sharper strategy.

10.3 Investors: Pipeline modeling

Forecasting win likelihood helps estimate future revenue streams for public contractors.

10.4 M&A due diligence

Assessing whether revenue is structurally defensible versus "lucky".

10.5 Government (theoretical)

Pattern visibility can reveal concentration and competition dynamics.

11 Implications

1. **Information asymmetry flattens:** new entrants gain visibility comparable to incumbents.
2. **Proposal economics shift:** fewer wasteful bids; more focused competition.
3. **Gaming is limited:** core signals (history, eligibility) are not easily manipulable.
4. **Transparency pressures outcomes:** concentration becomes visible and measurable.

12 What We Expect Going Forward

12.1 Expand beyond DLA

Generalize to GSA, VA, and other DoD components; likely requires per-agency models initially.

12.2 Incorporate richer signals

Potential improvements include transformer-based text understanding, attachment parsing, vendor registration/financial proxies, and protest history.

12.3 Real-time prediction

Score solicitations at posting time with live monitoring and low-latency inference.

12.4 Price modeling (the "holy grail")

If vendors share bid histories, price distributions could be modeled—creating a compounding data network effect.

13 Limitations and Honest Caveats

- **DLA-specific (for now).** Non-DLA predictions require additional training.
- **Commodity-focused.** Not designed for major weapons systems, services, or R&D.
- **Backward-looking.** Requires retraining as markets and policies shift.
- **Not a guarantee.** Predictions are probabilistic guidance.
- **No pricing visibility.** Especially limiting under LPTA dynamics.

14 Summary

Anvil o1 solves a data integration problem that unlocks prediction: link awards (FPDS) to solicitations (SAM.gov) to create supervised labels, then train a ranking model to predict likely winners. In DLA commodity procurement, the result is actionable: **56% Top-10 accuracy** on high-confidence linked pairs, enabling better bid decisions, competitive intelligence, and pipeline forecasting.

This LaTeX document is a faithful typeset of the user-provided brief.