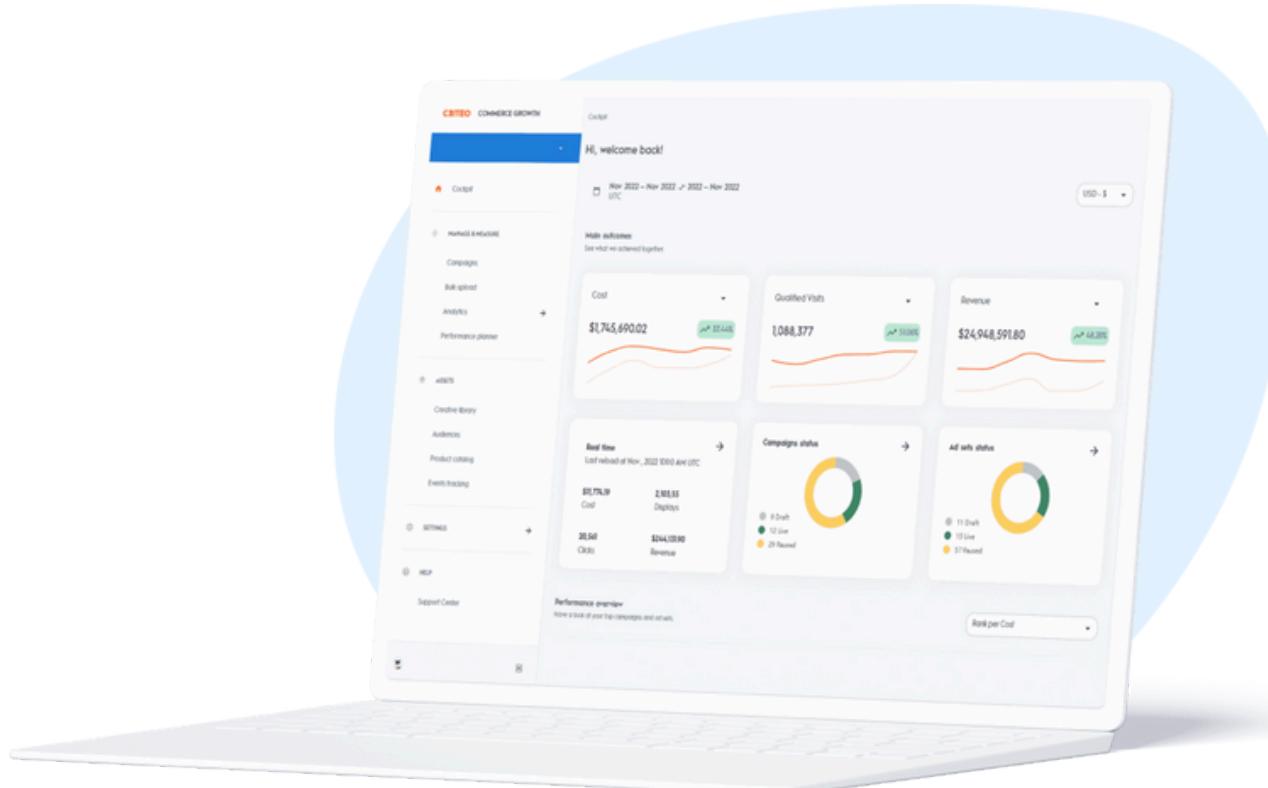


Bidding Model Evaluation at Criteo

Ecem Bayındır

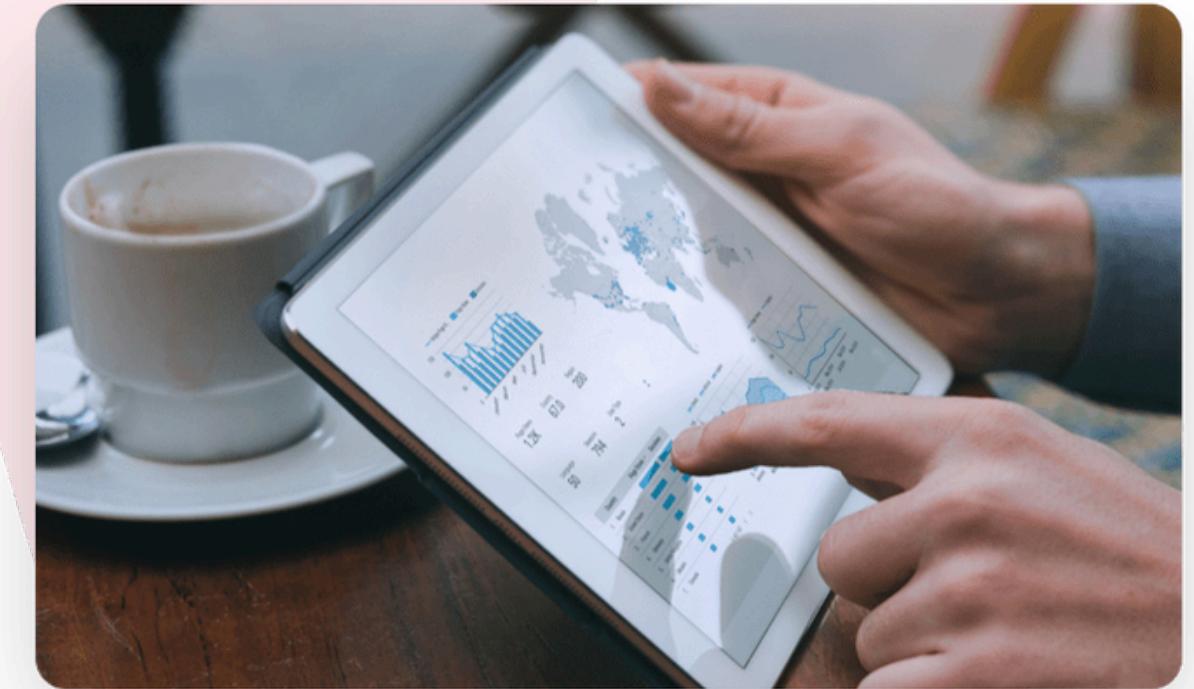
Product Analytics - Data Scientist Candidate

21 April 2025



Project Objective

- Investigate and compare two **model-based bidding** strategies:
 - Linear Regression Bidding Model
 - Deep Learning Bidding Model
- Evaluate models using **profitability, prediction performance, and business impact**
- Recommend a **strategy for production deployment** at Criteo



Business Context

- 1,000 display ad slots on CNN.com are auctioned for Criteo clients
- Two brands, Hollister and Fridge, bid using with **Linear Regression and Deep Learning Bidding Models**
- Clicks and conversions lead to **client's ROI** and **Criteo's revenue**



Dataset Overview

- 30-day sample of **Criteo's live traffic**
- **16.5M** impressions, **45K** conversions, **700** campaigns
- **Features** include:
 - Timestamp, User ID, Campaign ID, Conversion, Click, Cost, etc.
 - Categorical contextual features (cat1 to cat9).
- Public dataset [Link](#)
 - Released with AdKDD 2017 paper **by Criteo AI Lab**



Workflow Summary

- Data Cleaning & Preprocessing
- Feature Engineering
- Model Development (Linear Regression & Deep Learning)
- Threshold Optimization & Bidding Simulation
- A/B Testing (Simulated)
- Business Impact Evaluation



Data Cleaning and Processing

Data Cleaning & Processing

- Removed invalid or placeholder values (e.g. -1 for conversion timestamps, time since last click).
- Replaced missing time values with NaN; created is_first_click flag.
- Sampled 1M impressions stratified on target (click) to ensure class balance.
- Split into Train (70%), Validation (10%), Test (20%).

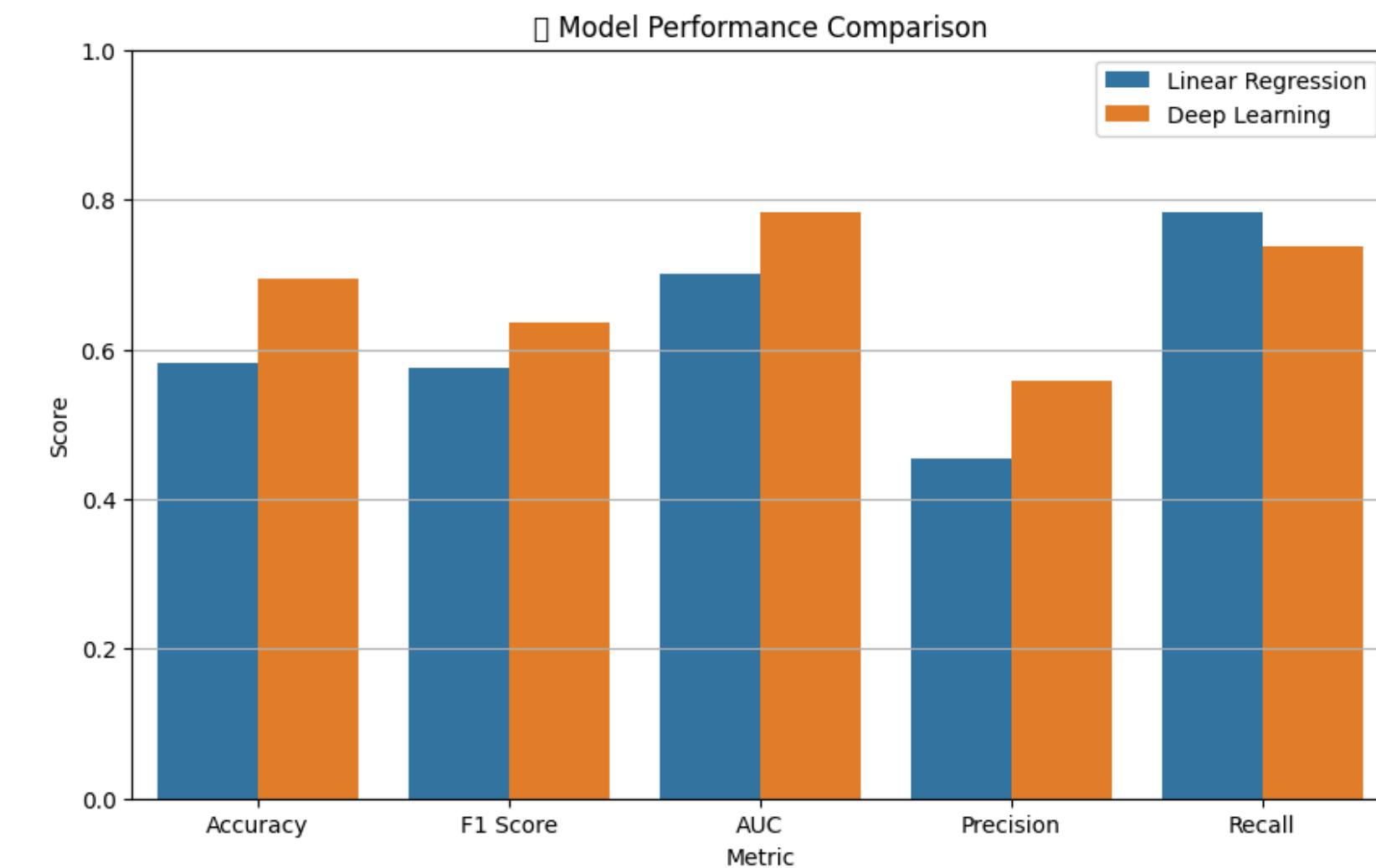


Feature Engineering

- Used all contextual features: cat1–cat9.
- Applied frequency encoding: cat2, cat3, cat7 and one-hot encoding for others
- Created: time_since_last_click_cleaned, is_first_click.

Model Performance Comparison

Metrics	Linear Regression	Deep Learning
Accuracy	0.582355	0.693485
F1 Score	0.575398	0.634860
Precision	0.454627	0.557126
Recall	0.783545	0.737803
AUC	0.701133	0.781959
Profit (\$)	\$ 50372.45	\$ 48510.15



- **DL** outperforms in all **prediction metrics**, **LR** delivers **higher profit**, highlighting a trade-off between accuracy and ROI.
- **Optimized threshold** to maximize F1 Score.
- **Simulated bids**: $\text{base_bid} \times \text{predicted_CTR}$.
- **Evaluated** with classification metrics + profit simulation.

A/B Testing Design (Simulated)

- **Hypotheses**

- H_0 : No statistically significant difference in avg. user-level profit between LR and DL
- H_1 : There is a statistically significant difference

- **Profit Formula:**

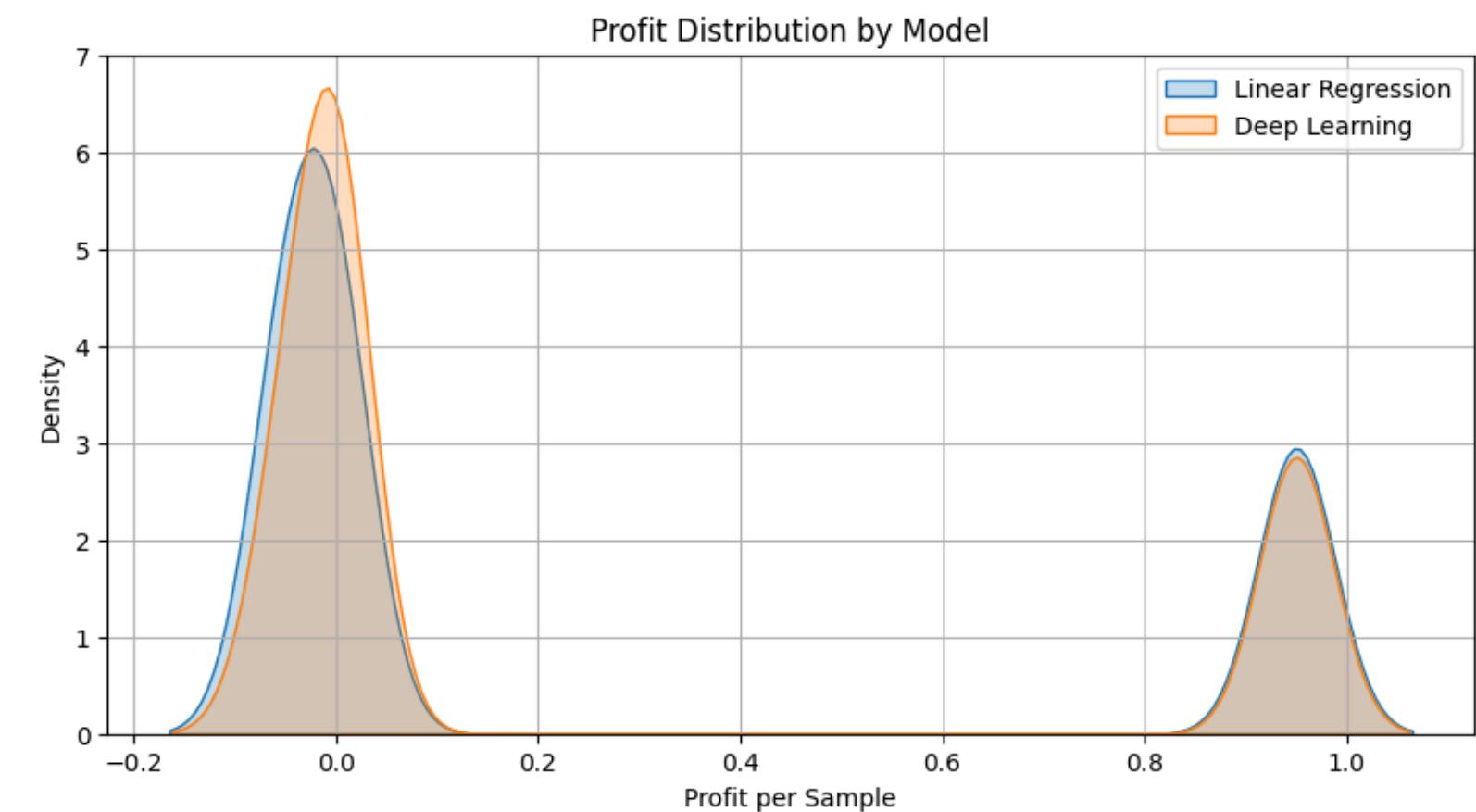
- profit = (conversion * 1.0) - (bid * 0.05)

- **Profit Distribution:** KDE plots show a bimodal shape:

- Peak at \$0: Bids with no conversions.
- Peak at \$0.95: Bids resulting in conversions.
- DL has a stronger skew toward higher-profit conversions.

- **Statistical Test:** Mann-Whitney U Test on 200000 samples:

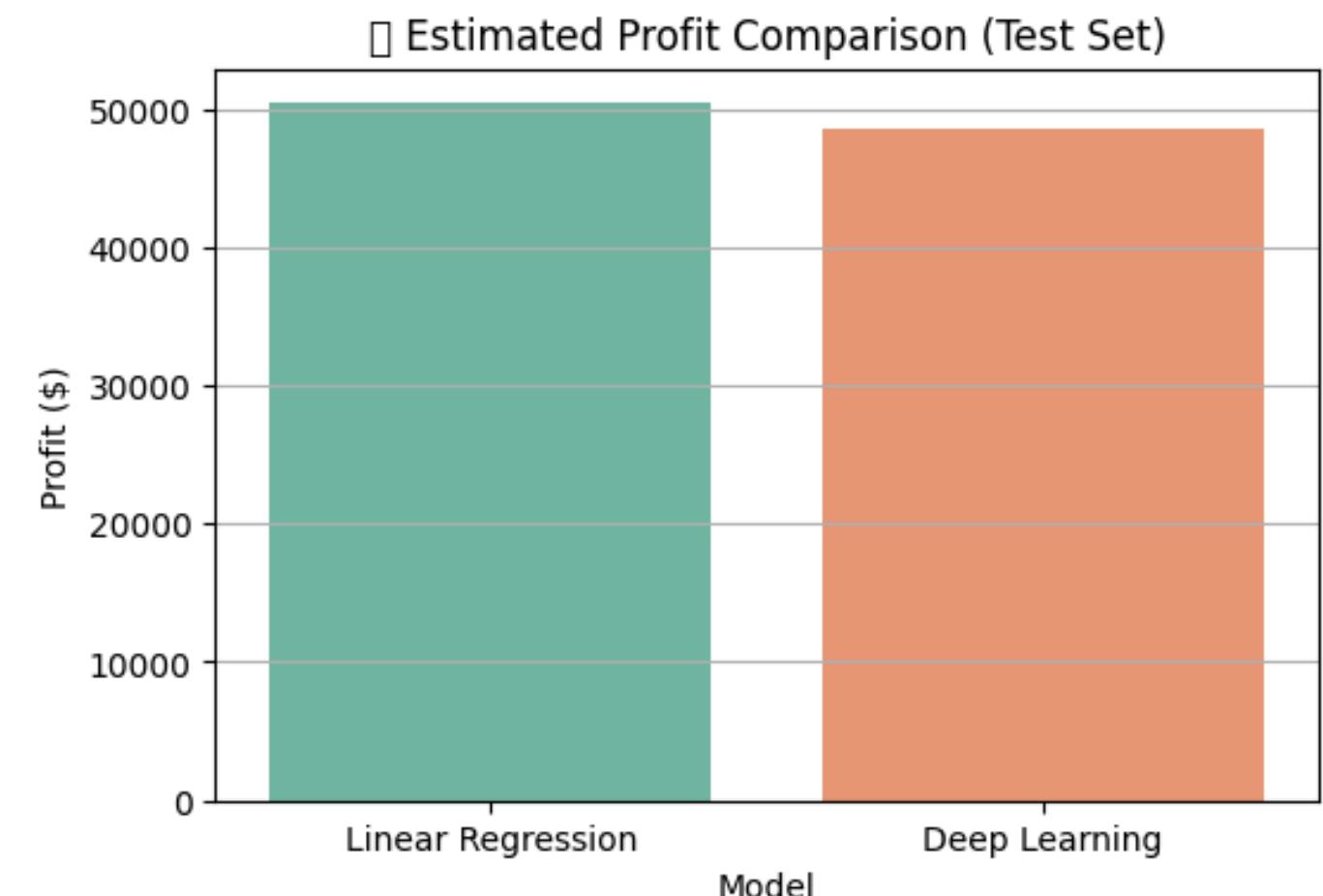
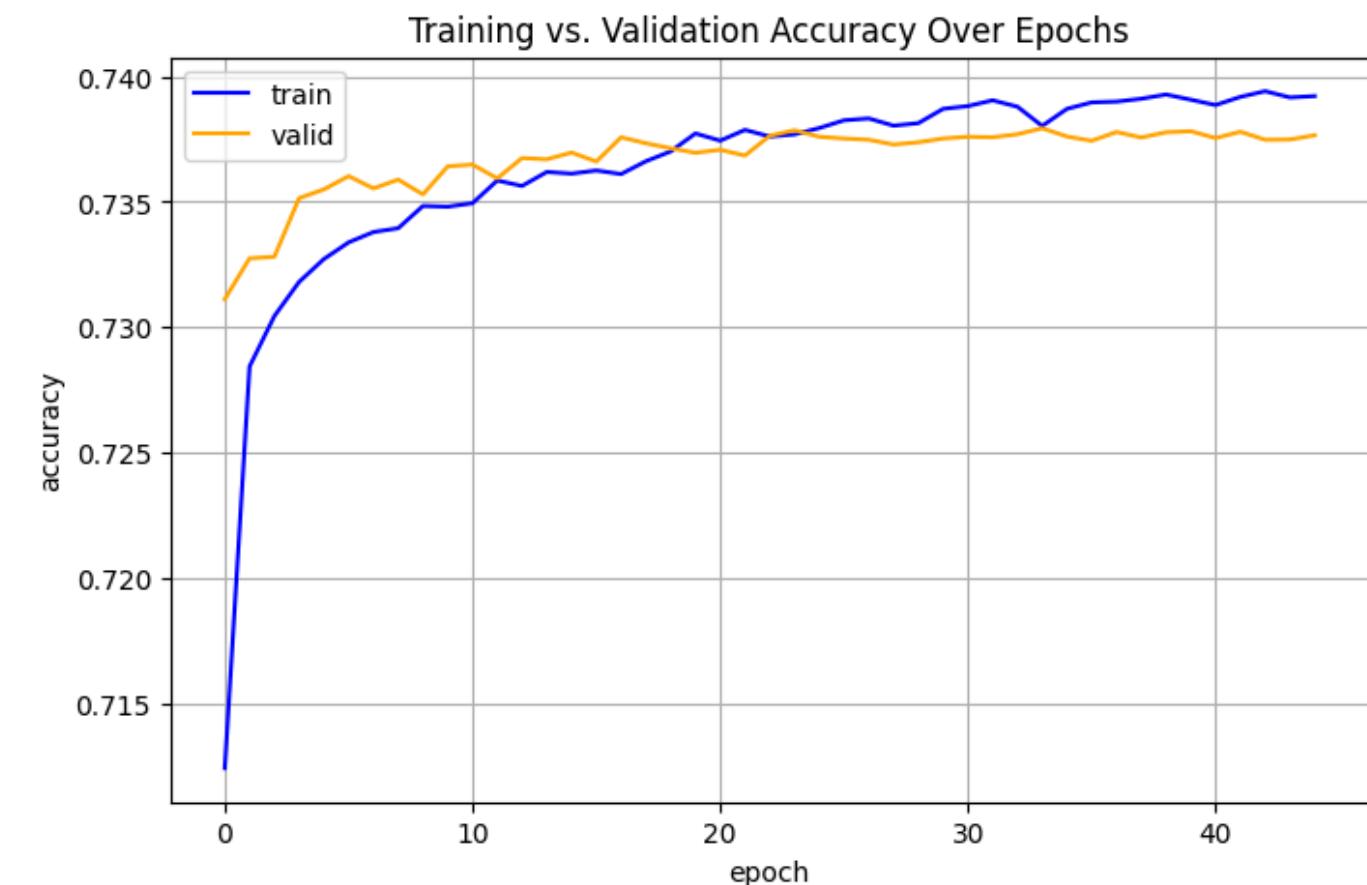
- U statistic: 18358902678.5.
- p-value: less than 0.000
- Alpha (α): 0.05
- Result: Statistically significant difference between models. (**Reject H_0**)



Key Outcome: Linear Regression profit (\$50372.45) exceeds Deep Learning (\$48075.55) by \$2296.90 (about 4.7 percent).

Conclusion

- **Deep Learning Model:**
 - Superior predictive power
 - Higher classification accuracy (**69.5% vs. 58.2%**)
 - Better precision and recall
- **Linear Regression Model:**
 - Slightly higher total profit (**\$50,372 vs. \$48,076**)
 - Statistically significant difference in profit distributions
 - More lightweight and easier to deploy
- **A/B test:**
 - (Mann-Whitney U, $p < 0.0001$) confirms a **statistically significant** difference in user-level profits between models



Recommendation

- **Short-Term:**
 - Deploy the Linear Regression model for production.
- **Long-Term:**
 - Continue iterating on the Deep Learning pipeline for future improvements.
- **Further Testing:**
 - Conduct online A/B testing and explore dynamic bidding thresholds before full deployment of the Deep Learning model.

