Predictive Pricing Model

Faizan Waheed, Sai Bhavya Sree Sirikonda, Shamika Karnik, Shoukath Ali Shaik, Swamini Sontakke

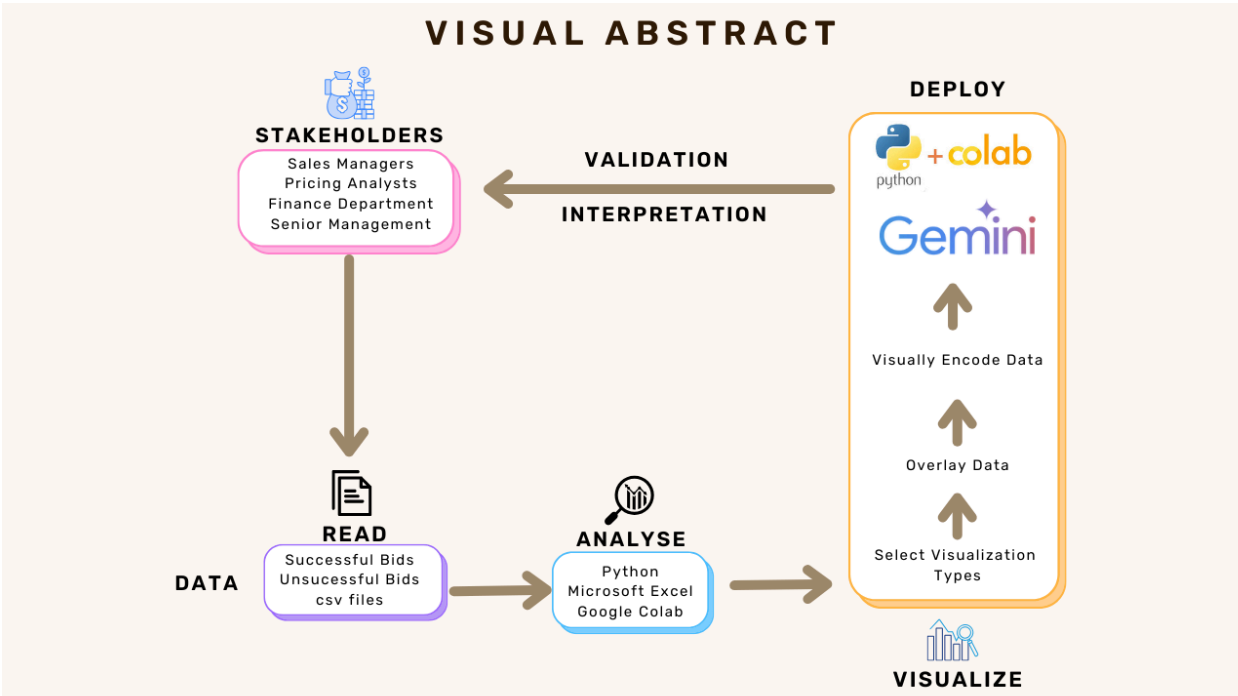


Fig. 1. This project addressed Opinion Route's high number of unsuccessful bids by analyzing client datasets using Random Forest modeling and Gemini AI tools, identifying key factors like CPI and IR, and providing actionable strategies through interactive visualizations and predictive analytics.

1. INTRODUCTION AND PRIOR WORK

Pricing optimization is critical in market research, where dynamic variables influence deal success. Traditional pricing models often rely on static benchmarks, failing to adapt to evolving market conditions. This project aims to develop a data-driven pricing system that enhances win rates by analyzing past sales and unsuccessful proposals, providing actionable insights for strategic decision-making.

Prior studies have applied machine learning to pricing, using regression, reinforcement learning, and probabilistic models to predict customer behavior and price sensitivity [1, 2]. However, many existing approaches suffer from overfitting, lack of adaptability across diverse client profiles, and insufficient differentiation between successful and unsuccessful proposals [3,4].

This project builds upon prior work by introducing a predictive framework that explicitly accounts for deal success factors. By leveraging data visualization to provide clear insights to ensure transparency, it will offer a scalable and informed pricing strategy, filling gaps in current market research pricing methodologies.

1.1 Stakeholder Groups

For this use case we identify the following stakeholder groups:

1. Sales and Account Managers
2. Pricing Analysts
3. Finance Teams
4. Marketing Teams
5. Senior Management

1.2 Stakeholder Needs

All the above mentioned stakeholder groups have the following needs:

1. **Sales & Account Managers:** Optimize deal pricing, understand lost deals, suggest competitive rates
2. **Pricing Analysts:** Develop data-driven pricing strategies using lost deal data
3. **Finance Team:** Ensure profitability by analyzing revenue, costs, and margin impact
4. **Marketing Team:** Identify pricing trends to refine promotional strategies
5. **Senior Management:** Strategic decision-making based on pricing performance

2. DATA ACQUISITION

The data we would be working on consists of 3 different files, one holds data about invoiced deals, and the other about the lost deals. It is provided by the client in the form of csv files.

2.1 Data Sources

The project relies on three datasets provided by the client:

1. **Deal Item Report (LOST):** Contains details of lost deals, including record IDs, quantities, rates, item amounts, deal names, and account names.
2. **Invoiced Jobs (2024):** Includes invoiced project details such as project codes, project names, invoiced dates, revenues, vendor costs, client names, and completion statuses.
3. **Account List with Segment:** Provides account-level details like record IDs, account names, client segment types, and assigned account owners, used for client segmentation and ownership tracking.

Direct links to raw data on Google Drive are available: https://drive.google.com/drive/u/2/folders/12zvOu9qpiO8AlQlUP2S-P2dd_q9PRCda

2.2 Data Description, Quality and Coverage

The **Deal Item Report (LOST)** dataset is structured, covering multiple clients and deals, with minor missing values that did not significantly affect analysis. Temporal coverage is limited to specific business periods.
The **Invoiced Jobs (2024)** dataset is well-organized, detailing invoiced projects across industries and geographies from early 2024. Some fields (like audience categories) have missing values but were manageable through imputation or exclusion.
The **Account List with Segment** dataset contains 153 fully populated records with no missing values, offering clean and consistent categorical data suitable for client segmentation and ownership analysis.

3. DATA ANALYSIS

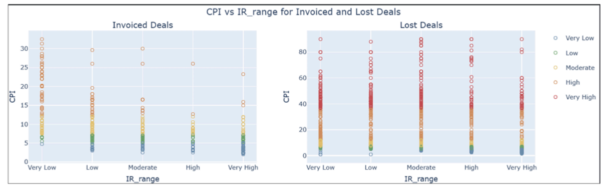
The data analysis focused on uncovering relationships between key bid attributes and success outcomes by leveraging client datasets, including lost deals, invoiced jobs, and account segment information. Microsoft Excel was used for initial data cleaning, exploration, and validation, while deeper statistical analysis and modeling were performed in Google Colab using Python libraries such as pandas, matplotlib, and scikit-learn. Visual comparisons between invoiced and lost deals revealed strong patterns, notably that higher CPI values were consistently linked to bid losses, while lower CPI and higher completes improved win rates. Segment-wise breakdowns also showed variations between consumer and B2B clients. Outlier detection and missing value handling ensured the reliability of the datasets. This analysis built a solid, data-driven foundation for predictive modeling and strategic recommendations aimed at helping OpinionRoute optimize their future bid strategies.

4. VISUALIZATIONS

In order to effectively convey the insights obtained from the Predictive Pricing Model, we utilize a variety of data visualization techniques.

4.1 CPI and IR Relationship

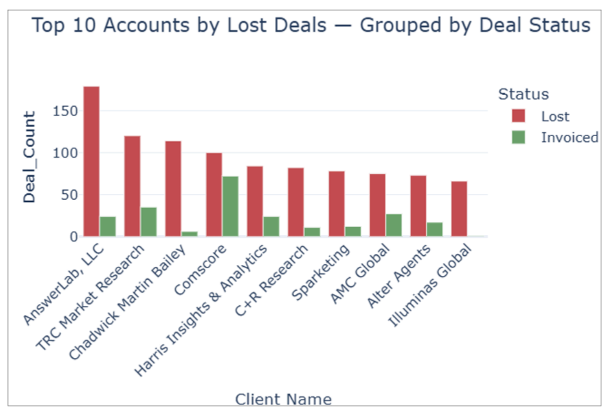
Our first visualization compares the relationship between CPI (Cost Per Interview) and IR_range (Incidence Rate) for invoiced and lost deals. On the left, the invoiced deals show that most CPI values are concentrated below 20 across all IR ranges, with relatively fewer deals at higher CPI levels. On the right, lost deals demonstrate a much wider spread in CPI, with many values exceeding 50, regardless of the IR range.



The color coding by CPI range indicates that higher CPI values are more commonly associated with lost deals. Overall, the chart suggests that while the incidence rate has some variation, high CPI is a stronger indicator of deal loss, emphasizing the importance of managing CPI to improve win rates.

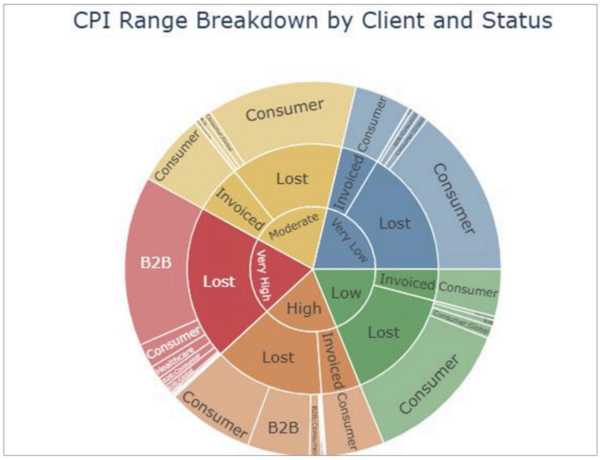
4.2 Clients with Highest Number of Lost Bids

Our second visualization is a column chart which shows the top 10 accounts with the highest number of lost deals, grouped by deal status into "Lost" and "Invoiced." The red bars represent lost deals, while the green bars represent invoiced (won) deals. AnswerLab, LLC leads with the highest number of lost deals, followed by TRC Market Research and Chadwick Martin Bailey. While all listed clients have a higher proportion of lost deals, a few accounts such as Comscore and AMC Global show relatively better invoicing performance compared to others. Overall, the chart highlights the concentration of lost opportunities among specific key accounts and suggests areas where focused improvement efforts could yield better outcomes.

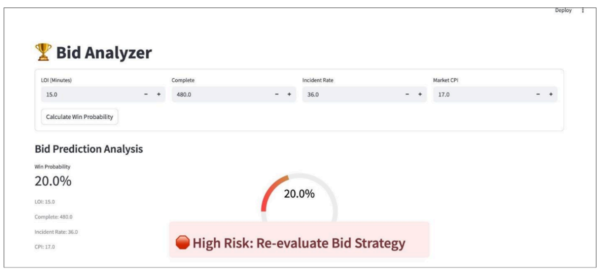


4.3 CPI Range and Client Type Impact on Deal Outcomes

This visualization presents a CPI range breakdown by client type and deal status using a sunburst chart. The inner layers represent different CPI ranges from Very Low to Very High, while the outer layers break down the deal status into Lost or Invoiced, further segmented by client types such as Consumer and B2B. A clear trend is visible: higher CPI ranges (High and Very High) are predominantly associated with lost deals, particularly among B2B clients. In contrast, lower CPI ranges (Very Low and Low) show a healthier mix of invoiced and lost deals, mainly among Consumer clients. This highlights that controlling CPI is critical for improving win rates, especially for B2B segments.



4.4 Bid Success Prediction Model Using ML and Gemini AI



We also developed a Bid Analyzer model on the specific needs of the client. The model uses machine learning techniques to predict the probability of bidding success by analyzing key inputs such as length of interview, required number of completes, incident rate, and market CPI. It evaluates whether a bid aligns more closely with historically won or lost bids and flags high-risk submissions. In addition to prediction, the system integrates Gemini AI tools to recommend targeted actions—such as reducing CPI, increasing the number of completes, and shortening the interview length—to improve the chances of winning. Overall, the Bid Analyzer not only forecasts success probabilities but also provides intelligent, actionable strategies to optimize bidding performance and drive better business outcomes.

5. USAGE AND CRITIQUE OF AI TOOLS

In the client project, AI tools were used to build a bid prediction model that estimated bidding success based on historical data and provided strategic recommendations through Gemini AI. This approach improved efficiency and decision-making by identifying patterns across key inputs like CPI, IR, and completes. To address potential biases in historical data, we carefully curated the training set to ensure a balanced representation of different bid types and market segments. For interpretability, we incorporated feature importance analysis to make the model's reasoning more transparent to users. These steps helped improve both the fairness and trustworthiness of the model, though further enhancements like SHAP-based explainability could be explored in future iterations.

6. INTERPRETATION OF RESULTS

The analysis revealed clear patterns linking bid success to key metrics like CPI, LOI, and IR, allowing users to interactively explore differences between invoice and lost bids. Segment-wise breakdowns further highlighted how client types and respondent profiles influenced outcomes. The Random Forest model delivered highly accurate predictions, while the dynamic gauge visualization made interpreting win probabilities intuitive. Importantly, the model not only predicted outcomes but also identified the most influential input factors, with Gemini AI tools generating actionable recommendations based on these insights. The results support the hypothesis that combining machine learning models with large language models enhances bid strategy planning, providing both accurate predictions and strategic guidance. Overall, the integrated use of interactive analytics, predictive modeling, and AI-driven interpretation created a comprehensive decision-support system, although future improvements could focus on incorporating real-time updates to address current data limitations..

ACKNOWLEDGEMENTS

We would like to thank our Professors, Katy Borner and Michael Ginda for providing us with the opportunity of working on client projects. The team would also like to thank Terence McCarron, the representative from Opinion Route for his valuable input and support throughout the project. We also extend our gratitude to Divya Prasanth Paraman, Teaching Assistant at Indiana University, for his continuous guidance and mentorship during the development of this work.

REFERENCES

1. K. Talluri and G. Van Ryzin, "Revenue Management and Pricing Analytics," *Operations Research*, vol. 50, no. 6, pp. 107-125, 2020.
2. C. Koenig and G. Meissner, "Dynamic Pricing in Competitive Markets," *Journal of Pricing Strategy*, vol. 18, no. 4, pp. 299-315, 2019.
3. S. Gupta and R. Rooderkerk, "AI-Powered Pricing: Opportunities and Challenges," *Marketing Science*, vol. 40, no. 3, pp. 472-490, 2021.
4. BESPOQUOTE, "13 Machine Learning Models for Predictive Pricing," [Online]. Available: https://bespoquote.com/13-machine-learning-models-for-predictive-pricing/.