MIT Sloan Proseminar

Factors in Deciding Buy vs Build

September 2021

Interpretable AI

What's Interpretable AI?

Interpretable AI makes performant machine learning software that creates models that are easily understandable.

We issue licenses as a software vendor, and also offer direct consulting services.

Dr. Jack Dunn, Partner
PhD from MIT
Software engineering @ Google

Dr. Daisy Zhuo, Partner
PhD from MIT
8+ years consulting

Software Offered by Interpretable AI

Python / R / Julia packages that can be installed just like any other packages

Optimal Imputation

Unlock the full power of data with missing values or quality issues

Optimal Feature Selection

Automatic selection of optimal features from the noise

Optimal Decision Trees

As powerful as black-box artificial intelligence with the interpretability of a single decision tree

Interpretable Matrix Completion

Powerful recommender system that gives detailed explanations for each suggestion

Current Users


Finance
Personalized Banking Products Recommendation


Finance
Marketing Recommendations that Maximize Fund Flow


Finance
Predicting Risk of Loan Default


Cybersecurity
Improving Malware Detection in Cybersecurity


Manufacturing
Understanding Machine Failures in Car Manufacturing Plants


Manufacturing
Interpretable Predictive Maintenance for Turbofans


Manufacturing
Predictive Quality Solution for Die Casting in Car Manufacturing


Manufacturing
Monitoring Hard Drives in a Data Center


Manufacturing
Optimal Experiment Design for Automotive Testing


Healthcare
Surgical Risk Calculator: POTTER


Healthcare
Screening Procedure for Pediatric Head Trauma


Healthcare
Mortality Risk Prediction Tool in Cancer Patients


Insurance
Optimizing Data Acquisition


Insurance
Robust Data Pipeline Design


Retail
Assortment Optimization with Consumer Preference Learning


Real Estate
Pricing for Real Estate Auctions

info@interpretable.ai

Interpretable AI

What makes a buy option more attractive than build?

Over the years, we have observed that what our clients get out of the partnership varies.

Today we want to share some success stories and failure modes, with a focus on when buy or build can be more beneficial.

Success Story 1

Real estate platform - pricing

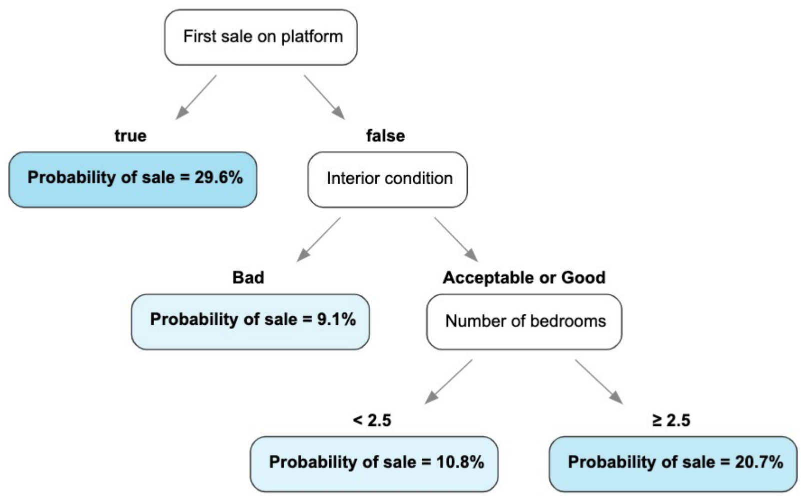
Real estate auction platform needs better pricing to increase sales rate

Our client was one of the largest online real-estate platforms.

It needed a better pricing model, which at the time was based on AVM (automated valuation models), but did not take market conditions into account.

As a result, their current pricing was often suboptimal, either leaving money on the table or too high to attract buyers

Real estate auction platform needs better pricing to increase sales rate



Example decision tree that predicts sales probability

We worked closely with their analytics team and domain experts to deliver a sales-probability driven pricing model.

In a pilot study, the client identified that the new pricing tool could result in 18% additional sales and 24% overall higher revenue.

Real estate auction platform needs better pricing to increase sales rate

The core ingredients to success we saw:

- **Stakeholder involvement**: the COO had been involved from day 1 and very actively participated in the steering committee meetings and communicated to the C-suite
- **Business-driven translator**: the head-of-analytics had keen business sense and did not just focus on the raw model performance but what it meant for the stakeholders
- **Feedback loop from users**: the portfolio managers reviewed the predictions and model logic to ensure the models make sense and as a result improved the weaker components of the model over time

Failure Story 1

Real estate platform – recommender system

Seeing the success, they want a recommender system

After Phase 1, all stakeholders were on board. Based on the success of the first project, they wanted to engage us more heavily.

They had confidence in us and trusted we can deliver, so we set up the project for Phase 2. The goal would be to recommend houses to customers based on their interactions on their website.

Seeing the success, they want a recommender system

As we started the project to look closer into the data, we identified:

- A certain class of properties are listed on the website for information only, so there is no interaction data. This strikes half of the properties

- Only about 1/3 of their users log in to the same account over time, building a profile
  - 1/3 don't log in, and 1/3 create new accounts regularly to hide their behavior

Together, this meant that only 1/3 of the users would have behavior profiles, and only half of the properties were eligible for recommendation. Thus, the market for the project was actually only ~1/6 of what they had planned for in the business case

This distorted the value proposition and caused them to put the brakes on.

Despite the good success of the first project, mis-sizing the follow-up hurt the momentum.

Failure Story 2

Simultaneous internal development

Chemical company wants to use IAI software

An ML lead at a major chemical company was interested in our software for his group. He has strong technical background and is familiar with the methodologies and our papers.

We evaluated a few use cases together in a POC, in which we found very good results and we hoped would lead to longer term licensing agreements, but the discussion slowed down.

Chemical company wants to use IAI software

A long time later, after leaving the company, the ML lead told us that he actually had been working with an engineer in parallel to recreate our algorithms from scratch. They had not pursued the licensing deal because they thought they could build it themselves.

6 months later, they determined that their implementation was too slow and did not scale to the problem sizes they needed, therefore they abandoned the internal build option.

Chemical company wants to use IAI software

Lessons:

- There may be details and know-how that the software vendors do not disclose fully even if their methodologies are published in papers (Deep Mind’s Alpha Go for example). Often reproducing these works seems deceptively easy, but is actually extremely difficult without an army of engineers.

- It can be difficult to know ahead of time whether the internal time investment is too risky compared to the known payoff of licensing the existing solution

Success Story 2

Insurance data company linking records

Insurance data company has data quality issue affecting its client relationship

Our client, an insurance data company, aggregates data from its clients and provides reports and recommendations for its clients.

Because the raw data is complex and messy, their reports are highly sensitive to data quality issues.

Insurance data company has data quality issue affecting its client relationship

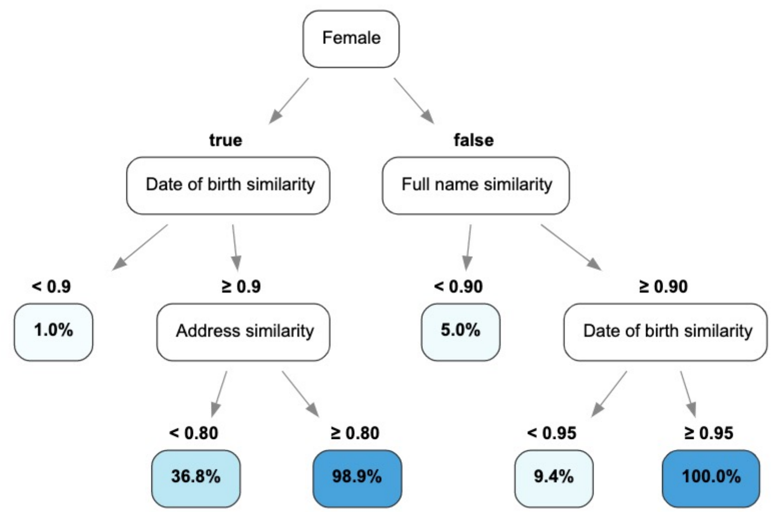
We identified that one major problem is record linking from difference sources, where often typos and missing values in SSN, Date of Birth, and Names prevent us from linking them to the same person, resulting in duplicate records and inaccurate followup analyses

Source	Name	Date of Birth	SSN
Eligibility	Daisy Zhuo	01/01/2000	000000001
Medical claims	Daisy Zhou	01/01/2000	123456789

Example table of two records that are the same person but wouldn't normally be linked

Insurance data company has data quality issue affecting its client relationship

We used Optimal Classification Trees, and in collaboration with assistance from their data experts, developed an automated approach with over 99.5% accuracy.



Example decision tree for record linkage

info@interpretable.ai

Insurance data company has data quality issue affecting its client relationship

The core ingredients to success we saw:

- **Well-scoped problem**: it does not aim to “solve all data quality issues”, but targets a well-defined and high-impact component of data quality
- **Data collection on demand**: we had access to their data experts who can label new records in areas where the model needs more help
- **Clearly defined metrics**: it was clear at the start that minimizing false positives is the first objective, and while trying to avoid introducing too many false negatives. This gives a clear direction when tuning the models to be most useful

Failure Story 3

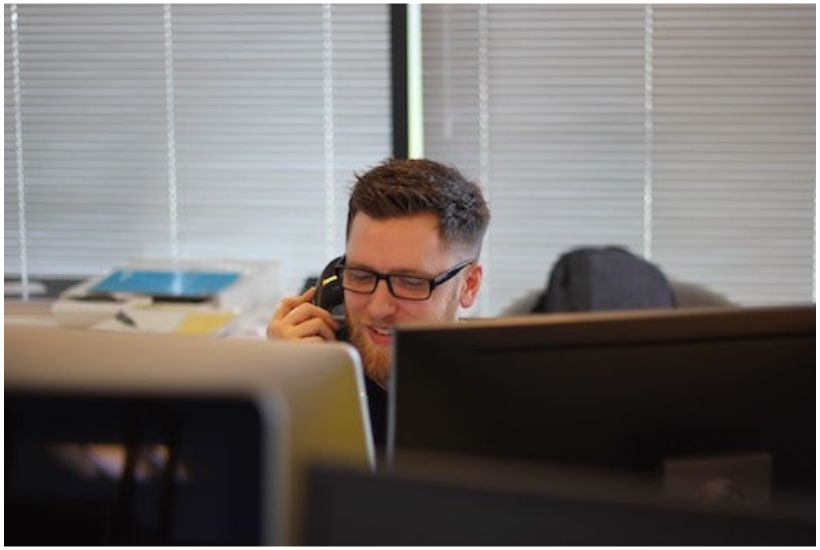
Financial company marketing strategy PoC

info@interpretable.ai

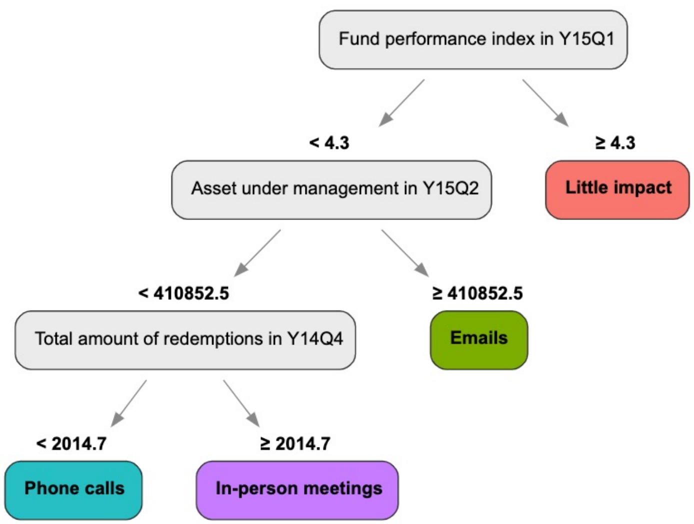
Asset management company needs to know who to market to and how

An asset management company has hundreds of fund managers that it works with, and it wants to learn best interaction strategies to maximize fund flow.

They have rich data detailing how customers respond to marketing efforts and we decided to do a PoC together.



Asset management company needs to know who to market to and how



Example tree outputs that recommends best interaction strategy

We developed a PoC product that recommends which interactions should be used to maximize their fund flow.

In a pilot study it showed a 8-15% increase in fund flow under the new data-driven strategy.

Asset management company needs to know who to market to and how

After the PoC, they decided not to expand the tool in continued collaboration with us. Instead, they assigned a few data scientists to work on creating a similar tool, using open source technologies such as CART to replace our proprietary components.

Despite their efforts, the new model did not deliver results as good as those from the pilot PoC. After a year's work, the project was abandoned.

Asset management company needs to know who to market to and how

Lesson:
After the success of a PoC, deciding whether you want to build it internally requires careful consideration. Some aspects to evaluate:

- Is it simply a matter of scaling? Or does it involve re-engineering or substituting major components? If so, how important were these components to the PoC results?

- Who are the people that were involved, and would they still be involved in the expansion? This includes both internal and external members

- There is the benefit of upskilling your internal work force if you choose to build. Ask these questions to make sure they are ready: did they learn sufficiently from the vendor during the collaboration? Can they manage unexpected behavior such as changes in data or metrics?

Summary of Failure Modes

- Mis-sizing the market
- Simultaneous internal development leading to duplicate work
- Swapping out the vendor solution after PoC without considering the change on results

Summary of Ingredients for Success

People
- Stakeholder involvement
- Business-driven translator
- Feedback loop from users

Problem
- Well-scoped problem
- Clearly defined metrics

Data
- Data collection on demand