Vectice White Paper

Evaluating Data Science Infrastructure During M&A Deals

Analyzing the Foundation: Assessing Data Science Infrastructure in Merger and Acquisition Transactions

by Colleen Qiu

Sharing experience from her Data Science Leadership at

metromile SAFEWAY TESLA

vectice

About Colleen Qiu

Colleen Qiu is the Head of Data @ Lindenlab with experience in AI, data science and analytics across various business domains at Fortune 500 firms and tech startups. She was involved in the M&A process of several companies, and summarized her valuable insights in this paper on how to evaluate the data infrastructure of a business.

- In 2002, she joined the **PayPal** Strategy team after **eBay** just acquired the company. Colleen provided PayPal user segmentation data and predictive risk scores to eBay CEO Meg Whitman. This data helped the corporate leadership understand their user communities and needs, and build a post-merge growth strategy for the company.

- In 2010, she joined **Chegg** as they transitioned from textbook rentals to building a college student platform. Chegg acquired **CourseRank** (course selection), **Cramster** (homework help), **Notehall** (notes sharing), **Zinch** (finding scholarship) and **3D3R** (ebook readers). She provided insights into customer lifetime value, market share opportunities and user audiences overlap, based on customer acquisition data and product data.

- In 2021, Colleen was part of the due diligence team for **Metromile** when **Lemonade** announced that it would fully acquire them. The pay-per-mile InsurTech company was acquired in an all-stock transaction. Insurance is built on large data, and Colleen helped showcase their data, the ML models in production and their impact on the business.

Introduction

Mergers & Acquisitions (M&A) are great strategies for companies to expand product and service portfolios. They allow companies to satisfy more customer demands with one-stop shopping and grow their revenue. M&A can also help businesses diversify in their markets. For example, a computer manufacturer might want to purchase an augmented reality startup to offer new camera features on their latest smartphone.

M&A can also help when a company wants to vertically integrate their production or operations. They can acquire smaller competitors or merge with partners to achieve new synergies and economies of scale. This is common for large manufacturers with global supply chains, but vertical integration also happens in software companies.

Evaluating Data Infrastructure in M&A

In this paper, we'll look at how to evaluate the data components of a business. Data and algorithms are becoming incredibly valuable assets. Data science is often embedded in the core product and operations nowadays. In order to properly assess the value of the company, it is important to demonstrate the breadth and depth of the AI solutions and how effective a company can apply AI solutions into the day-to-day business operation.

We'll take the perspectives of Buyers (how to evaluate data infrastructure) and Sellers (how to value your infrastructure when considering a sale). Both parties will have different priorities and valuations, but they'll have to come to a mutual agreement. This balancing act is the main reason why M&As are long and complex undertakings.

Based on my experience, I present the following 5 key areas of the data infrastructure from which all value is derived. How much each area is valued depends on multiple factors, but these are the fundamental areas that define the value. Evaluating will happen during due diligence and is carried out by data science and financial experts.

1. DS/ML Platform And Technical Stack

ML Stack Integration

The most fundamental data component is the ML stack that companies use to build, train, deploy, scale and monitor machine learning systems. An ML stack consists of of tools, frameworks, libraries, and components used to build and deploy machine learning systems.

Building a stack is becoming more accessible with ready-to-use SaaS tools available on the market. SaaS solutions are highly secure and offer economies of scale and ease of deployment. ML teams can also use powerful open-source applications and libraries.

Integration is Key

What matters most is how tightly these components are integrated. **Technical leads from the Buyer need to evaluate how data gets processed from feature stores to algorithms seamlessly in the post-integration world.** This work is typically supported by MLOps and is essential for consistently delivering ML solutions.

One amazing example is Uber's Michelangelo platform, which makes ML operations simpler, faster and more streamlined. Uber had trouble scaling their ML efforts, since engineers worked in different tools on one-off solutions. Michelangelo is their answer:

"Michelangelo is designed to cover the entire ML workflow: manage data, train, evaluate, and deploy models, make and monitor predictions. The system also supports traditional ML models, time series forecasting, and deep learning". (*Uber - Michelangelo ML Platform*)

Platform Consolidation During M&A

Not all companies have the resources to build such a massive in-house platform, but this type of vertical integration is extremely valuable during M&A deals. It shows that the company has invested in infrastructure needed to bring ML models to deployment.

Before a purchase, companies need to decide how they can merge two AI platforms into one. Naturally, there will be some overlap between systems, products and people. Generally, the Seller will lean towards the Buyer during integration. When Metromile was acquired by Lemonade, they needed their AI infrastructure to move towards Lemonade.

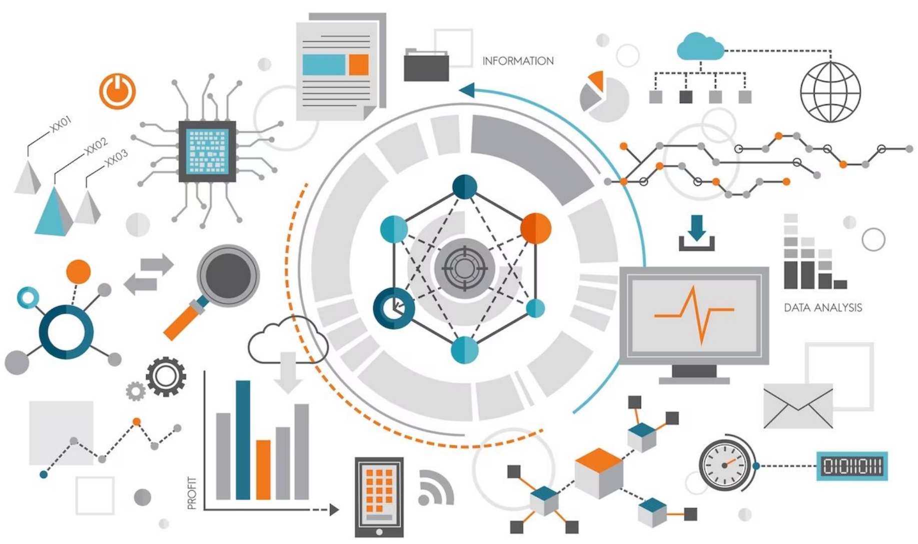
2. Models And Algorithms In Production

Measuring Business Impact

Next, Buyers and Sellers need to evaluate models and algorithms in production. **What matters most is not the number of active models, but the business impact they have.** Ideally, companies already measure business impact, but often they will learn more about their models during the M&A due diligence process. Impact can be measured in terms of ROI, operational efficiencies, or potential losses saved and risk mitigated.

For example, PayPal uses machine learning systems to detect fraudulent transactions.
ML predictions allow PayPal to cut costs from manual investigation, reduce the risk of scams, reduce false positives and negative experiences that lead to customer churn, or avoid the expensive process of recovering and reversing unauthorized transactions.

The PayPal team can quantify how many transactions are going through their model, show false positive rates and compare them to industry standards for fraud detection, and calculate the impact in terms of mitigated losses or increased revenue.



Demonstrating Impact Strategically

Executives should showcase the impact of models without presenting too many details. ML models are valuable intellectual property, so demonstrating the impact based upon metrics while minimizing details about the algorithms can avoid potential conflicts if the deal fails to materialize. This can be done by focusing on performance metrics instead.

Monitoring Models

Monitoring of active models is also extremely important for business governance. ML models are powerful tools, and companies are ultimately responsible for their output. **The Seller wants to demonstrate the tools and techniques used to monitor and re-train models, and show that they are compliant with rigorous privacy and security standards.**

3. Roadmap For Bringing Models To Production

Models in Experimental Phase

Machine learning model development often involves extensive experimentation, with only a fraction of prototypes progressing to production deployment. **Rigorous testing and validation is required before models can be integrated into business processes.** But models that are never deployed cannot have a measurable impact on the business.

Deployment Roadmap and MLOps

Sellers should aim to demonstrate a roadmap for how they are deploying models into the real world. This work is typically supported by MLOps, who should aim to build a standardized ML pipeline for the lifecycle of building, deploying and monitoring models.

Buyers can observe how actively teams publish results in academic papers and on technical blogs, particularly on solving different types of problems. If they frequently publish to internal and external audiences, it shows maturity of the team and suggests that they have a strong roadmap in place for deploying solutions.

Different Lead Times

Executives should keep in mind that engineering and ML teams have different lead times. Some problems (time series prediction) are fairly streamlined, whereas other problems (image recognition or neural networks) need much more lead time. As such, data science teams must be evaluated differently from traditional software teams

4. Quality of Big Data

Showcasing Data Quality

The next critical component is the amount and quality of data the Seller has. Since data science and machine learning are powered by data, the quality and amount of data is critical for building AI solutions. As the saying goes: “Garbage in, garbage out”.

Quality of data can be showcased by producing a data quality report that shows metrics for volume, accuracy, completeness and consistency. Usually, this is based on metadata so the confidential information remains hidden to the Buyer.

Data should be accurate, valid, timely, consistent, complete and unique. Data should also be centralized, so stakeholders and engineers can access them efficiently. What also matters is which sources companies are getting their data from and how frequently they can update data or increase their data collection.

Examples of Valuable Data

For example, Metromile is able to offer pay-per-mile insurance by collecting data from drivers in real-time. This data allows them to offer personalized packages and discount rates for safe drivers. While the drivers generate the data, Metromile collects, stores and analyzes the data, which provides enormous business value.

Feature Stores for Efficiency

Feature Stores are another component that shows investment and commitment from ML teams. Feature Stores are centralized repositories with labeled and structured features, which streamlines the process of selecting features and training the models. Teams using Feature Stores will be more successful in bringing models to production.

5. Talent & Team Culture

Team Culture and Collaboration

Having a strong team culture is essential for delivering business solutions. **Buyers want to look for teams that foster collaboration, value ownership and allow for safe experimentation and failure.**

A strong sign of a productive team culture is when teams are actively publishing results in the data science community. It suggests an ability to build solutions and willingness to share outcomes. Publications also showcase the quality of solutions and performance of the algorithms and models they are building.

Acqui-hires for Talent

Sometimes M&A happens solely to acquire talents. Talent evaluation during acqui-hires is based on tenure, skills, background and experience. For example, smaller teams with strong talent but a limited product and market share may be acquired and absorbed into a bigger company.

Integrating Teams

During a merger, companies need to decide how teams will be incorporated into the new organizational structure. With talent in high demand, companies will want to retain all data scientists.

They might want to move from a centralized to an embedded team structure, where smaller teams work closely together with specialized departments.

For example, Chegg has clearly defined teams that each serve a unique role in the company's strategy. Their data science and analytics team can be embedded to serve requests across all departments, while still remaining a separate division.

Conclusion

Mergers & Acquisitions are great strategies for companies to expand their product and service portfolios. M&A's create synergies and economies of scale that are otherwise not possible to achieve. During the due diligence process, Buyers and Sellers will need to evaluate and agree on how much their business is worth.

Since data science and machine learning are relatively new fields, many executives find it hard to evaluate these areas. **They cannot be evaluated using the same metrics as traditional software teams.** However, they have become some of the most valuable components of a business.

In this paper, I presented the 5 key areas from which all value is derived. First, the AI infrastructure and system integration is what creates a solid production pipeline. Next, what matters is the active models and their business impact. A roadmap should also be in place that supports building and deploying these models. Furthermore, the amount and quality of data should be analyzed, since all models are trained on this data. Lastly, talent and team culture can contribute tremendous value to an M&A deal.

In the next decades, data science will remain a key differentiator. Companies who are considering acquiring or merging with data science companies can use these 5 areas to evaluate what teams have to offer and how much they are willing to pay for it.

Introducing vectice - Deliver AI you can trust, faster

When it comes to M&A: Documented AI models = Valuable Models

Leveraging Vectice in the case of a M&A due diligence ensures:

- Centralized system of record encompassing models, datasets, code - only fully documented AI models hold true value.
- Consistent documentation, making it easy to evaluate asset quality and production readiness. In M&A, value come from the thoroughness and precision of documentation.
- Governance features like approval workflows and progress dashboards let buyers assess development processes.
- Searchability enables sellers to quickly showcase all documented models and team expertise.
- Templates and best practices that allow sellers to showcase they follow industry standards.

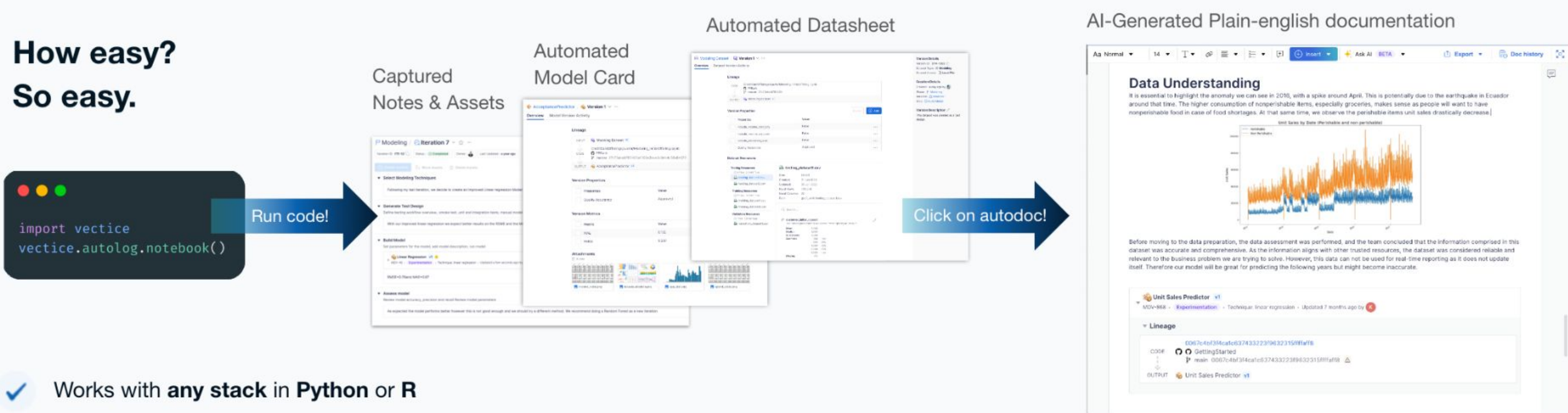
By auto-documenting assets and centralizing knowledge, Vectice simplifies data science infrastructure during M&A deals for both buyers and sellers.

No matter your stack, get started with vectice today!

Instant acceleration of your analysis/model to production

- ✓ Most automated documentation solution
- ✓ Optimized for data science & AI workflows: interactive, exploratory

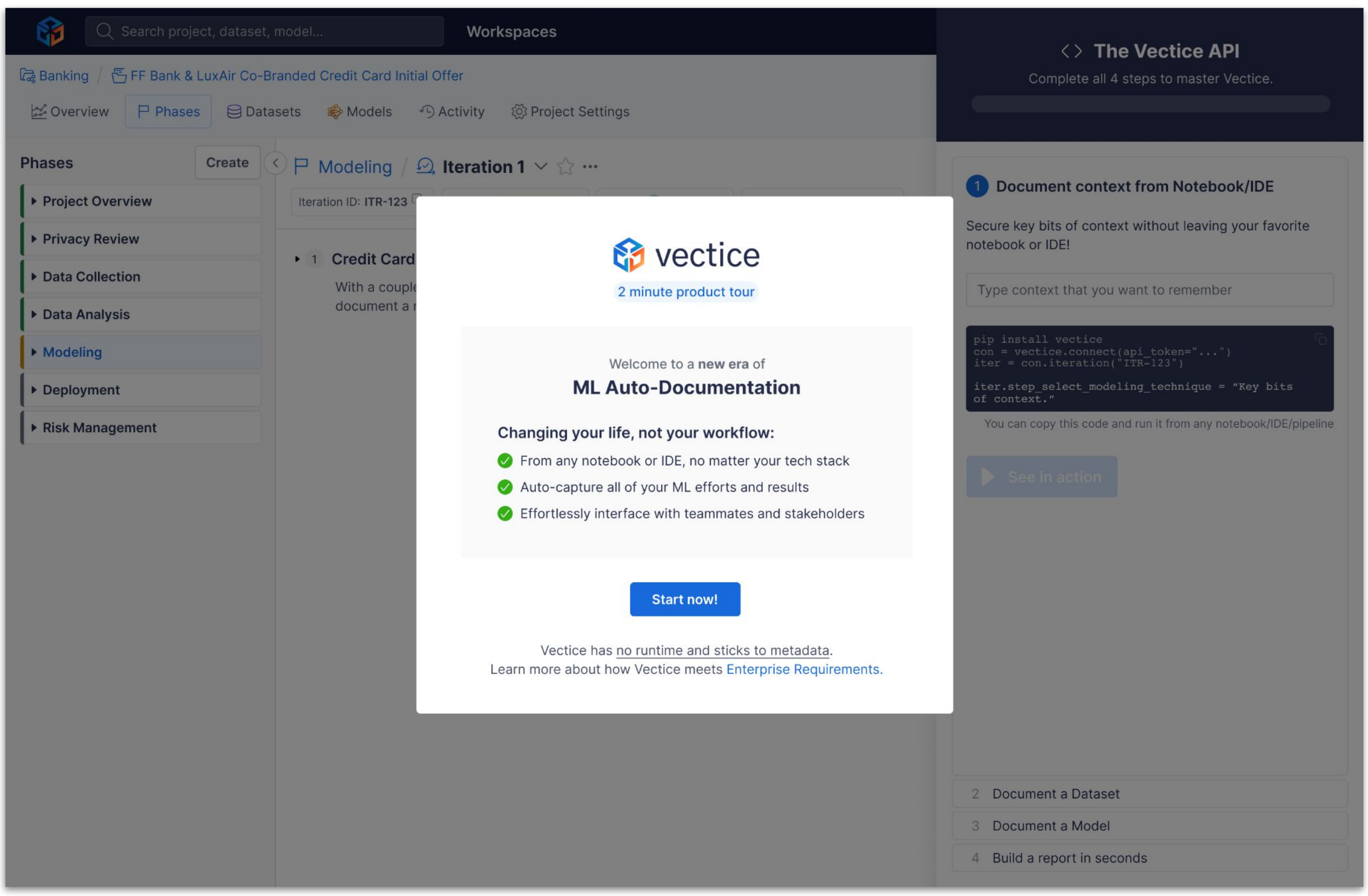
How easy?
So easy.



- ✓ Works with **any stack** in **Python** or **R**
- ✓ Including **MLflow**, **VertexAI**, **Comet ML** to document current & past work

With Vectice, your modeling team can effortlessly attain the documentation standards required for maintainability, explainability, and compliance with the ever-evolving AI regulatory landscape – all without the burden of error-prone manual labor.

With Vectice's unique technical and cross functional Auto-Documentation capabilities, you will accelerate your ML projects, delivering value faster with peace of mind.



Try Vectice for free - www.vectice.com

Learn more about Vectice and how it can support you

785 Market Street, Suite 700, San Francisco CA 94103 - United States

contact@vectice.com (650) 399 - 0114