

Measuring Success of Machine Learning Products

Business Performance vs Model Performance



Wilhem Kornhauser · [Follow](#)

Published in Towards Data Science · 4 min read · Jan 4, 2021



Failure is part of the learning process. Unfortunately, it frequents being part of the machine learning development process far too often. ML projects can be doomed from conception due to a misalignment between product metrics and model metrics. Today, there are many skilled individuals who can create highly accurate models, and poor modelling capabilities is not a common pitfall. Rather, there is a tendency for accurate models to be developed that are not useful for a product, thus failing to meet business objectives.

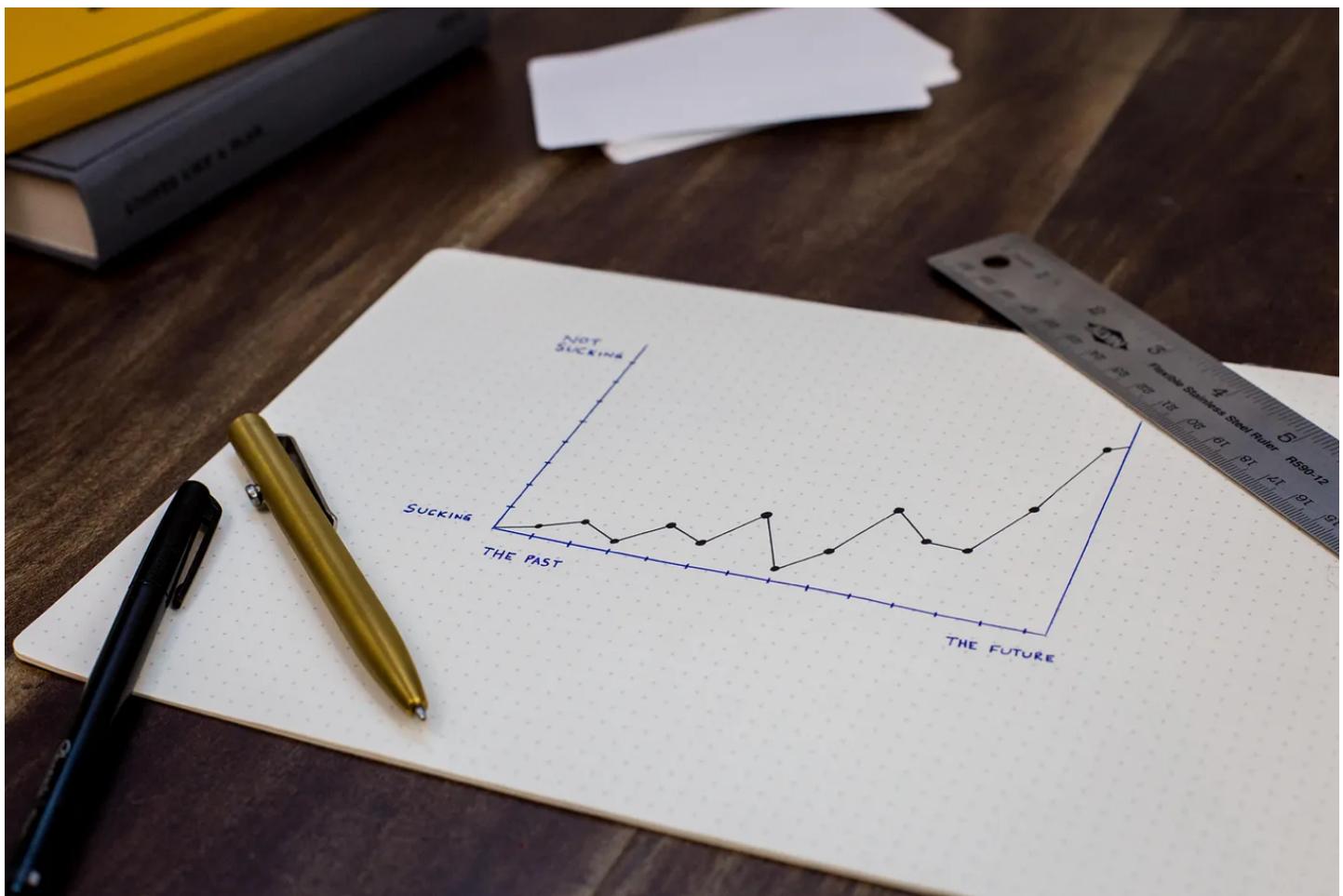


Image by [Isaac Smith](#) on [Unsplash](#)

Medium



Search



Write

Sign up

Sign in



Defining and Quantifying Success

In defining success, it is important to consider the differences between *business performance* and *model performance*. The most simple way to put this is that business performance is a function of many variables, not just model performance. With poor model performance, business performance will be inadequate, but good model performance does not guarantee good business performance!





Business Performance

To evaluate business performance, it is necessary to start with a product or feature goal. For example, increasing revenue of an e-commerce site. Once this is defined, a *product metric* should be assigned to evaluate success. This metric needs to be separate from any model metrics, only quantifying the product's success. Product metrics can vary, with metrics such as number of users a feature attracts or the click-through rate (CTR) of recommendations both representing valid examples.

At the end of the day (and fiscal period), product metrics are what matter. They represent the goals of the product. Any other metrics are simply to be considered tools available to optimize product metrics. Typically projects only aim to improve a single product metric, but their impact is frequently quantified with respect to numerous metrics. Some of these include *guardrail metrics*, which represent metrics that are not to fall below a certain point. For example, an ML project can have the objective of increasing a product metric like number of users while maintaining the stability of other metrics like average user session.

Model Performance

Measuring the effectiveness of an ML approach requires the tracking of model performance. Prior to deployment of a product utilizing ML, it is not possible to quantify product metrics. During the building of the ML product, *offline metrics* or *model metrics* are useful for defining success. To consider an offline metric to be of quality, it should be evaluated without

exposing the ML model to users. Furthermore, it should be highly correlated with product metrics and business goals.

Suppose you were developing a feature to offer suggestions to users while typing queries to an online retail store. The success of this feature can be measured using CTR (product metric). To create these suggestions, a model which predicts the words a user will type and displays these predictions can be built. By measuring the word-level accuracy (calculating how often the model predicts the correct next set of words), the performance of the model can be defined. In this scenario, the model would be required to have extreme accuracy to increase the product's CTR, as a single error in word prediction would render a suggestion useless.

A (better) approach would involve training a model that can take user input and perform classification into categories of your catalog, suggesting the top most likely predicted categories. Here, the number of categories in a catalog is significantly less than all the words in the English language, making this a much easier model metric to optimize. Moreover, the model only needs to correctly predict one category to generate a click.

Closing Remarks

Business performance seems to often get lost in the hype of ML model performance, but it is always important to realize what metrics need to be optimized.

Beyond just business performance and model performance, other metrics for success are required. Considerations include model freshness and speed. Model freshness is an important consideration as models age and data distributions change. Speed is always an important consideration with any software. Autonomous vehicles could never be achieved if it took even several seconds to process data and generate predictions.

For more on this topic, I highly recommend *Building Machine Learning Powered Applications* by Emmanuel Ameisen. It covers the skills necessary to design, build, and deploy applications powered by machine learning.

Machine Learning

Business

AI

Product Design

Software Development



Written by Wilhem Kornhauser

33 Followers · Writer for Towards Data Science

AI Engineer | Python Programmer| Engineering Graduate | Connect on LinkedIn!

<https://www.linkedin.com/in/wilhemkornhauser/>

Follow



More from Wilhem Kornhauser and Towards Data Science



 Wilhem Kornhauser in Towards Data Science

Algorithms, Data Structures, and Data Science

Why should you care about algorithms and data structures as a data scientist?

Dec 30, 2020  169  3

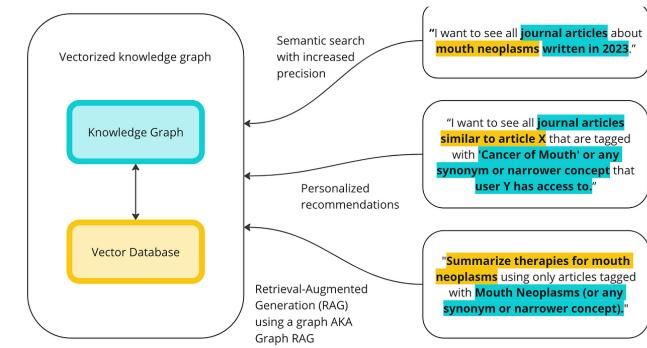


 Ahmed Besbes in Towards Data Science

What Nobody Tells You About RAGs

A deep dive into why RAG doesn't always work as expected: an overview of the...

 Aug 23  1.7K  25



 Steve Hedden in Towards Data Science

How to Implement Graph RAG Using Knowledge Graphs and...

A Step-by-Step Tutorial on Implementing Retrieval-Augmented Generation (RAG),...

Sep 6  1.2K  13



 Wilhem Kornhauser in CodeX

Algorithms, Data Structures, and Data Science

Why should you care about algorithms and data structures as a data scientist?

Jan 5, 2021  6



[See all from Wilhem Kornhauser](#)[See all from Towards Data Science](#)

Recommended from Medium



 Aparna Dhinakaran in Towards Data Science

Choosing Between LLM Agent Frameworks

The tradeoffs between building bespoke code-based agents and the major agent...

Sep 20  2K  21



Lists

AMAZON.COM
Software Development Engineer Seattle, WA
Mar. 2020 – May 2021

- Developed Amazon checkout and payment services to handle traffic of 10 Million daily global transactions
- Integrated Iframes for credit cards and bank accounts to secure 80% of all consumer traffic and prevent CSRF, cross-site scripting, and cookie-jacking
- Led Your Transactions implementation for JavaScript front-end framework to showcase consumer transactions and reduce call center costs by \$25 Million
- Recovered Saudi Arabia checkout failure impacting 4000+ customers due to incorrect GET form redirection

Projects

NinjaPrep.io (React)

- Platform to offer coding problem practice with built in code editor and written + video solutions in React
- Utilized Nginx to reverse proxy IP address on Digital Ocean hosts
- Developed using Styled-Components for 95% CSS styling to ensure proper CSS scoping
- Implemented Docker with Seccomp to safely run user submitted code with < 2.2s runtime

HeatMap (JavaScript)

- Visualized Google Takeout location data of location history using Google Maps API and Google Maps heatmap code with React
- Included local file system storage to reliably handle 5mb of location history data
- Implemented Express to include routing between pages and jQuery to parse Google Map and implement heatmap overlay

 Alexander Nguyen in Level Up Coding

The resume that got a software engineer a \$300,000 job at...

1-page. Well-formatted.

 Jun 1  23K  455





Predictive Modeling w/ Python

20 stories · 1561 saves



The New Chatbots: ChatGPT, Bard, and Beyond

12 stories · 467 saves



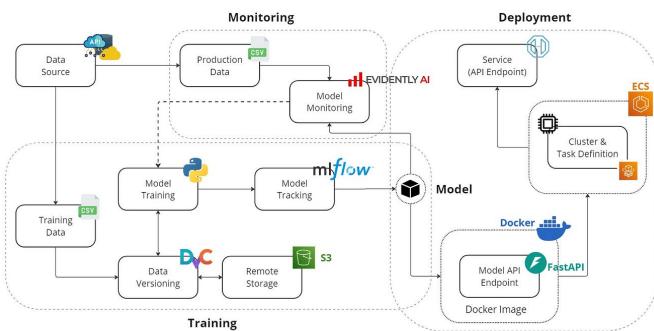
Natural Language Processing

1728 stories · 1302 saves



Practical Guides to Machine Learning

10 stories · 1894 saves

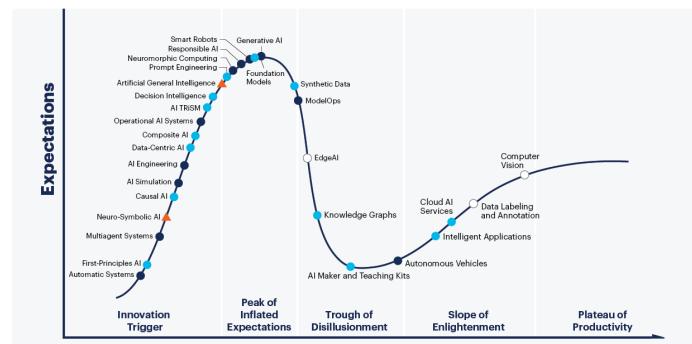


Prasad Mahamulkar

Machine Learning Operations (MLOps) For Beginners

End-to-end Project Implementation

Aug 29 1.1K 10



Vishal Rajput in AIGuys

Why GEN AI Boom Is Fading And What's Next?

Every technology has its hype and cool down period.

Sep 4 1.7K 56

RAG or RAG Fusion?



RAG

Struggling to succeed



RAG Fusion

Driving success upward



Surya Maddula in Towards AI

Not RAG, but RAG Fusion? Understanding Next-Gen Info...

AI and Natural Language Processing are advancing at an incredible pace, and now...

Sep 22 668 8

Analysing Tech Layoffs: Which Roles Were Hit Hardest?

In the aftermath of tech layoffs, it would appear that engineering was hit hardest, bu...

Sep 17 819 16

[See more recommendations](#)

[Help](#) [Status](#) [About](#) [Careers](#) [Press](#) [Blog](#) [Privacy](#) [Terms](#) [Text to speech](#) [Teams](#)