Why Great Models Fail

LESSONS FROM 9 YEARS OF DEPLOYING ML MODELS

MAY 2026

Megan Robertson

NDC TORONTO

MEGAN ROBERTSON

- Chief Machine Learning Engineer - MegRob DSA
- 9+ years experience in data science and machine learning
    - Building products to support stakeholder and business strategy
    - Intersection between technical and non-technical teams

“FAILURES” ARE A LEARNING OPPORTUNITY

HOW DO I KNOW THIS?

SOCIAL MEDIA ANALYSIS

What are the target customers talking about?

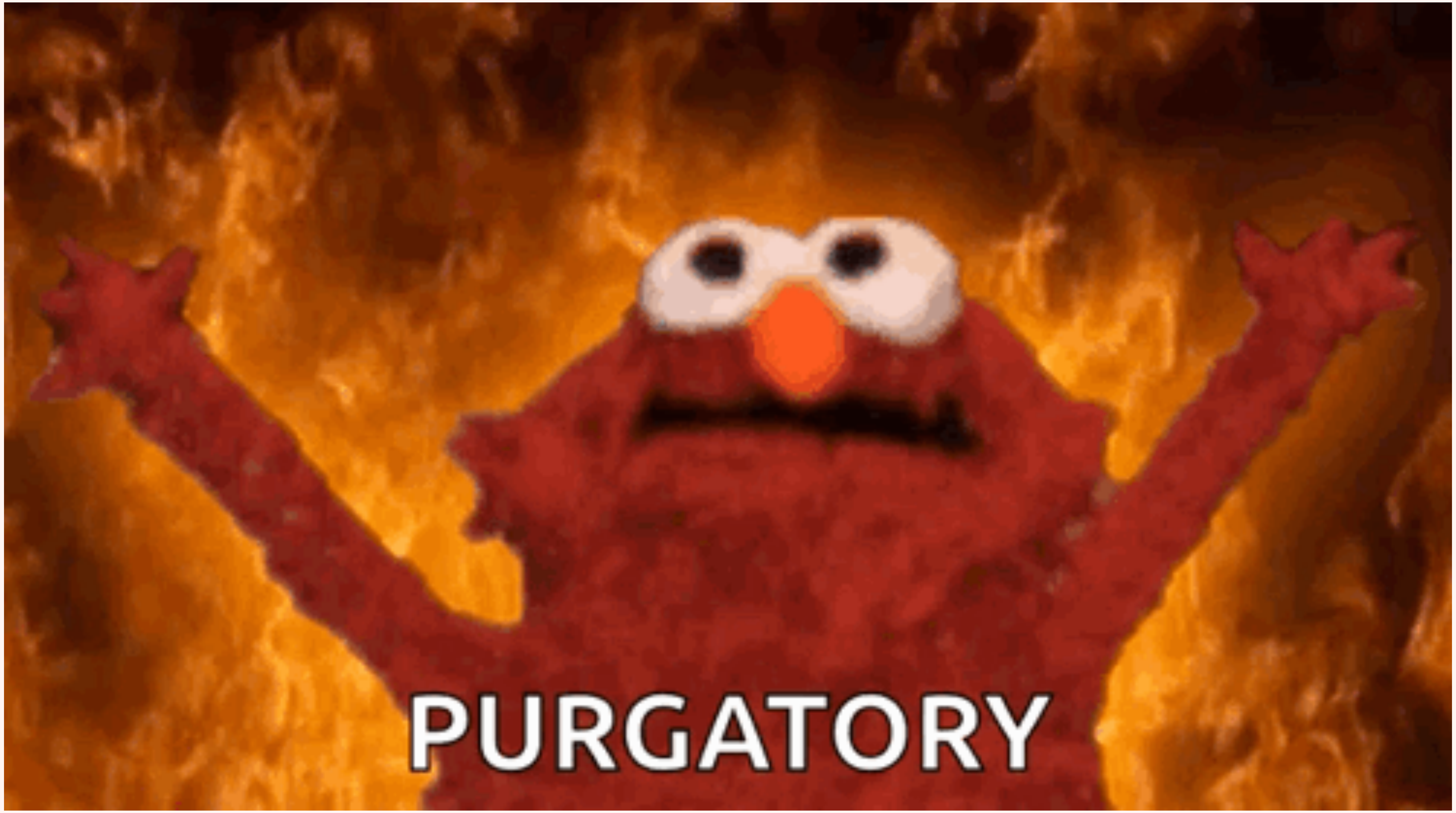
What are the trending topics of conversations?

SOCIAL MEDIA ANALYSIS

SHORTCOMINGS

Minimal Stakeholder Consultation

Expensive Data

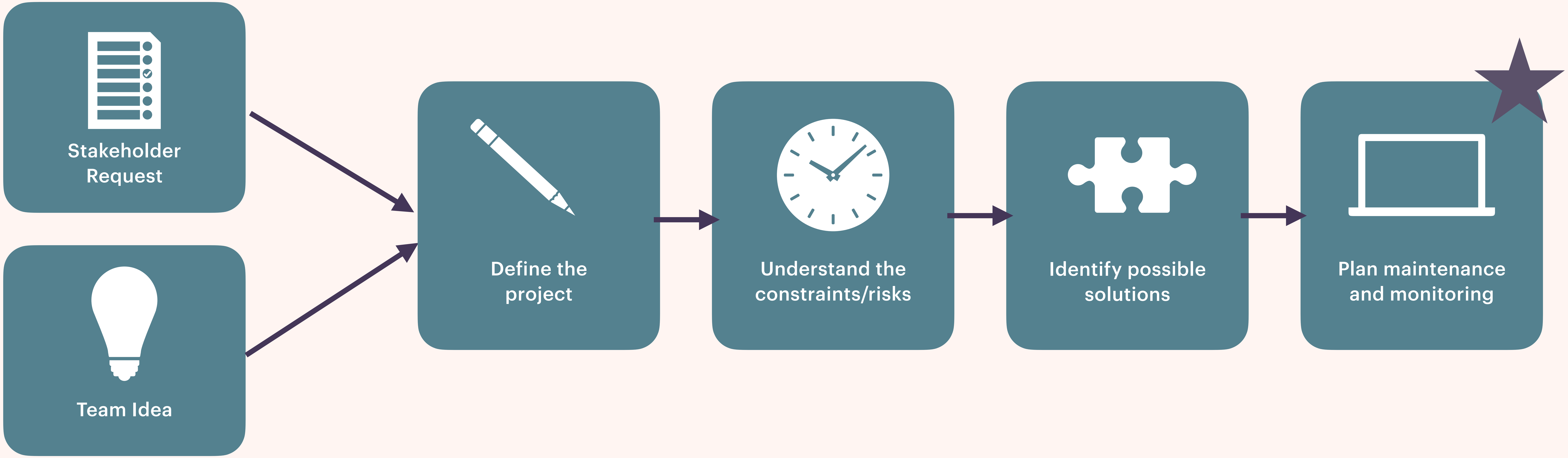

PURGATORY

WHY DO MODELS FAIL?

1. Scope of work not fully defined
2. The world changes over time

PROPER SCOPE IS THE FOUNDATION OF A SUCCESSFUL PROJECT

SCOPING A PROJECT



What is your value add?

DEFINE THE PROJECT

Identify the stakeholders and end users

Identify the problem or pain point

Define success

IDENTIFY STAKEHOLDERS AND END USERS

- Who will act on the predictions from the model?
- Who owns the decisions that are informed by the model?
- What workflows does this model touch/affect?

IDENTIFY PROBLEM TO SOLVE

- How does this ask help you?
- What do you wish you knew about your customer?
- What manual or semi-manual tasks takes up a lot of time?
- Do you feel like you’re guessing anywhere?

DEFINE SUCCESS

We want better marketing campaigns.	Unsubscribe rate, bounce rate, conversion rate
We need to keep users engaged in the app.	Daily active users, churn rate
We need more users to like our content.	Click-through rate (CTR), time on page

COMMON PITFALLS

Building what is cool to you

Evaluating on vibes

There isn't a stakeholder to use the model outputs.

The contribution of the model does not outweigh the costs of its maintenance.

You are unable to quantify the contributions of your model.

You cannot justify the cost of maintaining your model.

UNDERSTAND CONSTRAINTS AND RISKS

Identify available budget and bandwidth

Brainstorm effects of bad predictions or recommendations

IDENTIFY BUDGET AND BANDWIDTH

- What is the expected compute power needed?
- Is there a cost to acquire any additional data?
- How long will it take to build the model?
- What are the team’s other priorities and how does this project compare?

BRAINSTORM BAD OUTCOMES

- What population is harmed by a bad prediction?
- When is it “ok” for the model to be wrong? (Type I vs. Type II errors)
- What are the risks of an incorrect prediction - monetary, life/death, etc?

COMMON PITFALLS

Ignoring budget and bandwidth

Underestimating risks

The cost of the best performing model exceeds the computing resources available.

The team doesn't have the capacity to scale this into production.

A small mistake in the model can cause big harm to the company or customers.

The risk is not worth the model payoff.

IDENTIFY POSSIBLE SOLUTIONS

Understand existing capabilities and solutions

Define the minimum viable product

UNDERSTAND EXISTING CAPABILITIES

- Has someone tried to solve this problem before?
- What can you learn from previous attempts?
- Is there an existing tool that you can adapt?

DEFINE THE MVP

- What is the most basic solution?
- What would convince stakeholders this is working?
- How will the output be consumed?

COMMON PITFALLS

Reinventing the wheel

You waste resources and time without providing any value above existing projects.

Overcomplicated solutions

It takes too long to build a model and launch it into production. You introduce risk and cost.

MAINTENANCE AND MONITORING

Know where the model pipeline can fail

Plan robust monitoring system

Strategy for fixing failures

IDENTIFY FAILURE POINTS

Data Ingestion

- Broken pipelines
- Schema changes
- Duplicates
- Delayed jobs

Feature Generation

- Feature drift
- Timeout errors

Model Training

- Data leakage
- Overfitting

Predictions and Inference

- Model drift
- Corrupt files

Serving

- Compute availability
- Latency issues

Dependency failure

ROBUST MONITORING SYSTEM

Data Issues

- Summary metrics
- Distribution changes (Chi-square, Kolmogrov-Smirnoff)
- Missing values

Model Issues

- MAE, MSE, AUC
- Precision, recall, F1 score
- Semantic similarity, prompt injections

Infrastructure

- Timeout rate
- Latency summary
- SLA comparison

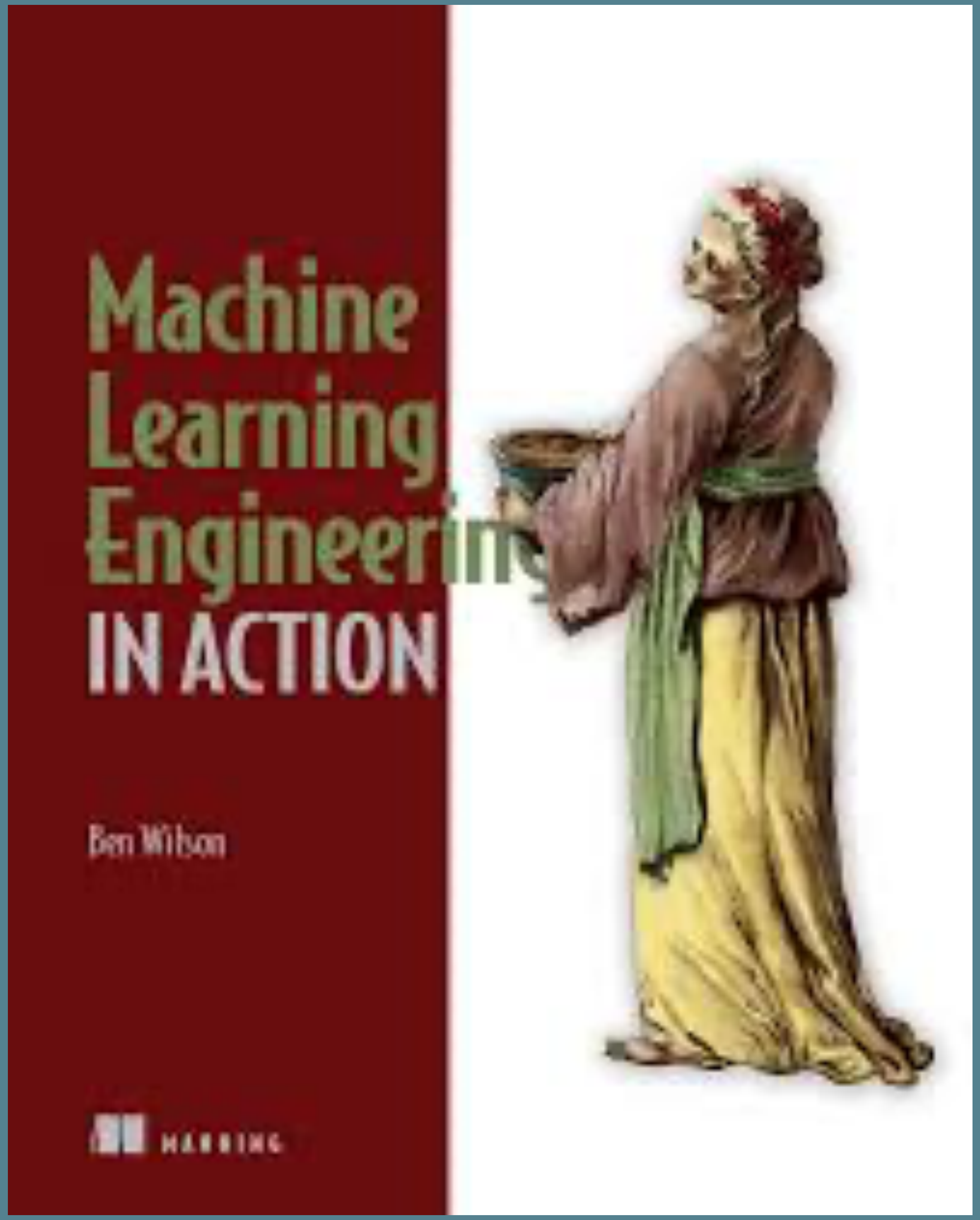
COMMON PITFALLS

Lack of adequate monitoring

You are unable to catch any error before it grows into a bigger problem. You lose stakeholder or customer trust in your team and model.

Ignoring model drift

Data changes over time - especially if you are using the model for intervention. The model only represents one point in time.



"MODELS WILL DRIFT. THERE IS NO SUCH THING AS STATIC IMPLEMENTATION."

WHY DO MODELS DRIFT?

Concept: Divergence between input and target variables

Data: Change in underlying distributions of features

Label: Distribution of target variables change over time

Reality:

ADDRESSING MODEL DRIFT

- Feature engineering
- Retrain model
- New algorithm
- Retire the model

KEY TAKEAWAYS

- A highly predictive model is not enough to succeed in industry.
- Stakeholders and subject matter experts should play a key role in project scope.
- A model represents a point in time, and the world changes over time.

MEGAN@MEGROBDATASCIENCE.COM

MEGROBDATASCIENCE.COM