Does a well-trained model always produce a strong performance in a live setting? Discuss plausible scenarios when a well-trained high-performant model actually underperforms in a production environment. What does that say about monitoring your model after training and through time?

No, just because you have a well-trained model does not mean it will perform well in a live setting and does not even mean it is a good model at all. There are a few reasons why this might occur for any given model.

1. Model was trained on bad data.

This is a very common occurrence and is why it is paramount to give your models good, reliable, and diverse data that is representative of true world populations you will want to employ the model on. Data scientists are said to spend the vast majority of their time in data collection and cleaning because your model can only be as good as your data. Michael Lones does a great job at explaining this in his article, “Why Doesn’t My Model Work?” He explains, “In the worst cases, misleading data can cause the phenomenon known as garbage in garbage out; that is, you can train a model, and potentially get very good performance on the test set, but the model has no real world utility” (Lones). He goes on to explain this was common over the COVID19 pandemic. Everyone was rushing to get models produced and it was later discovered that some of the publicly available data sources included overlapping records, mislabels, etc. and many of these models failed to produce valuable insight in real world applications. One example he gives is COVID19 imaging where hidden variables such as orientation of the person being imaged could impact the model (bias towards those lying down having covid because they are not able to sit up they are so sick). Computer vision models may use pixels in the background to classify images, even though the background is irrelevant in real life (for most problems). Other issues could include skewed data, where there is a misrepresentation of each class in the training set that does not represent real world ratios. Another common mistake made is overtraining and overfitting the model. Sometimes models are trained too much, and they can memorize the training data, and therefore are bad at inferring on never-before-seen data. This is common with smaller datasets, as it is hard to memorize millions of instances.

1. Data Leakage

Lones goes on to explain that data leakages can be the cause of a bad model performance. He explains, “This happens when the model training pipeline has access to information it shouldn’t have access to, particularly information that confers and advantage to the model. Most of the time, this manifests as information leaks from the test data…” (Lones). One example of how this might happen is preprocessing all the data together before a train – test split. If we were to perform feature selection, dimensionality reduction, scaling, etc. on the whole data set and then split, the train data will have parts (for lack of a better term) of the test data bias in it. There is knowledge of the test data present when training the model which can lead to a biased model that will perform badly on real world data. In time series and forecasting, if there are training examples that are ahead of some testing points (in time), this could cause the model to artificially improve results on the testing set because it trained on data on each end of that point in time (the model can almost fill in the blanks without predicting much). Lastly, tweaking models to perform well on the same testing set often leads to overfitting and artificially optimistic results. When making tweaks to a model, the training should start over and the training and testing data should be shuffled to avoid this happening. As models get more complicated, such as deep learning, data leakage becomes more common and a harder problem to track and fix in these black box settings.

1. The last main point Micheal Lones talks about is picking the wrong metric for evaluation. This point relates nicely to discussion 1 of this week where we talk about some of our favorite metrics. It is important to remember that not all metrics are useful for every model, and it is important to understand the context of the problem you are attempting to solve and what the model is showing you. For example, accuracy alone might not be a great metric for predicting cancer patients. If cancer is rare and is only in .1% of the population, you could easily get a 99% accuracy on a model by predicting no cancer for every person, but in practice this model would be bad at solving its problem. Conversely, if we predicted each patient to have cancer, then we have terrible accuracy, but a 100% true positive rate, so we would not miss any diagnoses. Context and the nature of the problem matters, and you must evaluate your model on appropriate metrics. It is easy to make any model look good by cherry picking certain metrics. Lones also explains, “Another problem is assuming that a single evaluation is sufficient to measure the performance of a model. Sometimes it is, but a lot of the time you’ll be working with models that are stochastic or unstable” (Lones). Oftentimes you need to evaluate multiple metrics that evaluate different areas of your model. Just relying on accuracy will probably get you into trouble eventually.

Lastly, the world changes. Models need constant updating and retraining. What might be true today in the world might not hold true tomorrow. Consumer patterns, demands, markets, resource availability, technology, etc. is always changing in the real world so it is wrong to think one model trained one time is going to hold its weight forever in predicting real world data.

Another article by Sahin Ahmed (medium.com) explains that models also can outdate themselves quickly. Models that are trained on consumer data can quickly become poor if consumer patterns shift (which is not uncommon). This is called data drift. His paper walks through several other “drift” phenomena such as feature drift, prediction drift, and outdated training data. Since this discussion is already a little long, I recommend you check out his paper for some new insights!

[Why Doesn’t My Model Work?](https://thegradient.pub/why-doesnt-my-model-work/)

[Why does machine learning model performance degrade, and how can we detect and prevent it? | by Sahin Ahmed, Data Scientist | Medium](https://medium.com/@sahin.samia/why-does-machine-learning-model-performance-degrade-and-how-can-we-detect-and-prevent-it-70f546a54548)

**Discussion 3** - Give examples of when a well-trained model may not end up benefiting a business problem. Likewise, there are situations when a weak model may provide a huge benefit or boost to the business. Envision and discuss an example of when these situations might manifest in an industry of your choice.

Model building is great and oftentimes leads to better business insights. However, there are situations where a well-trained model might not benefit a problem. Below I will go over some different scenarios where this may be the case. I am going to use sports for an example industry.

1. Insufficient Capital

Some models are great at identifying potential solutions to business problems, but those solutions may not be feasible. For example, if a pro hockey team creates a model to identify the best players in the league using advanced stats, metrics, movements, etc. of the players. The model is developed and performs great on historical data. The model is then used for current players and the model gives you a list of 5 players that can instantly make your team 20% better: Connor McDavid, Nathan MacKinnon, Cale Makar, Leon Draisaitl, and Connor Hellebuyck. This is great insight; however these players are widely considered the best at their positions. It would be great to add these players to the team, but it would be practically impossible to acquire these players via trade. Our team does not have the capital to enact that insight. If any of those players reach free agency, we most likely have to overpay them based on true market value and if we are a small market team we may not even have the funds to do so. This can be seen in baseball a lot. Any team would kill to have Juan Soto on their team, but only the Mets were willing to pay him 700 million dollars on his contract. Some teams like the Athletics and the Marlins don’t have enough money to even consider something like this.

Meg Rivera goes over 7 reasons your business model might fail in her article from thekickassentrepreneur.com. Reason number 5 is “insufficient capital in hand”.

1. Another way in which a great model may not end up benefitting a business problem is when the model gives a solution but not a direct action. It is easy to identify problems, but harder to find ways to fix or improve these problems. If a team is trying to identify ways to increase revenue it might find that the ticket sales/revenue is down, and the team should focus efforts on that area. However, if the stadium has been at 70% capacity for years this might be something the team is already aware of. The team now has to decide how to market better, what demographics to go after, and other actions such as dynamic pricing to increase sales. This is not to say the first model is not useful as it can serve as a starting point for further action. However, the first model did not help the problem as much as one would hope in a perfect world. It identified a problem which created a number of new questions (this is common in data science).

Conversely, there are situations where a weak model can help a business problem a lot.

An article by the chicagobooth.edu explains that many real world economic models rely on weak signals. “Weak signals are prevalent in economic data. For example, changes to personal income, the unemployment rate, or corporate bond spreads are not seemingly relevant to someone trying to predict a move in industrial production. But such data could be helpful in combination… after all, personal income changes are tied to consumer demand. Corporate bond spreads signal shifts in business borrowing costs…” (Monika Brown). The paper goes on to explain that sometimes models with more weak predictors can be more powerful than one that focuses too much on one or two powerful predictors. Your weaker model might uncover insights that were not thought of before. Even if a weak business model for a large company shows insights in areas the company can cut costs by 2-3%, this could result in millions of dollars saved across the whole company. It is hard to come up with a very concrete example of this.

To bring it back to sports, lets say a model shows a weak correlation between people under 20 years old and lower ticket sales. This weak correlation probably is not because that age group does not like sports, but maybe because they cannot afford the high-priced games on the weekend. A business could look into this weak predictor to attempt to enact larger change across their customer base.

[Turning Weak Signals into Strong Predictions | Chicago Booth Review](https://www.chicagobooth.edu/review/turning-weak-signals-into-strong-predictions)

[7 Reasons Your Business Model Might Fail | The Kickass Entrepreneur](https://thekickassentrepreneur.com/reasons-business-model-might-fail/)

[Common Reasons For The Failure Of Business Models | Founder's Guide](https://foundersguide.com/common-reasons-for-the-failure-of-business-models/)