



Understanding the Bias-Variance Tradeoff

Seema Singh

May 21, 2018 · 4 min read



Whenever we discuss model prediction, it's important to understand prediction errors (bias and variance). There is a tradeoff between a model's ability to minimize bias and variance. Gaining a proper understanding of these errors would help us not only to build accurate models but also to avoid the mistake of overfitting and underfitting.

So let's start with the basics and see how they make difference to our machine learning Models.

What is bias?

Bias is the difference between the average prediction of our model and the correct value which we are trying to predict. Model with high bias pays very little attention to the training data and oversimplifies the model. It always leads to high error on training and test data.

What is variance?

Variance is the variability of model prediction for a given data point or a value which tells us spread of our data. Model with high variance pays a lot of attention to training data and does not generalize on the data which it hasn't seen before. As a result, such models perform very well on training data but has high error rates on test data.

Mathematically

Let the variable we are trying to predict as Y and other covariates as X. We assume there is a relationship between the two such that

$$Y=f(X) + e$$

Where e is the error term and it's normally distributed with a mean of 0.

We will make a model $\hat{f}(X)$ of $f(X)$ using linear regression or any other modeling technique.

So the expected squared error at a point x is

$$Err(x) = E \left[\left(Y - \hat{f}(x) \right)^2 \right]$$

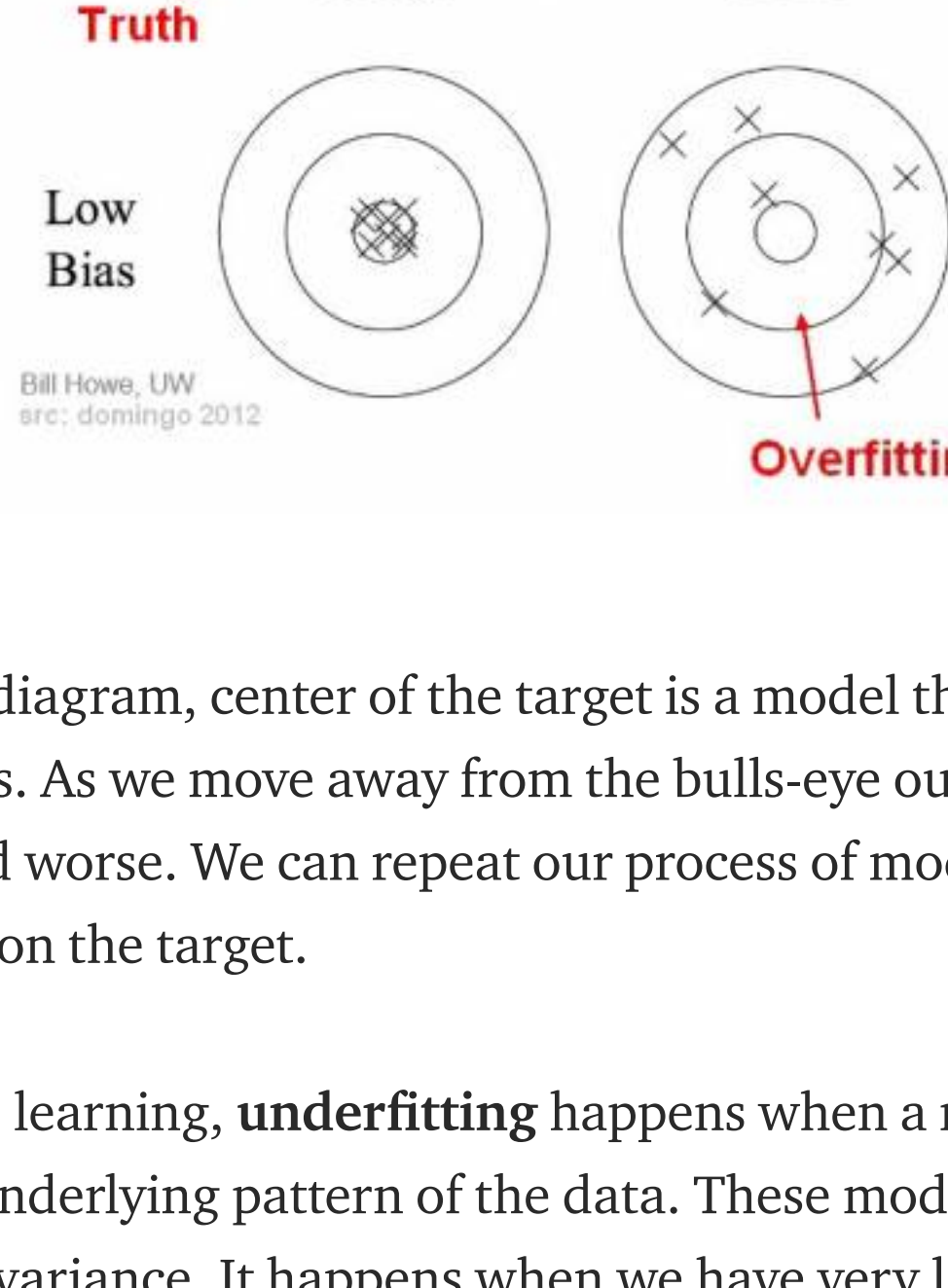
The $Err(x)$ can be further decomposed as

$$Err(x) = \left(E[\hat{f}(x)] - f(x) \right)^2 + E \left[\left(\hat{f}(x) - E[\hat{f}(x)] \right)^2 \right] + \sigma_e^2$$
$$Err(x) = Bias^2 + Variance + Irreducible Error$$

$Err(x)$ is the sum of $Bias^2$, variance and the irreducible error.

Irreducible error is the error that can't be reduced by creating good models. It is a measure of the amount of noise in our data. Here it is important to understand that no matter how good we make our model, our data will have certain amount of noise or irreducible error that can not be removed.

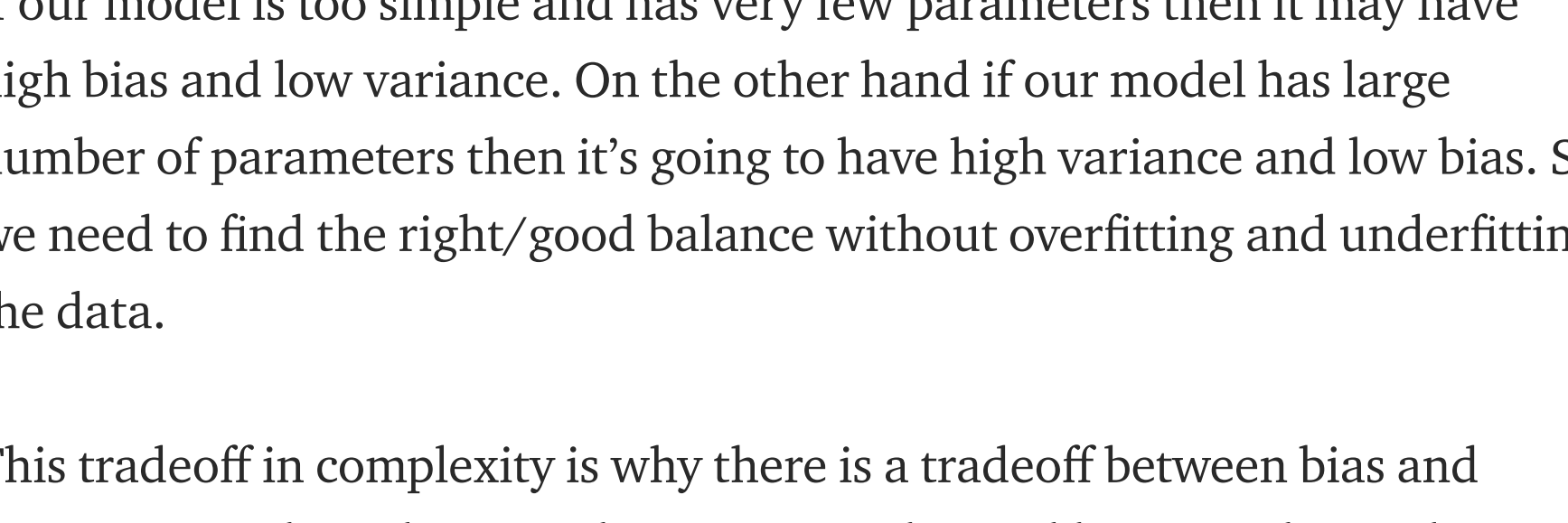
Bias and variance using bulls-eye diagram



In the above diagram, center of the target is a model that perfectly predicts correct values. As we move away from the bulls-eye our predictions become get worse and worse. We can repeat our process of model building to get separate hits on the target.

In supervised learning, **underfitting** happens when a model unable to capture the underlying pattern of the data. These models usually have high bias and low variance. It happens when we have very less amount of data to build an accurate model or when we try to build a linear model with a nonlinear data. Also, these kind of models are very simple to capture the complex patterns in data like Linear and logistic regression.

In supervised learning, **overfitting** happens when our model captures the noise along with the underlying pattern in data. It happens when we train our model a lot over noisy dataset. These models have low bias and high variance. These models are very complex like Decision trees which are prone to overfitting.



Why is Bias Variance Tradeoff?

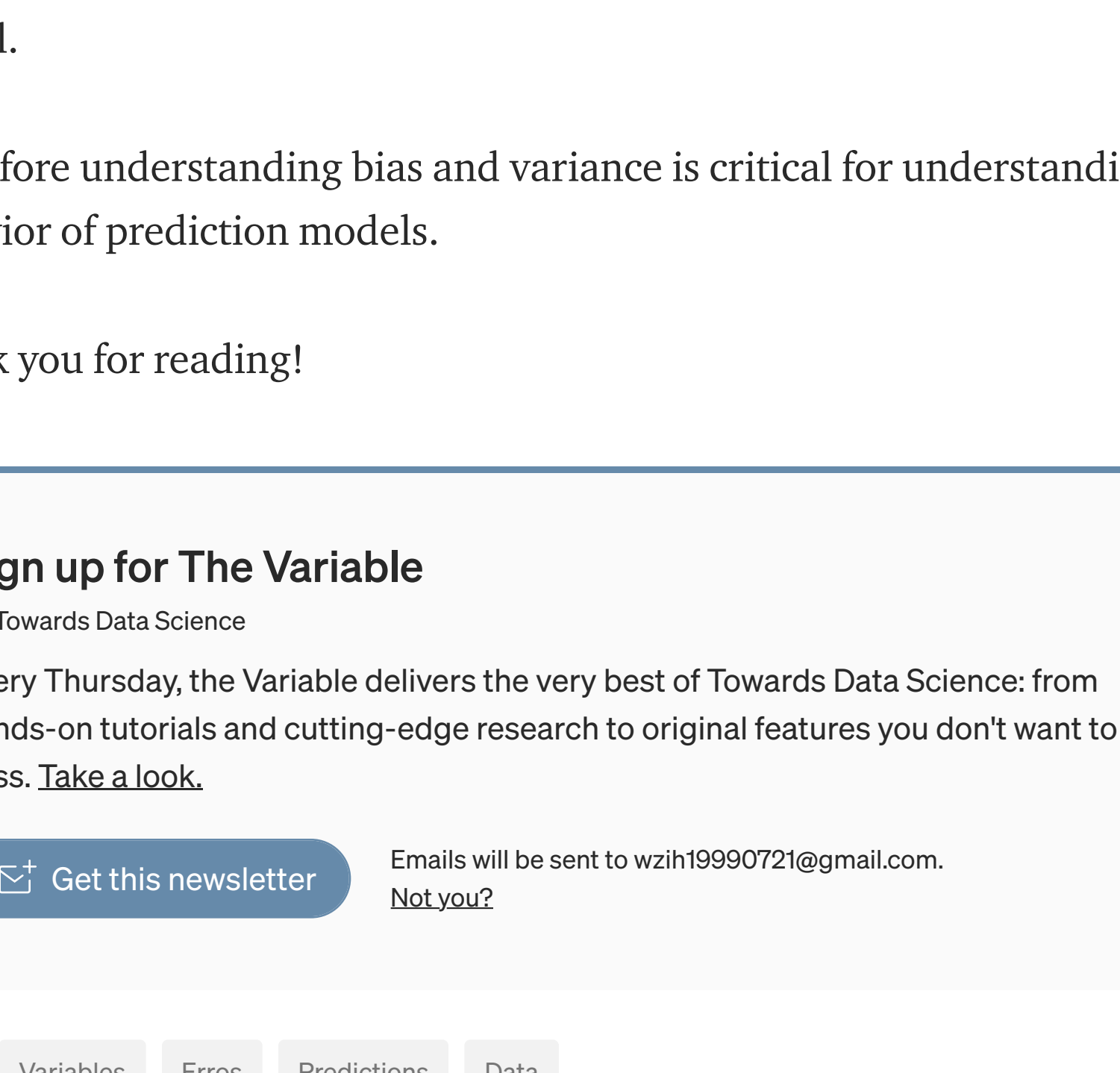
If our model is too simple and has very few parameters then it may have high bias and low variance. On the other hand if our model has large number of parameters then it's going to have high variance and low bias. So we need to find the right/good balance without overfitting and underfitting the data.

This tradeoff in complexity is why there is a tradeoff between bias and variance. An algorithm can't be more complex and less complex at the same time.

Total Error

To build a good model, we need to find a good balance between bias and variance such that it minimizes the total error.

$$\text{Total Error} = Bias^2 + Variance + Irreducible Error$$



An optimal balance of bias and variance would never overfit or underfit the model.

Therefore understanding bias and variance is critical for understanding the behavior of prediction models.

Thank you for reading!

Sign up for The Variable

By Towards Data Science

Every Thursday, the Variable delivers the very best of Towards Data Science: from hands-on tutorials and cutting-edge research to original features you don't want to miss. [Take a look.](#)

Get this newsletter

Emails will be sent to w2h19990721@gmail.com. Not you?

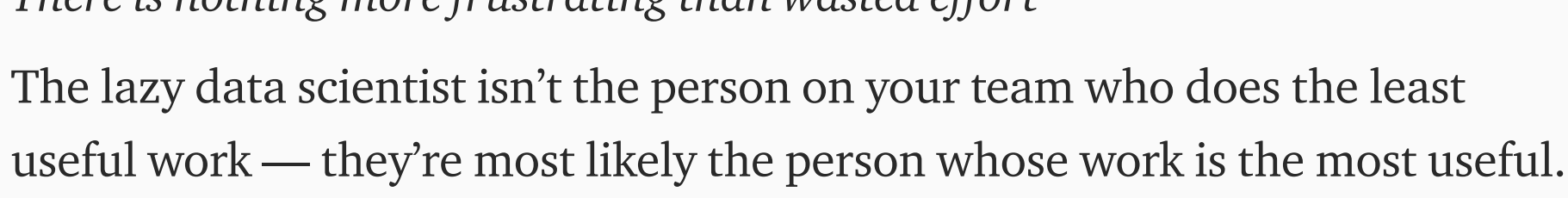
Bias

Variables

Errors

Predictions

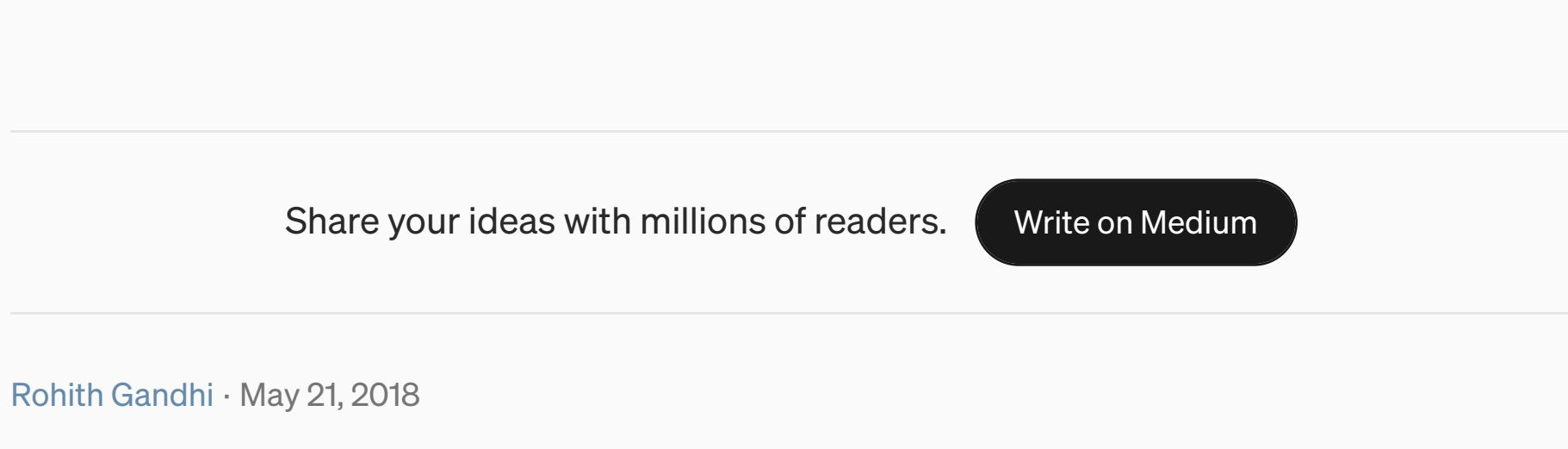
Data



More from Towards Data Science

Your home for data science. A Medium publication sharing concepts, ideas and codes.

Robert de Graaf · May 21, 2018



The Lazy Data Scientist

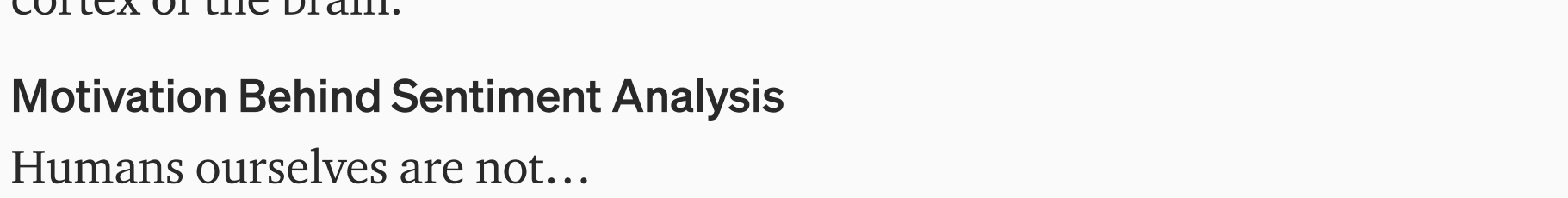
There is nothing more frustrating than wasted effort

The lazy data scientist isn't the person on your team who does the least useful work — they're most likely the person whose work is the most useful. They are simply the person who most hates to see effort go to waste.

Working on the wrong problem

The worst kind of wasted effort is fruitful effort applied to the wrong problem — it's simply heartbreaking, considering what could have been if the effort had been applied correctly. ...

Read more · 4 min read



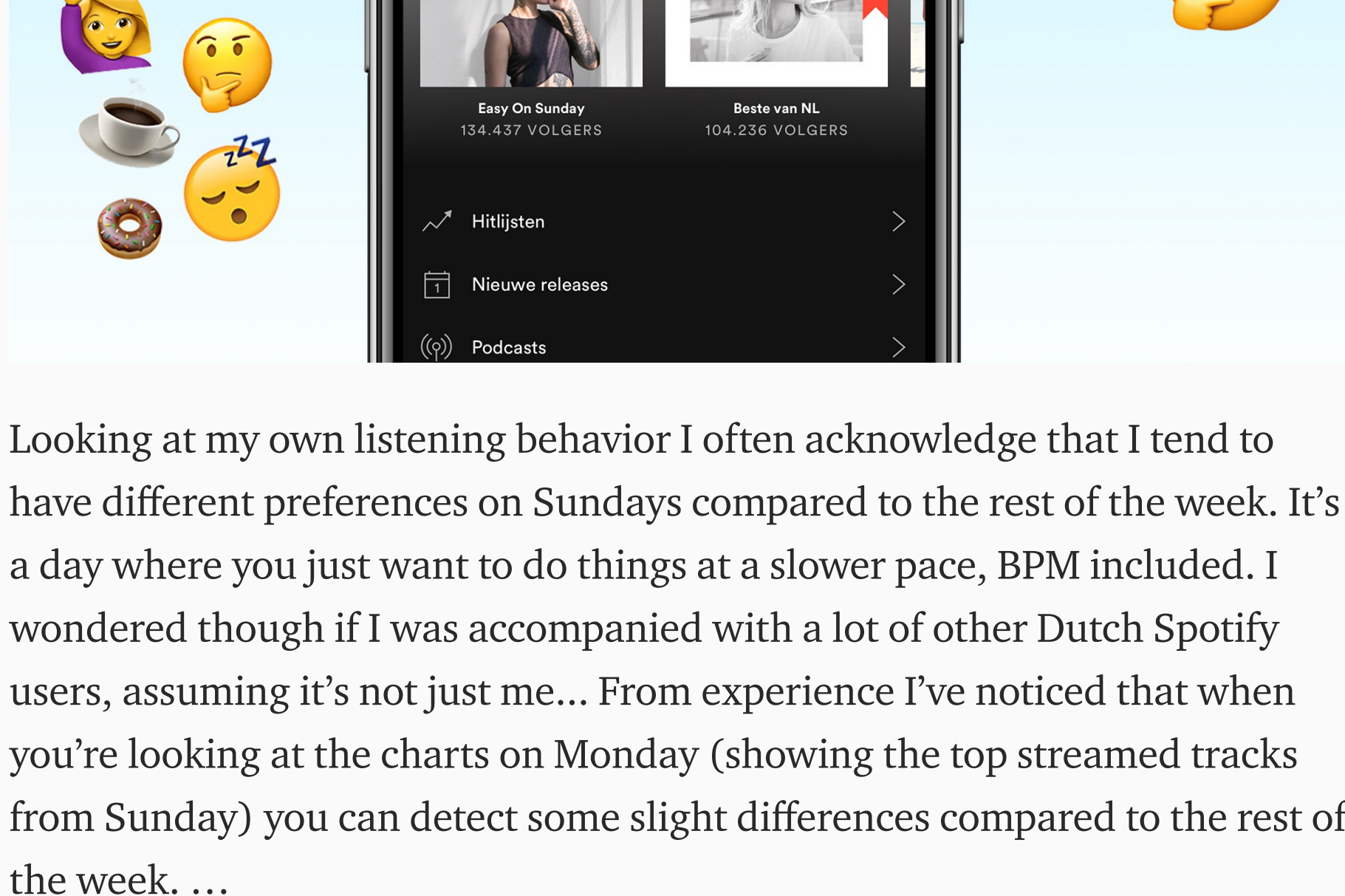
Share your ideas with millions of readers.

Write on Medium

Rohith Gandhi · May 21, 2018

Sentiment Analysis through LSTMs

Build your own model in less than 50 lines of code



Looking at my own listening behavior I often acknowledge that I tend to have different preferences on Sundays compared to the rest of the week. It's a day where you just want to do things at a slower pace, BPM included. I wondered though if I was accompanied with a lot of other Dutch Spotify users, assuming it's not just me... From experience I've noticed that when you're looking at the charts on Monday (showing the top streamed tracks from Sunday) you can detect some slight differences compared to the rest of the week. ...

Read more · 5 min read



Carolina Bento · May 21, 2018

Hypothesis Testing In Real Life

Real world problems solved with Math

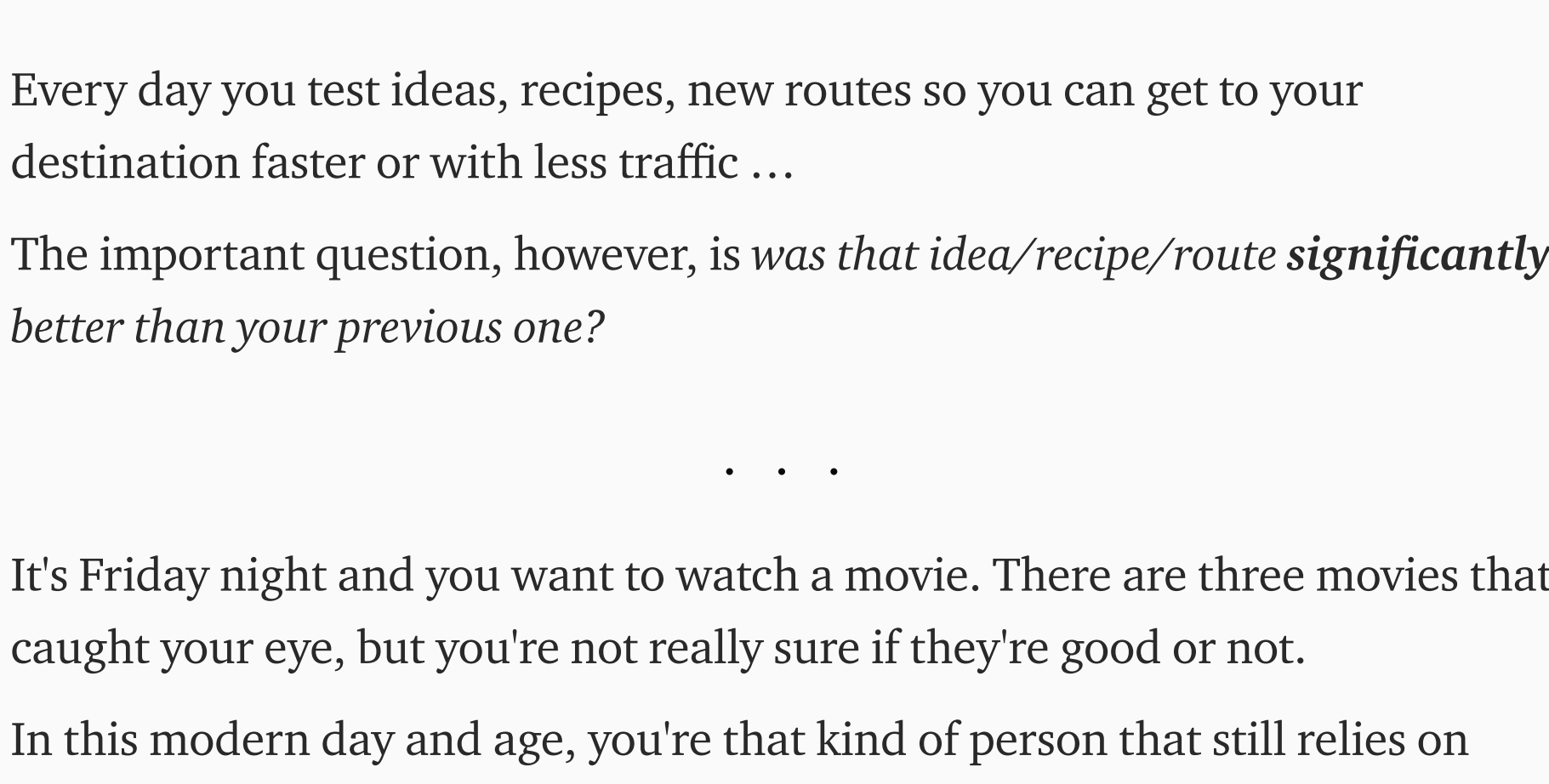


Photo by JESHOOTS.COM on Unsplash

Every day you test ideas, recipes, new routes so you can get to your destination faster or with less traffic ...

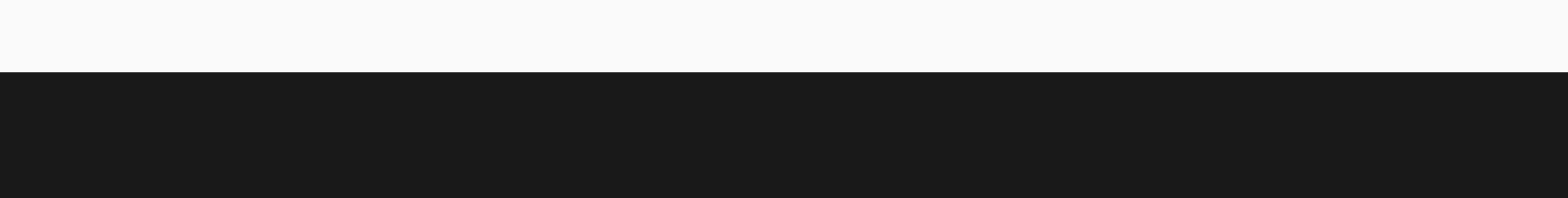
The important question, however, is *was that idea/recipe/route significantly better than your previous one?*

...

It's Friday night and you want to watch a movie. There are five movies that caught your eye, but you're not really sure if they're good or not.

In this modern day and age, you're that kind of person that still relies on family and friends for recommendations. ...

Read more · 10 min read



Read more from Towards Data Science