|  |  |
| --- | --- |
| |  | | --- | | **Interview Query Solution #32 | Food Delivery Times** | |

|  |  |
| --- | --- |
| |  | | --- | |  | |

|  |  |
| --- | --- |
| |  | | --- | | Here's the question and answer for yesterdays machine learning question.  This question was asked by: **Postmates** | |

|  |  |
| --- | --- |
| |  | | --- | | Let's say that you work as a data scientist at a food delivery company. You are tasked with building a model to predict food preparation times at a restaurant from the time when a customer sends in the order to when the meal is ready for the driver to pick up.  What are some reasons why measuring bias would be important in building this particular model? | |

|  |  |
| --- | --- |
| |  | | --- | | **Solution:** | |

|  |  |
| --- | --- |
| |  | | --- | | Let's add some more context to what is needed to build the model. The question takes a few assumptions into account here.  The first assumption is knowledge about what kind of model would be used. Since we're measuring food cooking times, it's assumed the model would be a **regression model** as the prediction would be the time it takes from order placed to food readiness to be picked up.  Okay, given we know it's a regression model, now let's figure out what it means to measure bias. Error due to bias is defined as the difference between the expected (or average) prediction of our model and the correct value which we are trying to predict. Bias measures how far off in general these models' predictions are from the correct value.  So for our case of food preparation times, the bias within a regression model will be the absolute value of the average residuals of the expected times. For example if we have some existing data and we see:  Chinese Restaurant | Actual Prep Time: 10 min | Predicted: 12 Italian Restaurant   | Actual Prep Time: 15 min | Predicted: 11 Pizza Shop             | Actual Prep Time: 22 min | Predicted: 19  We'll calculate an average value of the residuals as:  ((12 - 10) + (15-11) + (22 - 19)) / 3 = **3 minutes error on average**.  Okay so our bias is three minutes on average given the model in this example scenario. This matters because it represents the average value of the additional time that the estimate will be wrong when the driver arrives.  Given this knowledge, if applied to food prep, there's a clear reason why calculating the bias would be important. What exactly happens contextually in either situation for the bias in either direction?  If the model predicts the time estimate as being three minutes earlier than it is, then the delivery person will arrive "earlier" and be waiting for three minutes before the food comes out. However if the time estimate is three minutes later than actual, the deliver person will arrive "late" and the food **will be colder.**  Therefore bias matters in modeling out food prep times by the trade-off between either increasing food hot and readiness versus delivery person wait times. The trade-off is how we can use business decisions on **customer satisfaction in food delivery versus delivery person satisfaction** in determining the exact threshold that we want to set the bias in our model. | |