Module 2

Types of Bias:

Selection Bias

Selection bias happens when the data used in training is not large or representative enough and results in a misrepresentation of the true population. Sampling bias as a part of the overall selection bias refers to any sampling method which fails to attain true data randomization before selection. For example, a voice recognition technology is trained only with the audio language data generated by individuals with a British accent. This model will have difficulty in voice recognition when an individual with an Indian accent interacts with it. This results in lower accuracy in performance.

Overfitting and Underfitting

When a model gets trained with large amounts of data, it also starts learning from the noise and inaccurate data entries in the dataset. Consequently, the model does not categorize the data correctly, because of too many details and noise. This situation is referred to as overfitting.

When a machine learning model cannot capture the underlying trend of the data, underfitting occurs. Underfitting is commonly observed when there is less data to build an accurate model or when a linear model is being built with non-linear data. Underfitting creates a model with high bias. A model with a high bias may not be flexible enough when predicting outcomes.

Outliers

Outliers are extreme data points in a dataset that are exceptionally far from the mainstream of the data. Outliers can be caused by measurement/input error or data corruption. If an experiment's results are aimed at making decisions based on the average, then extreme data points will alter this decision, causing bias in the output.

Measurement Bias

Measurement bias is linked to underlying problems with the accuracy of the training data and how it is measured or assessed. An experiment containing invalid measurement or data collection methods will create measurement bias and biased output. For example, when testing a new feature on a mobile app that is available both for Android and iPhone users, if you perform the experiment only with the subset of iPhone users the results cannot be truly reflective, and thus introduces measurement bias in the experiment.

Module 2

Recall Bias

Recall bias in data commonly takes place in the data labeling stage when labels are inconsistently given based on subjective observations. This is also known as the false-positive rate. In machine learning, recall is defined as the rate of how many unseen points a model labeled accurately over the total number of observations. Let's say a group of test subjects share how many calories they consumed per day over the last week. As they cannot recall the precise amount, they will provide an estimation. These estimates take away from the true values, resulting in a recall bias.

Observer Bias

Observer bias, or confirmation bias, occurs when the conductor of the experiment integrates their expected outcome into the study. It can happen if a researcher starts on a project with subjective thoughts about their study, knowingly or unconsciously. An example can be seen in data labeling tasks where one data worker chooses a different label based on their subjective thoughts as opposed to other workers who follow the provided objective guidelines. Imagine the guidelines suggest that all tomato images should be tagged as fruit, yet one labeler believes that it should be classified as a vegetable and labels it accordingly. This would result in inaccurate data.

Exclusion Bias

During data pre-processing, features that are considered irrelevant end up being removed. This can consist of removing null values, outliers, or other extraneous data points. The removal process may lead to exclusion bias and the removed features may end up being underrepresented when the data is applied to a real-world problem and result in the loss of the true accuracy of the data collected. Let's imagine that referral rates from the English and Sinhala versions of the website are being compared. 98% of the clicks come from the English version and 2% come from the Sinhala version. One can choose to leave the 2% out, thinking it would not affect the final analysis. By doing so, one may miss that Sinhala clicks have a higher conversion rate compared to the English website clicks. This would lead to exclusion bias and delivers an inaccurate representation of the collected data.

Racial Bias

Racial bias, or demographic bias, occurs when the training data is reflecting a certain demographic, such as a particular race. When a model is trained on racially biased data, the outcome itself can be skewed. Imagine that image data used in the training of self-driving cars mostly features Caucasian individuals. This would mean that self-driving cars will be more likely to recognize Caucasian pedestrians than darker-skinned pedestrians, resulting in less safety for darker-skinned individuals as the technology becomes more widespread. Other forms of demographic bias include class and gender bias, which affect training outcomes in similar ways.

Module 2

Association Bias

Association bias skews, misleads, or distorts the way a machine learning model learns to associate certain features to be true based on the training data. Essentially, this reinforces a cultural bias if the data was not collected thoughtfully. If the training dataset labels all pilots as men and all flight attendants as women, for that specific model female pilots and male flight attendants won't exist, hence creating an association bias.