Feature Importance :- How individual features contribute to the final model.

Statistical bias :- Training data does not represent accurately the problem space.
Models are as good as the data they are trained on.

AWS Sage Maker , Clarify.

Statistical bias :-

→ Tendency of a statistic to overestimate or underestimate a parameter.

→ Some elements of dataset are heavily represented than others

Eg :- 1) Fraud detection, most txns are non-fradulent

Remedy :- Include more fradulent txns.

2) Product Reviews,

say product category A has more number of reviews and fewer examples of other categories say B, C.

The model can perform well on product category A examples but not that well on B, C.

Causes :-

Bias on human generated content especially on.

1) Activity Bias :- - Social Media Content

Say we have collected 10 years of data from FB. This contains data from very small percentage of actual population, who are active on social media. ∴ Data not representative of entire population.

2.) Societal bias :- Bias on human generated content based on the preconceived notions that exist in the society. Everyone has some unconscious bias.

3.) Selection bias :- Bias introduced by ML system itself.
Feedback loop. [The ML app can give some options to select. Based on user i/p, it used as training data and introduces feedback loops]

In streaming service, we may give feedback about liking a movie. Then the service might provide some relevant movies. There is a chance that some movies might not be shown at all.

4.) Data Drift :- This occurs when the data distribution is significantly different from the initial training data when the model is trained on.

Data Drift / Data Shift

Covariant Drift

The distribution of independent features may change

Say interest rate changed before / after recession in a financial application

Prior Probability Drift

The distribution of target variables may change

Concept Drift

The relationship b/w the features and labels may change.

Measure the statistical bias :-

Facet – a sensitive feature for which we want to analyze for imbalances.

Class Imbalance (CI) :-

measures the imbalance in number of members b/w different facet values.

Eg:- Does product_category 'A' has disproportionately large number of examples than any other category.

Difference in Proportion of Labels :- (DPL)

- measures the imbalance of positive outcomes b/w different facet values.

Eg:- Does a product category 'A' has disproportionately higher ratings than others

Tools to identify bias

Data Wrangler
- sources
- viz
- tranform
- statistical bias report ✓
- feature importance

- explore the data in more visual format,
- configure what goes into bias report using drop-down options
- launch the processing job with just a click.

* It uses just a subset of data to detect bias

Clarify
- statistical bias report ✓
- model bias report [b/w trained and deployed]
- explainability
- drift

```
from sagemaker import clarify
clarify_processor = clarify.SageMakerClarifyProcessor(
    instant_count=1,
    instance_type=
)
```
↳ construct that allows to scale the bias detection process into a distributed cluster.

```
clarify.DataConfig(...
    ...
)
clarify.BiasConfig(
    :
    facet_name=
)
```
→ for which facet you are measuring the bias.

Clarify:-
- API based approach
- ability to scale out the bias detection process.
- better for large volumes of data

```
clarify_processor.run_pre_training_bias(
        data_config=          ,
        data_bias_config=     ,
        methods = ['CI','DPL',...]
        .
        .
        .
    )
```

Feature Importance :-

- how valuable a feature is relative to other features

Eg:- Predict sentiment for a product → which features play a role?

SHAP :- Shapley Additional Explanations.

explains how the features are related in the prediction of the model.

Local vs Global explanations.

Local exploration - focuses on indicating how the individual feature corresponds to the final model.

Global explanation :- how the data in entirety contributes to the model.

SHAP framework - very extensive in nature. It considers all combinations of feature values with all possible outcomes for the ML model.
- time intensive but consistent in local accuracy.

AWS SM → New data flow
↓ import data
raw data from S3
↓ add analysis to data
Quick Model (choose analysis name, Label).

Data Wrangler uses a subset of data.
(70% train, 30% test)

- It ran random-cut forest and provides f1 score, importance score on the test set.