Feature Importance

The importance of a feature is the increase in the prediction error of the model after we permuted the feature's values, which breaks the relationship between the feature and the true outcome.

We measure the importance of a feature by calculating the increase in the model's prediction error after permuting the feature. A feature is "important" if shuffling its values increases the model error, because in this case the model relied on the feature for the prediction. A feature is "unimportant" if shuffling its values leaves the model error unchanged, because in this case the model ignored the feature for the prediction.

Data scientists often focus on optimizing model performance. It therefore important to understand how the features in our model contribute to prediction.some model biases are socially or legally unacceptable, e.g in the context of sensitive automated decision making the General Data Protection Regulation (GDPR) stipulates the right to know what an automated decision was based on.Hence the need to understand how your model contributes to model prediction.

Classes of feature importance:

1. ensemble tree specific feature importance (local model-specific),
2. permuted feature importance (global model-agnostic) - gives us a measure of uncertainty.
3. LIME (local model-agnostic) - can answer "what would happen if"
4. Shapley values (local model-agnostic) - an be used to explain which feature(s) contribute most to a specific prediction.

This is a **tree-specific feature importance measure** and computes the average reduction in impurity across all trees in the forest due to each feature. That is, features that tend to split nodes closer to the root of a tree will result in a larger importance value.

A model-agnostic approach is **permutation feature importance**. After evaluating the performance of your model, you permute the values of a feature of interest and reevaluate model performance. An advantage of permutation is that it gives us a measure of uncertainty. Its main downside of is that it can be computationally demanding if the number of features is large.

Local interpretable model-agnostic explanations (LIME) is a technique aimed at explaining which features are most important in specific areas of the feature space. This can be used to assess for a specific subject which features contributed most to its prediction.
LIME's local surrogate models were designed to answer "what would happen if"

Shapley values can be used to assess local feature importance, they can be used to explain which feature(s) contribute most to a specific prediction.

Types of feature importance;

1. model specific or model-agnostic
2. global or local feature importance where Local measures focus on the contribution of features for a specific prediction, whereas global measures take all predictions into account. can be used to explain why a specific person was denied a loan from credit scoring model.

References:

https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-9-307
http://blog.datadive.net/interpreting-random-forests/
https://explained.ai/rf-importance/index.html

https://christophm.github.io/interpretable-ml-book/feature-importance.html on) and Transformation (eg. extracting a the day from dates or difference between two columns)