# Paper Summary: Wager & Athey (2018)

# Review Structure

If the paper introduces a new method and you want to implement it, the following alterations can be made to the summary structure:

1. Introduction: Briefly introduce the new method and explain the problem it aims to solve. Mention the key features of the method, such as its advantages over existing methods.
2. Methodology: Provide a detailed description of the new method, including its steps and procedures. Emphasise the key differences between the new method and existing methods. If the paper provides any performance metrics, explain how to measure the new method's performance.
3. Results: If the paper presents any results, describe them briefly. If the results are unavailable, explain how to evaluate the new method's performance and what performance improvements can be expected.
4. Implementation: Discuss the practical aspects of implementing the new method, such as the necessary software and hardware requirements, any data preparation or preprocessing steps that need to be taken, and any challenges that may arise during implementation.
5. Evaluation: Explain the effectiveness of the new method in a real-world scenario. Describe any experiments or tests that need to be conducted and what metrics should be used to evaluate the method's performance.
6. Conclusion: Summarise the key benefits of the new method and its potential impact on the field. Discuss any limitations or future research directions.
7. References: Include a list of references cited in the paper, following the appropriate citation style.

Overall, the goal of the summary is to provide a clear and concise description of the new method and how to implement it. It should also guide evaluating its effectiveness and measuring its impact.

# Introduction

**Briefly introduce the new method and explain the problem it aims to solve. Mention the key features of the method, such as its advantages over existing methods.**

The paper introduces a nonparametric method for estimating heterogeneous treatment effects in large datasets. The proposed method uses ideas from the machine learning literature to improve the performance of classical methods with many covariates.

Classical approaches to the nonparametric estimation of heterogeneous treatment effects, such as nearest neighbour matching, kernel methods, and series estimation, perform well in applications with a small number of covariates but quickly break down as the number of covariates increases. The proposed method focuses on a family of algorithms introduced by Breiman (2001a) that allow for flexible modelling of interactions in high dimensions. It builds many regression trees and averages predictions, like kernels and nearest neighbour matching. However, the random forests used in this method have a data-driven way of determining which nearby observations receive more weight, which is essential in environments with complex interactions among covariates.

The proposed method provides a data-driven way of determining which nearby observations receive more weight, which is essential in environments with complex interactions among covariates. This new method has the potential to facilitate the exploration of heterogeneous treatment effects and provide researchers with more robust insights into treatment effects. However, the authors must show that the estimator is consistent with a well-understood asymptotic sampling distribution to establish confidence intervals for causal inference. The article's main contribution is an asymptotic normality theory enabling statistical inference.

# Methodology

**Provide a detailed description of the new method, including its steps and procedures. Emphasise the key differences between the new method and existing methods. If the paper provides any performance metrics, explain how to measure the new method's performance.**

* Treatment Estimation with Unconfoundedness
  + The goal is to estimate the causal treatment effect
  + We cannot directly train machine learning on outcomes as we cannot observe treatment and control differences. We only observe one outcome.
  + Without further restrictions on the data-generating distribution, we cannot estimate the true causal effect.
  + We, therefore, assume unconfoundedness. Treatment assignments independent of potential outcomes conditional on X
  + Under the assumption of unconfoundedness, we can treat nearby observations in x space as coming from a randomised experiment
  + Nearest neighbour and other local methods are consistent estimators in this situation
  + If we assume unconfoundedness, we need to estimate the propensity of receiving treatment
  + If we had access to a simple, unbiased estimator for receiving treatment. This observation motivates propensity weighting approaches.
  + Early machine learning approaches estimate the propensity of receiving treatment using neural networks of random forests
  + In this approach, the authors take an indirect approach. We show that under regularity assumptions, causal forests can use the unconfoundedness assumption to achieve consistency without needing to estimate the propensity score strictly.
* **From Regression Trees to Causal Trees and Forests**
  + At a high level, trees and forests can be considered nearest neighbour methods with an adaptive neighbourhood metric.
  + Classical k-nearest neighbour methods such as k-nearest neighbours seek the k-closest points to x according to some pre-specified distance measure.
  + Tree-based methods seek to find training examples close to x.
  + Closeness, in this instance, is defined with respect to a decision tree
  + The closeness of points is defined with respect to a decision tree
  + The advantage of tress is that leaves can be narrower along the directions where the signal is changing fast and more comprehensive along other directions, potentially leading to a substantial increase in power
  + Authors seek to build causal trees that resemble regression analogues as closely as possible
  + Start by recursively splitting the feature space into a set of L leaves, each of which only contains a few samples
  + At each test point x, we evaluate the prediction by identifying the leaf containing x
  + This strategy is well-motivated if we believe the leaf L to be small enough that the responses Y inside the leaf are roughly identically distributed.
  + Several procedures exist to split the decision tree
  + In the context of causal trees, we want pairs small enough, so that assignment and outcome pairs act as though they come from a randomised experiment
  + We can therefore estimate the causal effect as the difference between the assigned and unassigned parties (trees)
  + Such trees can be used to grow causal forests that are consistent
  + Given a procedure for generating a single tree, the forest can aggregate predictions by averaging them
  + Individual causal trees in the forest are built using random subsamples of training examples
  + The advantage of a forest is that it is not always clear what the best causal tree is.
  + It is often best to generate many decent-looking trees and average their predictions instead of seeking a single highly optimised tree
  + This aggregation reduces variance and smooths sharp decision boundaries.

**Set of conditions under which predictions made by random forests are both asymptotically unbiased and Gaussian**

* **Compare performance with a classical nearest k-neighbour matching**
* **Causal forest dominates in terms of both bias and variance**

# Results

**If the paper presents any results, describe them briefly. If the results are unavailable, explain how to evaluate the new method's performance and what performance improvements can be expected.**

# Implementation

**Discuss the practical aspects of implementing the new method, such as the necessary software and hardware requirements, any data preparation or preprocessing steps, and any challenges that may arise during implementation.**

# Evaluation

**Explain how to evaluate the new method's effectiveness in a real-world scenario. Describe any experiments or tests that need to be conducted and what metrics should be used to evaluate the method's performance.**

# Conclusion

**Summarise the key benefits of the new method and its potential impact on the field. Discuss any limitations or future research directions.**

The article introduces a new nonparametric method for estimating treatment effects that allow for data-driven selection and maintains unbiased point estimates and valid confidence intervals. The method is adaptive and can handle large-scale applications. The proposed method uses a combination of honest trees and subsampling mechanisms of random forests to address selection bias. The new method has better mean-square error than classical methods while achieving nominal coverage in moderate sample sizes.

However, the current results only provide pointwise confidence intervals, and future research is needed to extend the theory to global functional estimation. Additionally, nearest-neighbour nonparametric estimators suffer from bias at the boundaries of the support feature space. A systematic approach to trimming at the boundaries and correcting this bias would improve confidence interval coverage.

Overall, this new method has the potential to significantly impact the field of treatment effect estimation by providing more accurate and precise estimates in large-scale applications. However, further research is needed to address the limitations and expand the method's scope.

# References

**Include a list of references cited in the paper, following the appropriate citation style.**