**TPOT(Tree based Pipeline Optimization Technique)**

Developed by Dr. Randal Olsan in 2015

It uses Expression Tree combined with stochastic search based algorithms(Genetic Programing) and uses Scikit learn as the main ML menu.

Links of papers:

[first paper](https://link.springer.com/chapter/10.1007/978-3-319-31204-0_9)

[second paper](http://dl.acm.org/citation.cfm?id=2908918)

[2017 paper](http://dl.acm.org/citation.cfm?id=3071178.3071212)

[a paper](https://academic.oup.com/bioinformatics/article/36/6/1772/5614811)

[book chapter](https://link.springer.com/chapter/10.1007/978-3-030-05318-5_8)

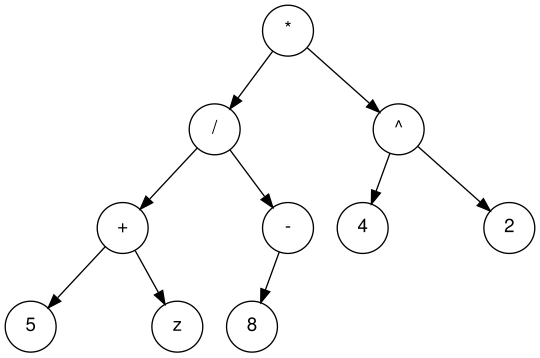
[a paper](https://www.sciencedirect.com/science/article/abs/pii/S0735109720302631?via%3Dihub)

[latest paper](https://academic.oup.com/bioinformatics/article/36/1/250/5511404)

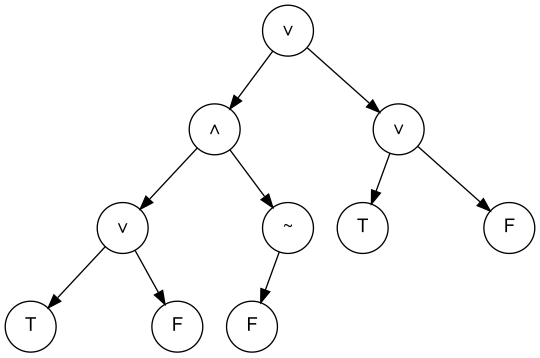
[recent review](https://link.springer.com/article/10.1007/s00439-021-02393-x)

**Expression tree:**

Used to represent expressions with the help of tree based data structure.



Binary algebraic expression tree.



Binary boolean expression tree.

**Genetic Programming Algorithm**

It is an evolutionary algorithm. It is inspired from Darwin’s principle of natural evolution. Traditionally, it evolves computer programs by modifying representation in memory based expression trees.

Steps to implement genetic algorithms:

1. Initialization:

Creation of initial population. A GP(Genetic Programming) tree contains a root node, functional nodes with children, and terminal nodes to a certain depth.

* 1. Grow: It creates individuals sequentially. It creates GP trees where branches of the tree may have varied depth.
  2. Full: It creates GP trees where all branches of the GP tree have the same depth.
  3. Ramped half and half: It creates (md-1) parts with md is the maximum depth of the tree. One half of each part is created by the grow method while the other half is created by the full method.

1. Selection:

Procedure of selecting certain individuals from the current generation to serve as parents of future generations. Selection often uses probabilistic methods to ensure the best performing individuals have a high probability of being selected.

Methods used in selection:

1. Elitism: It involves selecting the best n individuals of the current generation for seeding the next generation.
2. Other methods:

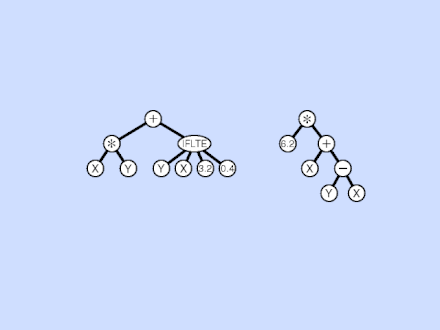
Tournament selection, lexicase selection, and fitness proportionate selection.

1. Cross Over:

Two fit individuals are chosen from the selected population to cross and produce one or two children.

Steps involved in cross-over phase of tree genetic programming:

1. Represent each individual as an inverted lisp or tree with a root node at the top.
2. Randomly select subtree from each of the trees.
3. In case of one child cross-over, Delete the selected subtree of one tree and replace it with the selected subtree of another tree.
4. In case of two child cross-over, Replace both selected subtrees of both trees with each other.



Genetic programming subtree cross-over.

1. Replication:

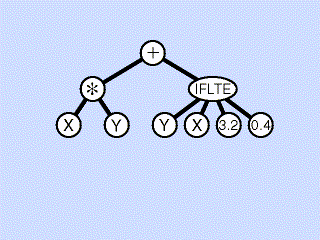
Some of the candidates which do not take part in cross-over steps are copied into the next generation as a child generated by asexual reproduction.

1. Mutation:

This is used to include genetic changes between generations. Mutation processes start from a fit syntactically correct parent and aim to randomly create a syntactically correct child.

Types of mutations:

* Replace a subtree with randomly generated subtree.
* Select a leaf node and replace it with a randomly selected leaf node.
* Select an internal function and replace it with another function of the same arity.
* Hoist mutation: Select a subtree and replace it with its child subtree.
* Subtree mutations based on selected function and terminal sets either increase or decrease the size of the tree.
* Some subtree mutations try to carefully control the size of replacement subtree and thus control the size of the tree.



Subtree based mutation.