

**GTC 2017**  
**Silicon Valley, California**

---

**An Approach to a High Performance Decision Tree Optimization  
within a Deep Learning Framework for Investment and Risk  
Management**

*Yigal Jhirad and Blay Tarnoff*  
*May 9, 2017*

# GTC 2017: Table of Contents

## **I. Deep Learning in Finance**

- Deep Learning
- Machine Learning Topography
- Neural Networks
- Augmented Decision Tree Models

## **II. Parallel Implementation**

## **III. Summary**

## **IV. Author Biographies**

**DISCLAIMER:** This presentation is for information purposes only. The presenter accepts no liability for the content of this presentation, or for the consequences of any actions taken on the basis of the information provided. Although the information in this presentation is considered to be accurate, this is not a representation that it is complete or should be relied upon as a sole resource, as the information contained herein is subject to change.

---

# Deep Learning

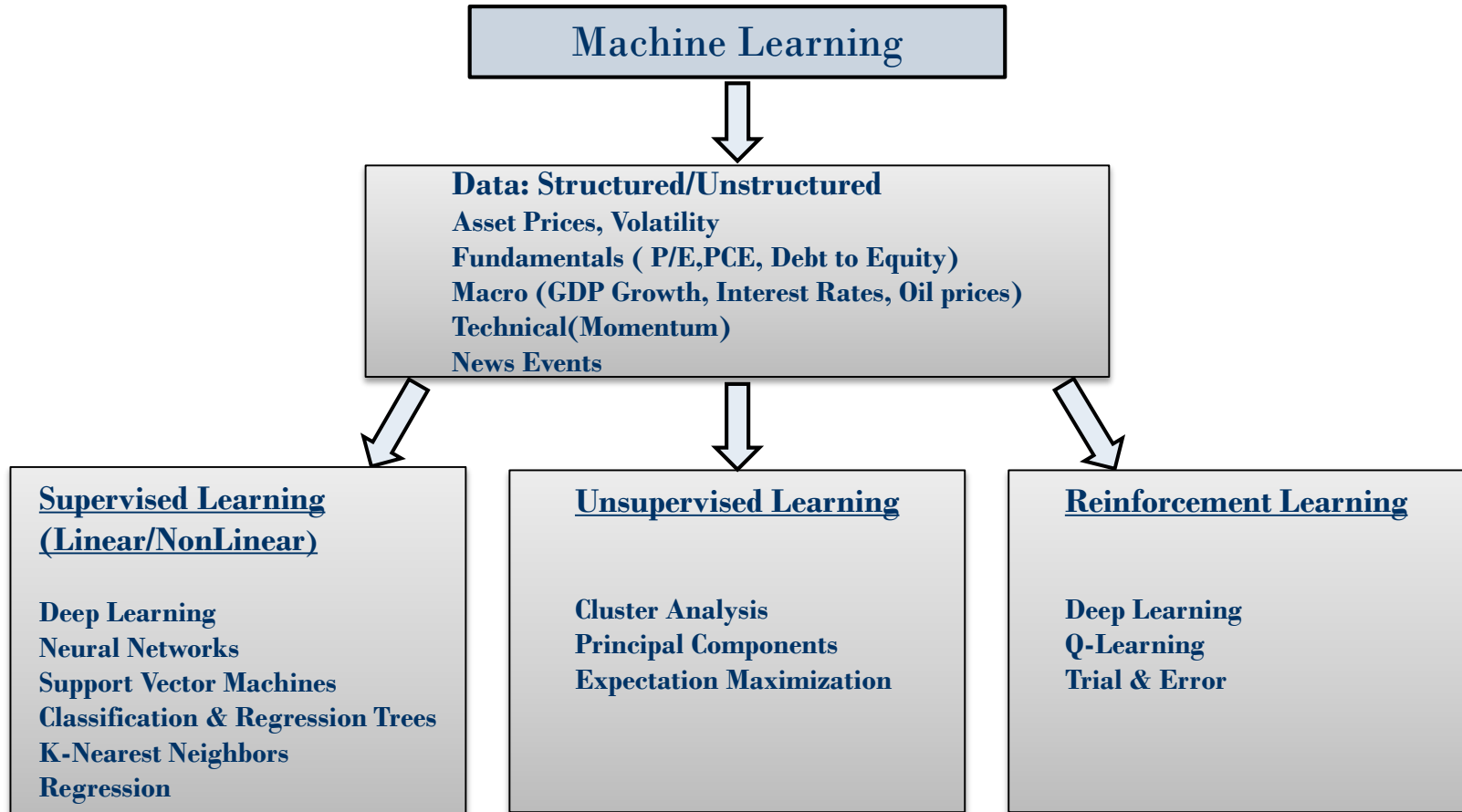
- **Investment & Risk Management**

- Forecast Market Returns, Volatility, Liquidity, Economic Cycles
- Opportunity for deeper integration of models into Investment and risk processes
- Big Data including Time Series Data, Interday, and Intraday

- **Challenges include state dependency and stochastic nature of markets**

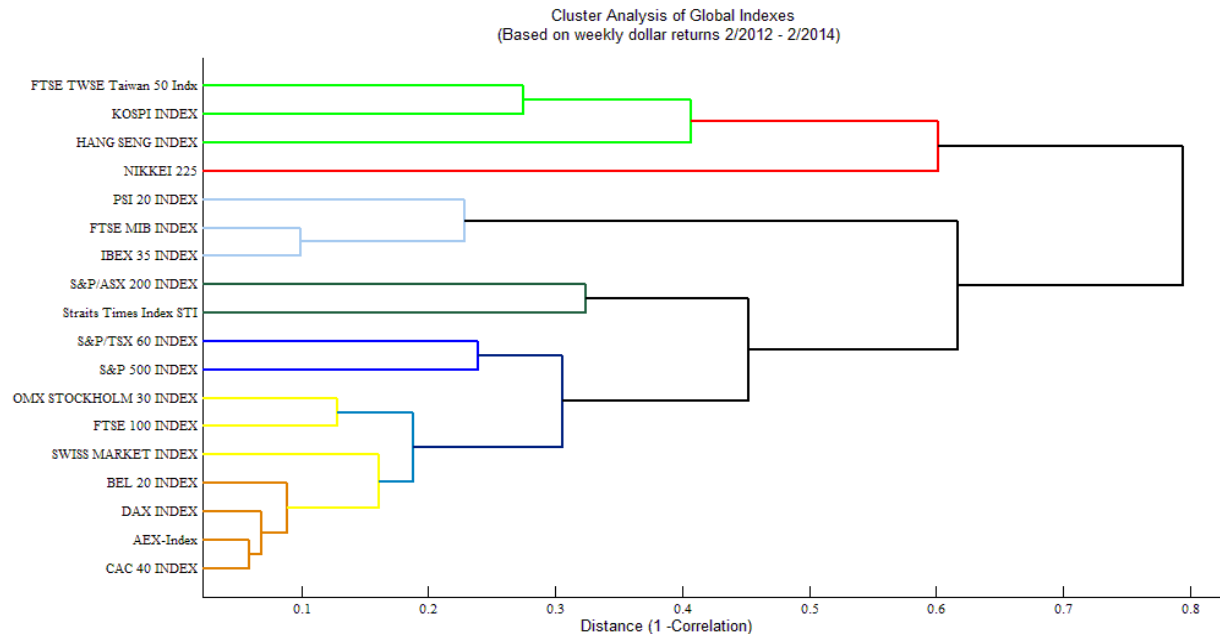
- Time series, Overfitting
- Generalization of data to produce accurate out of sample predictions

# Artificial Intelligence



# Unsupervised Learning: Cluster & Cointegration Analysis

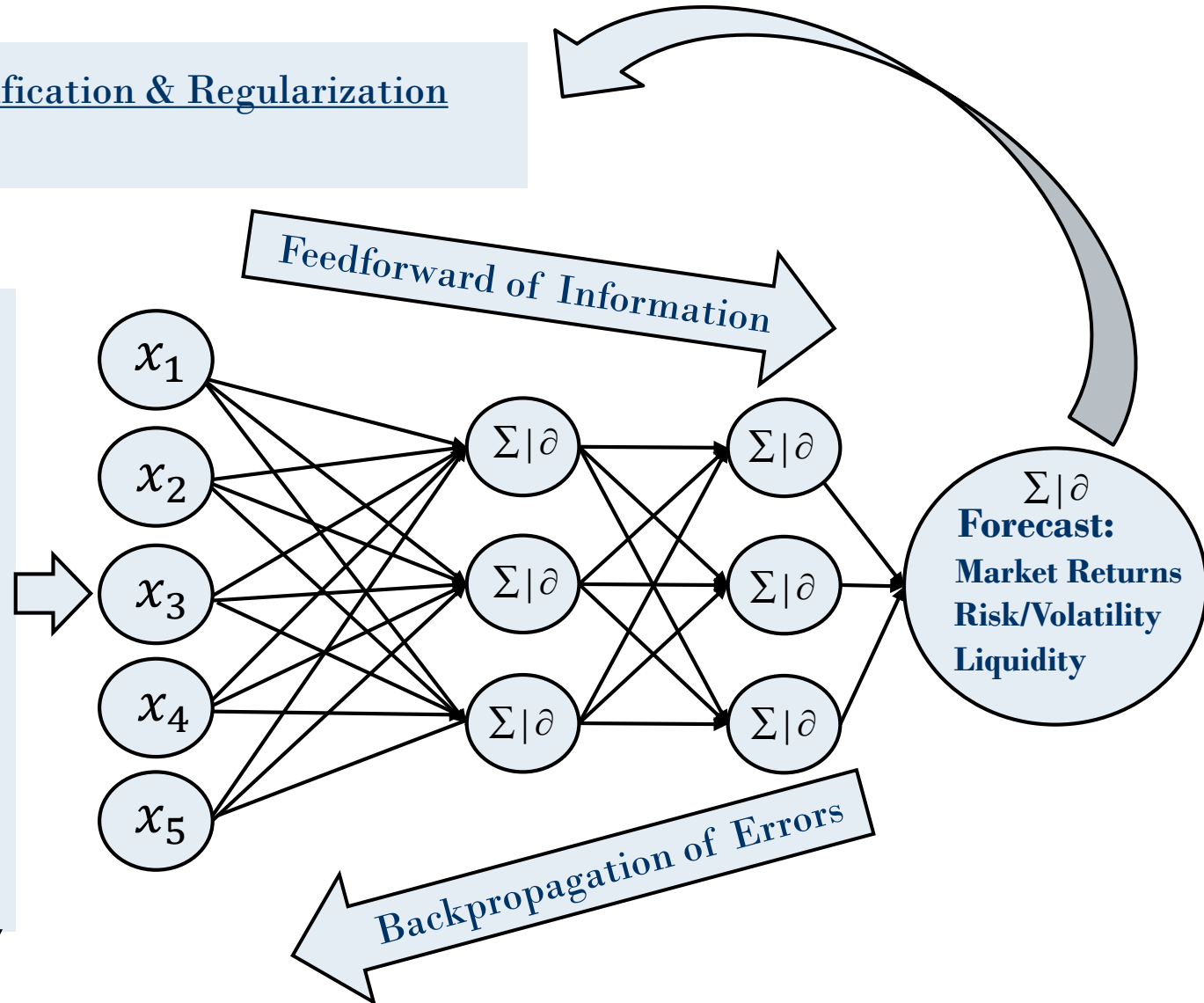
- **Cluster Analysis: A multivariate technique designed to identify relationships and cohesion**
  - Factor Analysis, Risk Model Development
- **Correlation Analysis: Pairwise analysis of data across assets. Each pairwise comparison can be run in parallel.**
  - Use Correlation or Cointegration as primary input to cluster analysis
  - Apply proprietary signal filter to remove selected data and reduce spurious correlations



# Supervised Learning: Neural Networks

Feature(Factor)Identification & Regularization  
Decision Trees

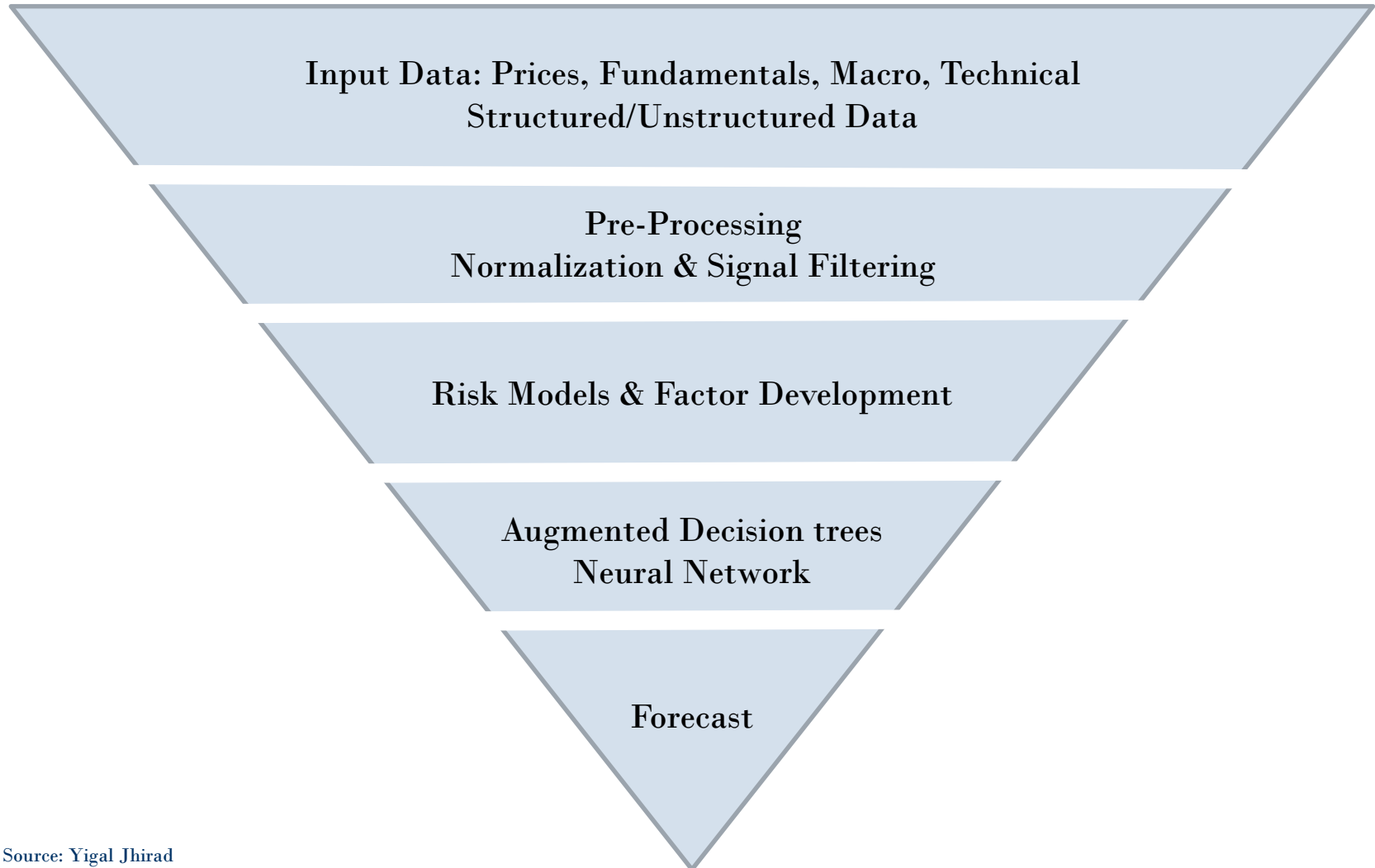
**Inputs:**  
Fundamental/Macro/Technical  
Price/Earnings  
Momentum/RSI  
Realized & Implied Volatility  
Value vs Growth  
GDP Growth/Interest Rates  
Dollar Strength  
Credit Spreads



# Augmented Decision Trees Models

- **Decision Trees**
  - Decision Trees can be more intuitive
  - Integrated feature (factor) selection
  - Utilize classification vs. regression tree to eliminate instability of point estimates
  - Non-parametric and effectively processes non-linear relationships
  - Robust to outliers
  - Purity (e.g. Entropy, Gini Index )
- **Propose an Augmented Decision Tree model that can help drive deep learning by identifying appropriate factors across market regimes**
  - Enhance construction by utilizing Optimization with added penalty function
  - Drive a Deep Learning process to create more robust prediction models
- **CUDA leverages GPU Hardware providing computational power to drive optimization algorithms**

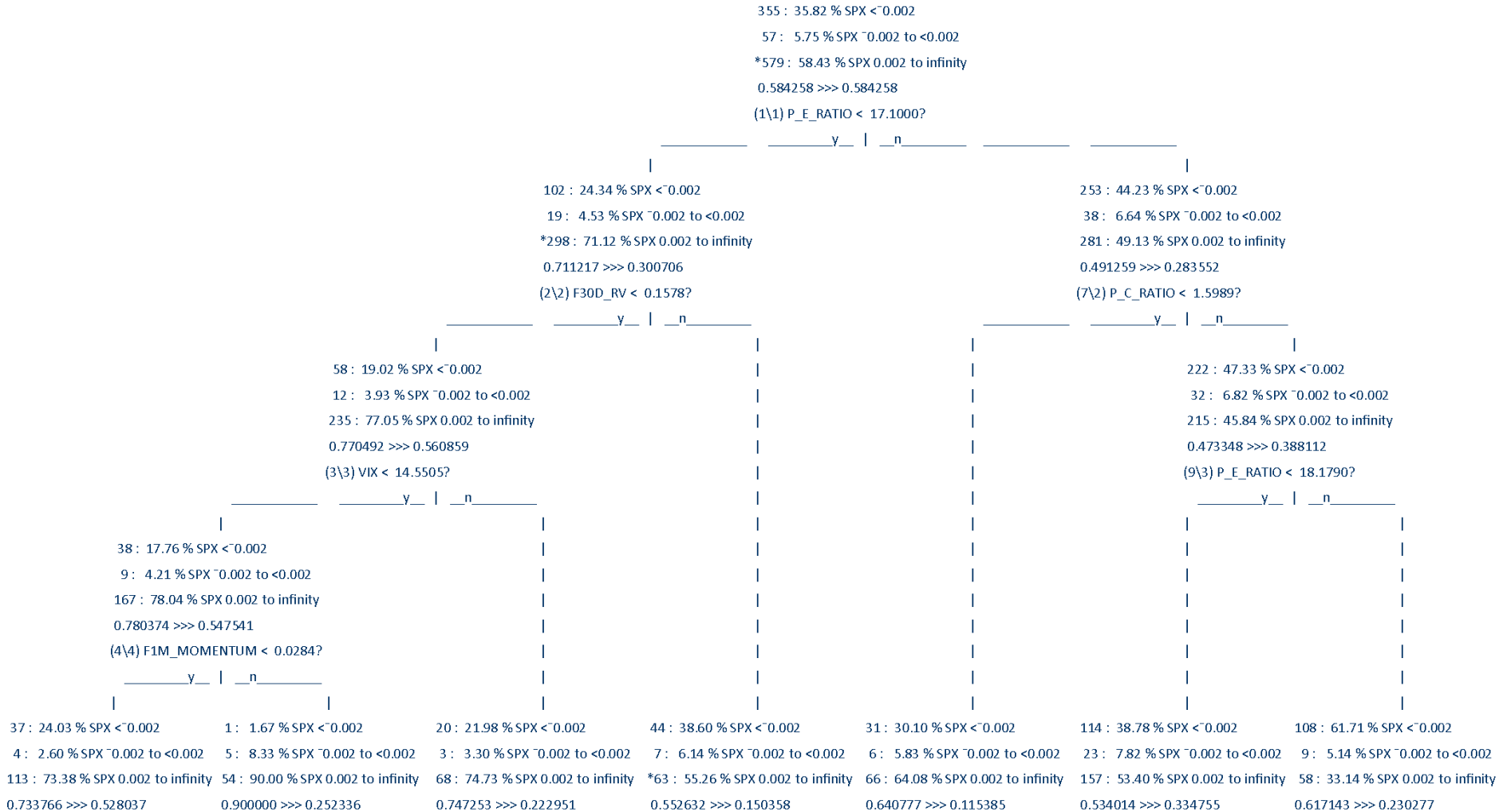
# Workflow



Source: Yigal Jhirad



# Decision Tree



# GPU Overview

- **Objective to create a tool that will produce decision trees for use in external, wrapper processes**
- **Solution leverages the power of recursive dynamic parallelism**
- **Engine: heart of the process**
- **Transparent, understandable, fast**
- **Layered control, driven by invoking application**
- **Can be used in neural network, optimization, risk assessment, other**

# General Philosophy and Approach to GPU Programming

- **Avoid black box: GPU process should be straightforward and transparent, to produce predictable, understandable results**
- **Leverage power of GPU to reach where otherwise not possible**
- **Call GPU process iteratively from external, wrapper processes that use those results intelligently**

# Nature of the Task: Generate Decision Tree

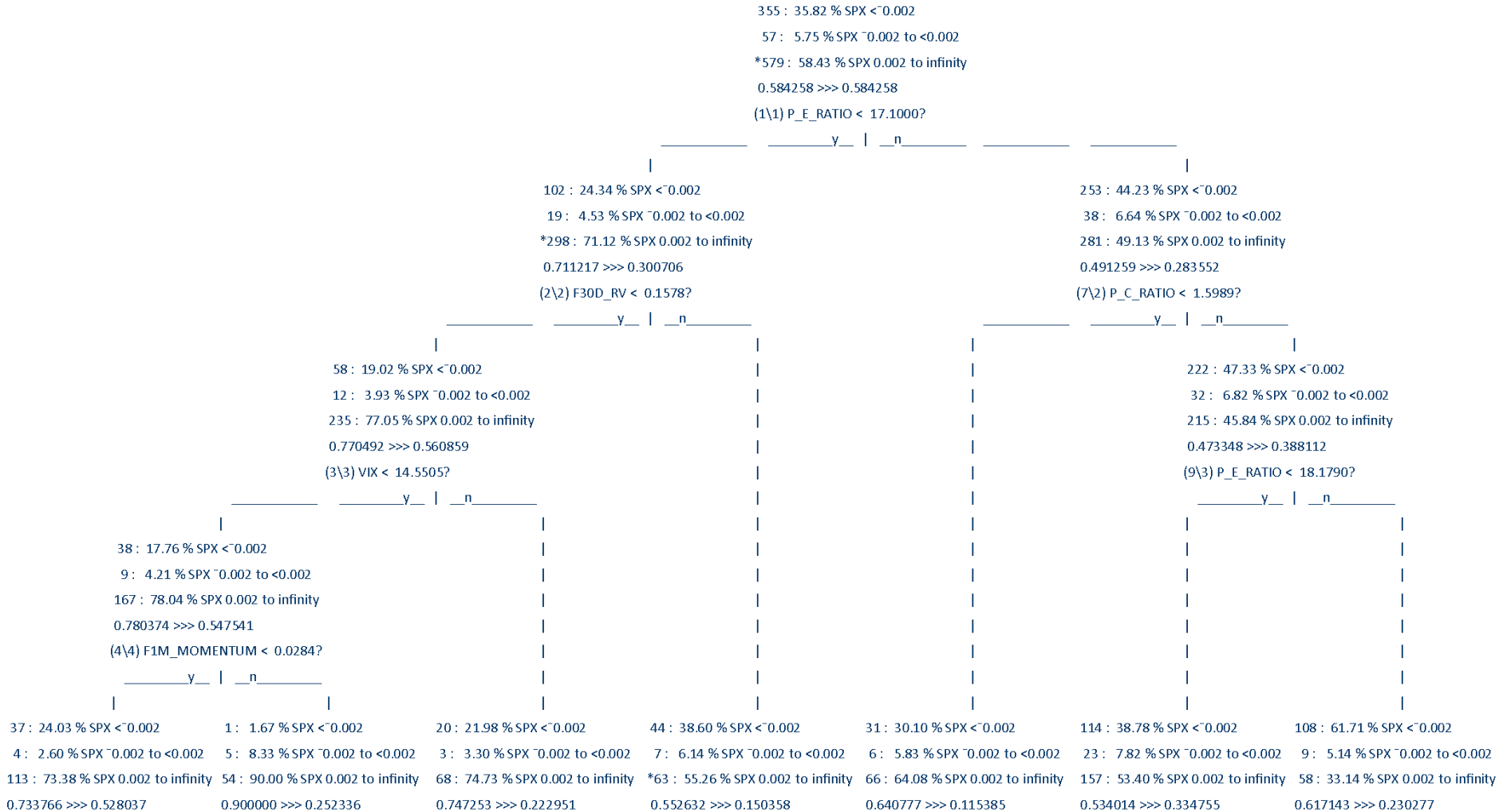
- Given a set of underlying “factors” and a corresponding time-shifted “class”, produce the “best” decision tree
- Underlying factors presumed to have predictive power
- Underlying factors and time-shifted class comprised of timeseries vectors

Factors	2/3/2004	2/4/2004	2/5/2004	2/6/2004	2/9/2004	2/10/2004	2/11/2004	2/12/2004	2/13/2004	2/17/2004
F1M_MOMENTUM	0.02	0.02	0.01	0.02	0.01	0.01	0.03	0.02	0.02	0.02
P_E_RATIO	20.91	20.74	20.77	21.03	20.98	21.08	21.31	21.21	21.09	21.30
VIX	17.34	17.87	17.71	16.00	16.39	15.94	15.39	15.31	15.58	15.40
F1W_MOMENTUM	-0.01	0.00	0.00	0.01	0.00	0.01	0.03	0.02	0.00	0.02
F30D_RV	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.12
IV_RV_1	1.35	1.44	1.44	1.17	1.33	1.24	1.14	1.18	1.25	1.16 . . .
F1M_UPSIDE_SKEW	0.06	0.06	0.06	0.05	0.08	0.07	0.06	0.07	0.08	0.08
F1M_DOWNSIDE_SKEW	0.17	0.17	0.18	0.15	0.17	0.16	0.14	0.13	0.16	0.13
P_C_RATIO	1.64	1.68	1.71	1.74	1.76	1.74	1.68	1.72	1.74	1.82
OPEN_INTEREST	0.96	0.58	0.65	1.43	1.19	1.75	3.97	2.82	1.69	3.99
BB_UPPER_BAND	0.99	0.98	0.98	0.99	0.99	0.99	1.00	1.00	0.99	1.00
BB_LOWER_BAND	1.02	1.01	1.01	1.02	1.02	1.02	1.03	1.03	1.02	1.03

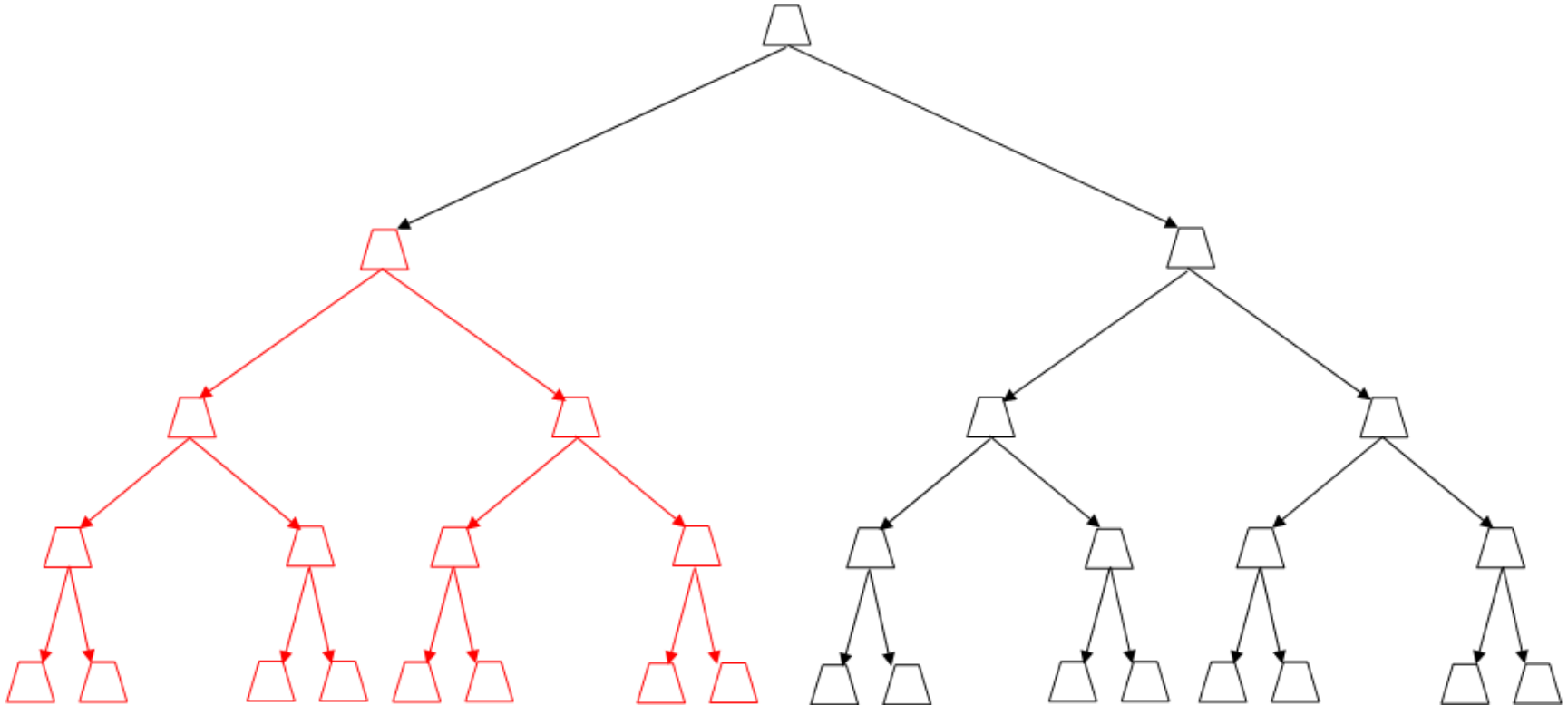
  

Class	3/4/2004	3/5/2004	3/8/2004	3/9/2004	3/10/2004	3/11/2004	3/12/2004	3/15/2004	3/16/2004	3/17/2004
SPX	0.02	0.02	0.01	-0.02	-0.03	-0.02	-0.05	-0.04	-0.02	-0.03 . . .

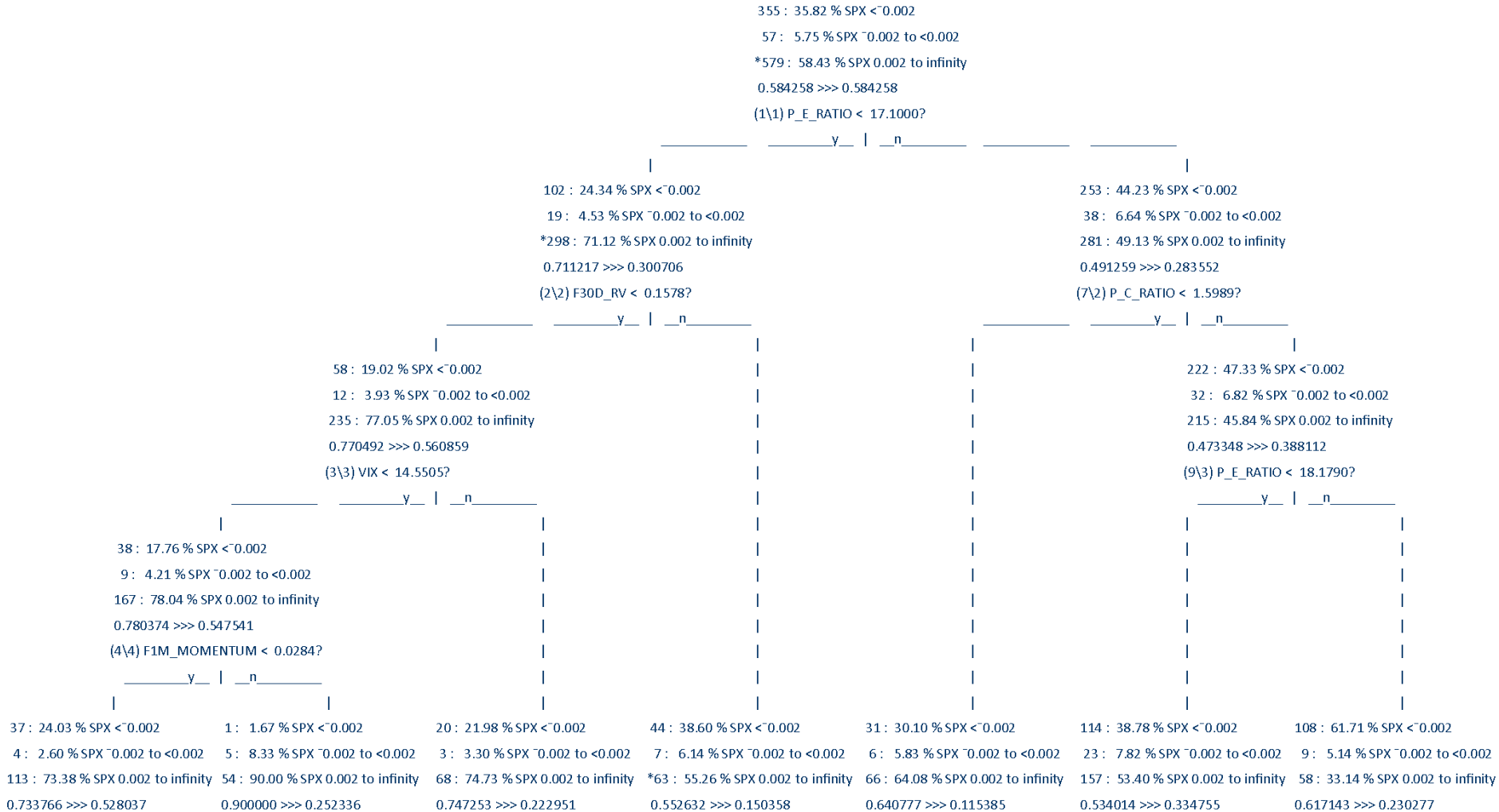
# Nature of the Task: Generate Decision Tree



# Nature of the Task: Naturally Recursive



# Approach: Pre-process to Convert Continuous Problem to Discrete



# Approach: Pre-process to Convert Continuous Problem to Discrete

Factors	2/3/2004	2/4/2004	2/5/2004	2/6/2004	2/9/2004	2/10/2004	2/11/2004	2/12/2004	2/13/2004	2/17/2004
F1M_MOMENTUM	0.02	0.02	0.01	0.02	0.01	0.01	0.03	0.02	0.02	0.02
P_E_RATIO	20.91	20.74	20.77	21.03	20.98	21.08	21.31	21.21	21.09	21.30
VIX	17.34	17.87	17.71	16.00	16.39	15.94	15.39	15.31	15.58	15.40
F1W_MOMENTUM	-0.01	0.00	0.00	0.01	0.00	0.01	0.03	0.02	0.00	0.02
F30D_RV	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.12
IV_RV_1	1.35	1.44	1.44	1.17	1.33	1.24	1.14	1.18	1.25	1.16 . . .
F1M_UPSIDE_SKEW	0.06	0.06	0.06	0.05	0.08	0.07	0.06	0.07	0.08	0.08
F1M_DOWNSIDE_SKEW	0.17	0.17	0.18	0.15	0.17	0.16	0.14	0.13	0.16	0.13
P_C_RATIO	1.64	1.68	1.71	1.74	1.76	1.74	1.68	1.72	1.74	1.82
OPEN_INTEREST	0.96	0.58	0.65	1.43	1.19	1.75	3.97	2.82	1.69	3.99
BB_UPPER_BAND	0.99	0.98	0.98	0.99	0.99	0.99	1.00	1.00	0.99	1.00
BB_LOWER_BAND	1.02	1.01	1.01	1.02	1.02	1.02	1.03	1.03	1.02	1.03

Class	3/4/2004	3/5/2004	3/8/2004	3/9/2004	3/10/2004	3/11/2004	3/12/2004	3/15/2004	3/16/2004	3/17/2004
SPX	0.02	0.02	0.01	-0.02	-0.03	-0.02	-0.05	-0.04	-0.02	-0.03 . . .

- **Where to divide each factor given as input parameter, based on standard deviations, iterative observations, or other criteria**
- **Where to divide each class also given as input parameter, based on desired signal**



# Approach: Pre-process to Convert Continuous Problem to Discrete

Factors	2/3/2004	2/4/2004	2/5/2004	2/6/2004	2/9/2004	2/10/2004	2/11/2004	2/12/2004	2/13/2004	2/17/2004
F1M_MOMENTUM	6	6	5	6	6	6	7	6	6	6
P_E_RATIO	9	9	9	9	9	9	9	9	9	9
VIX	5	5	5	5	5	5	5	5	5	5
F1W_MOMENTUM	5	5	5	6	6	6	8	7	6	7
F30D_RV	5	5	5	5	5	5	5	5	5	5
IV_RV_1	7	7	7	5	6	6	5	5	6	5 . . .
F1M_UPSIDE_SKEW	2	1	2	1	3	2	2	2	3	3
F1M_DOWNSIDE_SKEW	4	4	5	3	4	4	3	2	4	2
P_C_RATIO	5	5	6	6	6	6	5	6	6	7
OPEN_INTEREST	5	4	4	5	5	5	6	6	5	6
BB_UPPER_BAND	6	6	6	7	6	7	7	7	6	7
BB_LOWER_BAND	4	4	4	5	4	5	5	5	5	5

Class	3/4/2004	3/5/2004	3/8/2004	3/9/2004	3/10/2004	3/11/2004	3/12/2004	3/15/2004	3/16/2004	3/17/2004
SPX	2	2	2	0	0	0	0	0	0	0 . . .

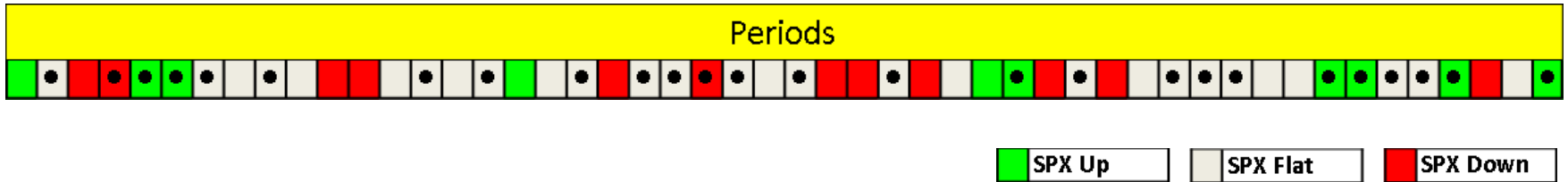
- **Conversion to discrete integer input for wrapper control, simplicity, speed and accuracy**

# Approach: Exhaustive Search

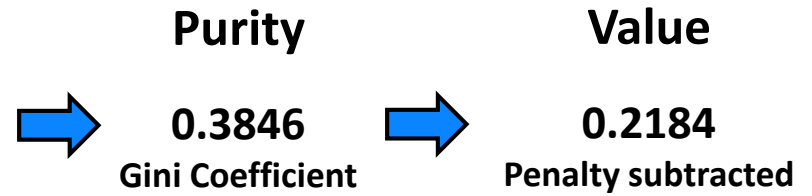
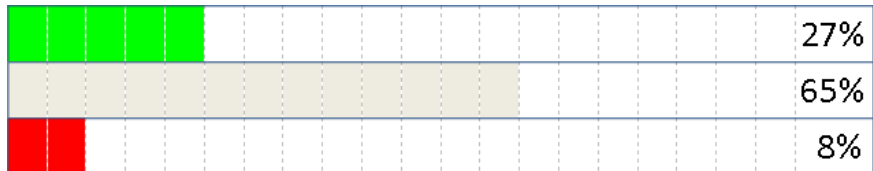
<u>Factor eligible for bifurcation at each node</u>	<u>Number of potential bifurcation points at each node</u>
F1M_MOMENTUM	11
P_E_RATIO	11
VIX	11
F1W_MOMENTUM	11
F30D_RV	11
IV_RV_1	11
F1M_UPSIDE_SKEW	11
F1M_DOWNSIDE_SKEW	11
P_C_RATIO	11
OPEN_INTEREST	11
BB_UPPER_BAND	11
BB_LOWER_BAND	11

- **At any given node, any of the factors may be split at any (pre-determined) point**
- **Total potential bifurcation points at any node = sum of potential bifurcation points for all factors, 132 in this example**

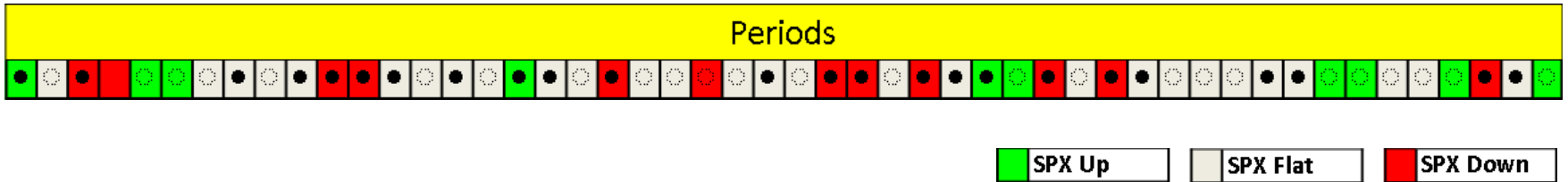
# Basic Algorithm: Leaf Level



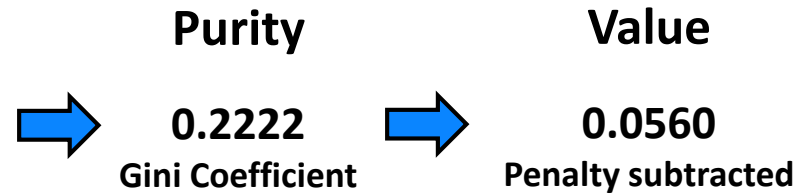
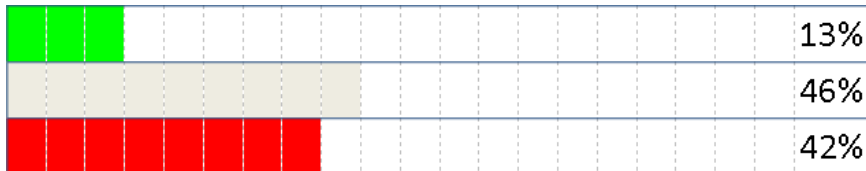
**P\_E\_RATIO < 17.1000**



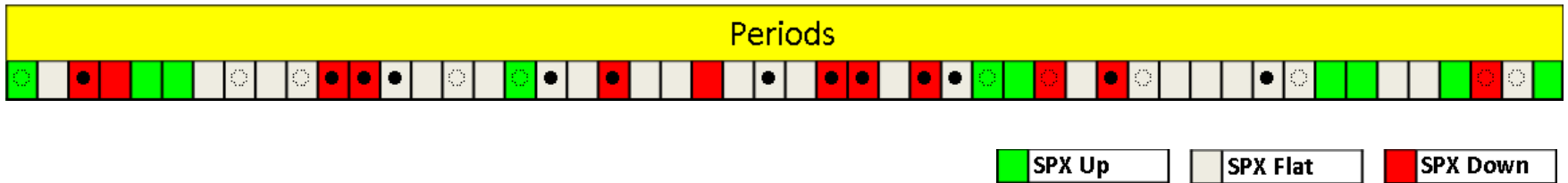
# Basic Algorithm: Leaf Level



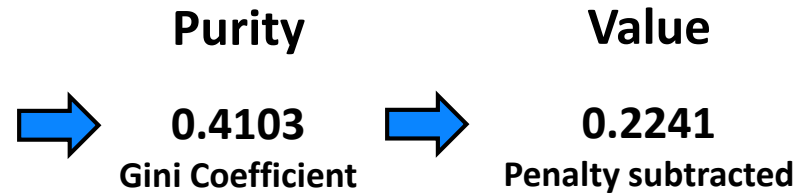
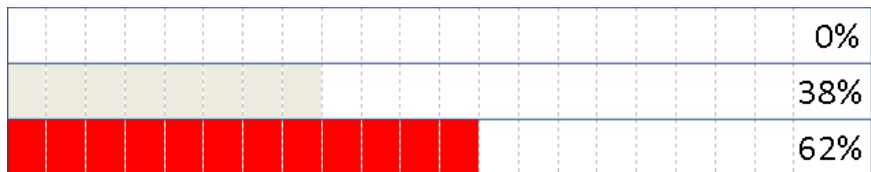
**P\_E\_RATIO > 17.1000**



# Basic Algorithm: Leaf Level



$P\_E\_RATIO > 17.1000$  &  $P\_C\_RATIO < 1.5989$



## Basic Algorithm: Leaf Level



## Basic Algorithm: Node Level

- **At every node, test every potential bifurcation point**
- **Determine how effectively the class is divided (“purity”) as a result of each side of the bifurcation. Adjust the two purities to achieve two values for this node**
- **For each side of the split, recursively invoke the tree generation function to produce a depth – 1 tree and select the tree with the highest value (if this is not a leaf node)**
- **If each sub-tree does not improve the value, ignore it. Otherwise, return the sub-tree and update the value**
- **Combine the two values**

# CUDA Implementation

- **Use recursive dynamic parallelism**
- **At each node, recursively invoke the tree generation function, which dynamically launches a kernel**
- **Dynamic launch of one block per potential bifurcation: each block builds a sub-tree**
- **Parent block picks the sub-tree which has the highest value**
- **CUDA automatically handles resource allocation; capacity never exceeded despite exponential dynamic launches**
- **Speed enables two or more additional levels of depth, depending on number of potential bifurcations**



# Shortcut Parameters

- **Exponential node proliferation: potential  $N^{depth}$  blocks launched, for number of potential bifurcation points N, represents serious time constraint, even given speed of CUDA. Shortcuts end processing early or enable more depths to be reached**
  - ❖ **maxDepth**                      **All nodes are leaf nodes at this depth**
  - ❖ **minPop**                              **Abort processing if a bifurcation results in too few periods**
  - ❖ **purityCeil**                          **Do not recursively generate sub-trees if purity is good enough**
  - ❖ **factorMaxUse**                      **Eliminate over-used factors from consideration**
  - ❖ **factorChoices**                    **Limit consideration of factors at any depth**
  - ❖ **factors**                                **Universe of input factors is, in itself, an implied limit**
  - ❖ **exhaustiveSearch**                **Perform exhaustive, random sampling or superficial search at each depth**

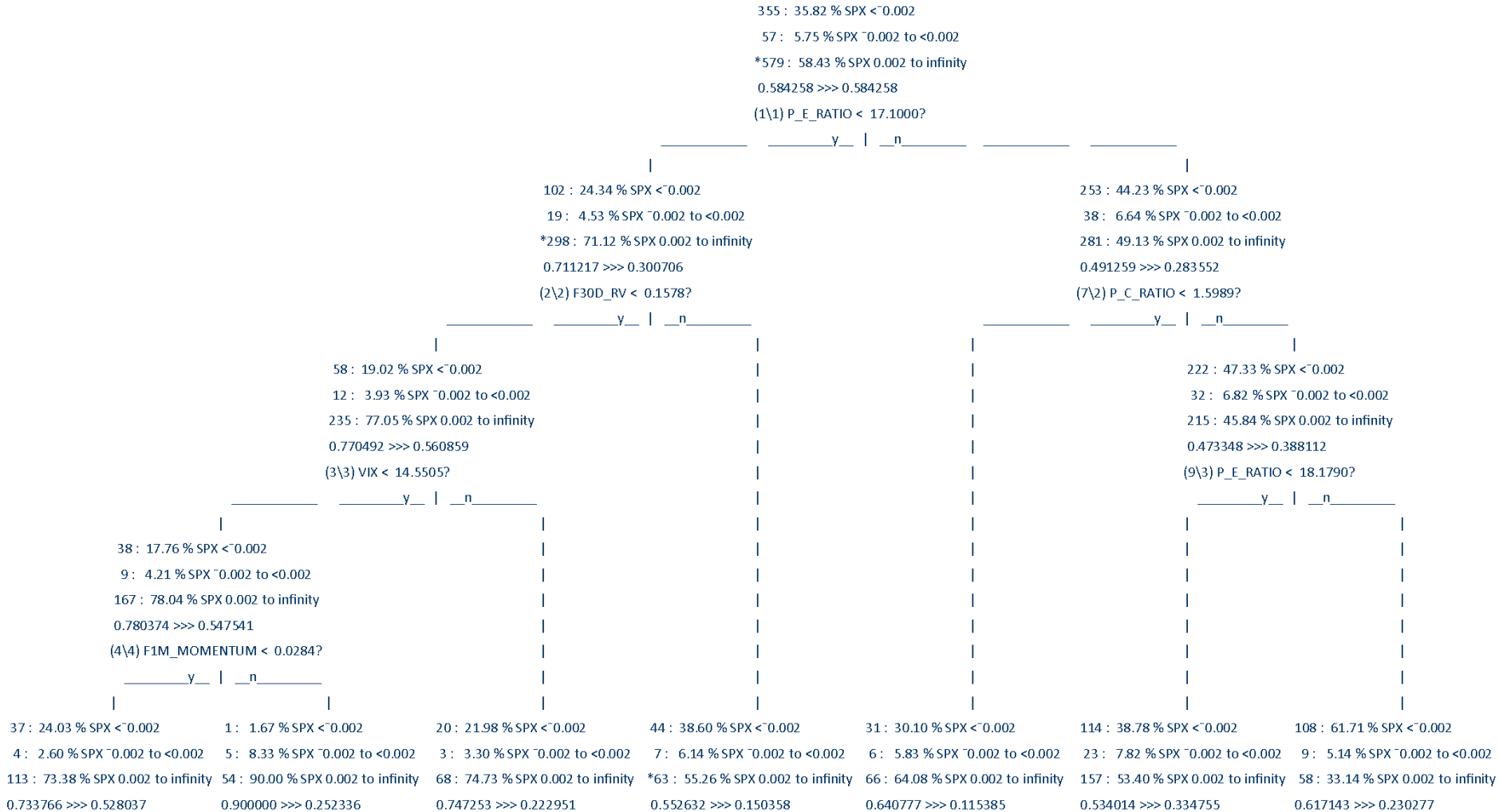
# Purity Rules

- **purityMethod** defines how the class-wise breakdown of a bifurcated set of periods is assessed. Sample methods of determining purity from populations of periods grouped by class:
  - ❖ **Top Weighting**      Purity is the weight of the class containing the most periods
  - ❖ **Gini Index**      Purity is variance among classes: sum of squares calculation
  - ❖ **Gini Coefficient**      Purity is average of all pairwise combinations of differences
  - ❖ **Entropy**      Purity is average of weights  $\times \log_2(\text{weights})$  of all classes
- **purityCombMethod** defines how to combine the purities of two sub-trees to produce an overall purity for the node. Sample methods are simple average, minimum, maximum, weighted average
- **Penalty parameters** control function to reduce importance of purity in valuing the various potential bifurcations at any node

# Control Parameters

- ❖ **purityMethod**      **Rule for determining purity of a bifurcation**
- ❖ **purityCombMeth**      **Rule for combining purities of two sub-nodes**
- ❖ **classUse**      **Classes to consider in purity calculation**
- ❖ **Factor divisions**      **How factors are divided is an implied control**
- ❖ **Class divisions**      **How classes are divided is an implied control**
- ❖ **Class time shift**      **How classes are time shifted is an implied control**
- ❖ **Penalty**      **Reduce effect of purity in valuing bifurcation**

# Decision Tree



# Summary

- **Augmented Decision Trees can facilitate identification of appropriate factors across regimes**
- **Leverage power of CUDA using recursive dynamic parallelism**
- **Augmented Optimized Decision Tree Models can be integrated into a broader deep learning framework creating models that are more well behaved over time**
- **Application in Investment and Risk Management**

## Author Biographies



- **Yigal D. Jhirad**, Senior Vice President, is Director of Quantitative and Derivatives Strategies and a Portfolio Manager for Cohen & Steers' options and real assets strategies. Mr. Jhirad heads the firm's Investment Risk Committee. Prior to joining the firm in 2007, Mr. Jhirad was an executive director in the institutional equities division of Morgan Stanley, where he headed the company's portfolio and derivatives strategies effort. He was responsible for developing, implementing and marketing quantitative and derivatives products to a broad array of institutional clients, including hedge funds, active and passive funds, pension funds and endowments. Mr. Jhirad holds a BS from the Wharton School. He is a Financial Risk Manager (FRM), as Certified by the Global Association of Risk Professionals.



- **Blay A. Tarnoff** is a senior applications developer and database architect. He specializes in array programming and database design and development. He has developed equity and derivatives applications for program trading, proprietary trading, quantitative strategy, and risk management. He is currently a consultant at Cohen & Steers and was previously at Morgan Stanley.