Machine Learning in Risk Modeling

History + Overview

Maroon Capital Board Presentation

Analyst: Matthew Li

Table of Contents

Previous Financial Risk Modeling Techniques

A mix of historical, probabilistic, and correlation analysis

Monte Carlo

- ↳ Invented in Monaco in the 1940s through roulette
- ↳ Basic principle is in ergodicity: the statistical behavior of a moving point in an enclosed system
- ↳ 3 main assumptions: returns are normally distributed, expected returns are constant over time, all return parameters are known[1]

Historical Simulation

- ↳ Using historical data to find probability distributions and risk correlations in past data
- ↳ Assumes that previous indicators of risk will persist

Parametric Models

- ↳ Specific probability distributions for inputs
  - e.g., normal, lognormal distributions
- ↳ Allow for further extrapolation beyond historical data through closed-form expressions

Value at Risk (VaR)

- ↳ Statistic that calculates the possible financial losses of a set of assets within a time frame
- ↳ Can be computed using historical, variance-covariance, and Monte Carlo methods
- ↳ Reduces chance of firm holding many highly correlated assets

1. https://macabacus.com/blog/financial-risk-modeling-management-strategies
2. https://aws.amazon.com/whatis/montecarlosimulation/#:~:text=The%20Monte%20Carlo%20simulation%20is%20a%20probabilistic%20model%20that%20can,home%20and%20office%20is%20fixed.

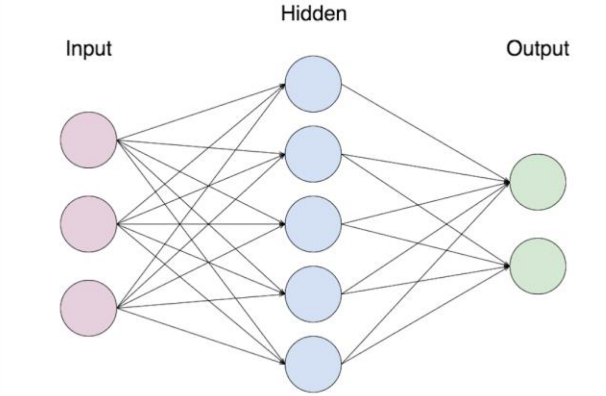
Machine Learning Techniques and Models

Machine Learning is a subset of statistics leading to revolutionary regressions and modeling

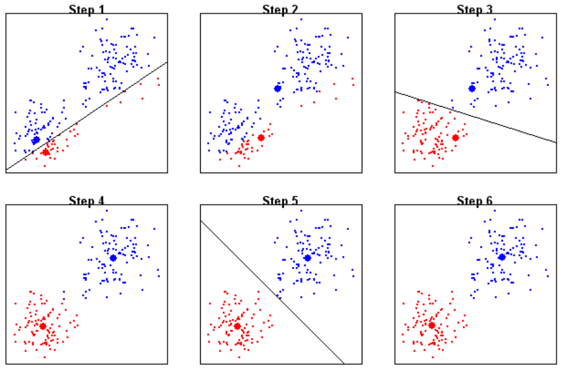
Important Techniques and Features

- Supervised learning – using multiple input variables to model out an output and check back for accuracy to revise model parameters
- Unsupervised learning – using data to predict and identify structures and patterns
- Better than linear regression since the models can point out non-linear relationships
- Linear methods include: partial least squares, principal component analysis
- Non-linear methods include: penalized regression, least absolute shrinkage and selection operator (LASSO), elastic nets
- Problem of overfitting for overly complex models

Neural Network Node Structure



K-Means Clustering Algorithm



1. https://www.iif.com/portals/0/Files/private/32370132_van_liebergen_-_machine_learning_in_compliance_risk_management.pdf

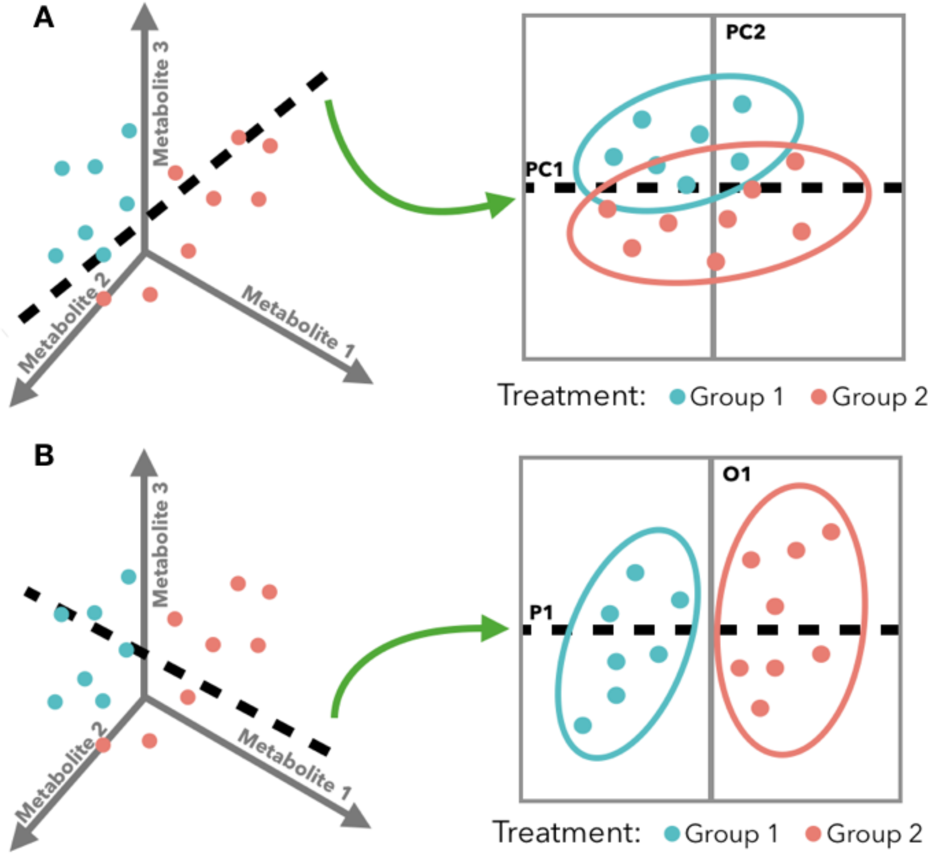
Supervised Linear Regression

Principal Components, Ridge, Partial Least Squares, LASSO

Basic Idea

- One of the most popular types of algorithms due to wide range of use cases
- Similar to regular statistical linear regression models
- Used to simulate mathematical relationship between variables for continuous predictors
- Principal Components Analysis (PCA) – used to represent a multivariate data table as a smaller set of variables to better observe trends
- Ridge Regression (L2) – regression across highly correlated variables using ridge estimators instead of ordinary least squares, creating lower, biased variance
- Partial Least Squares – similar to PCA, but instead of reducing dimensionality, it translates variables to a new space, making it a bilinear factor model

Partial Least Squares Example



1. https://ericrscott.com/project/pls-ecology/

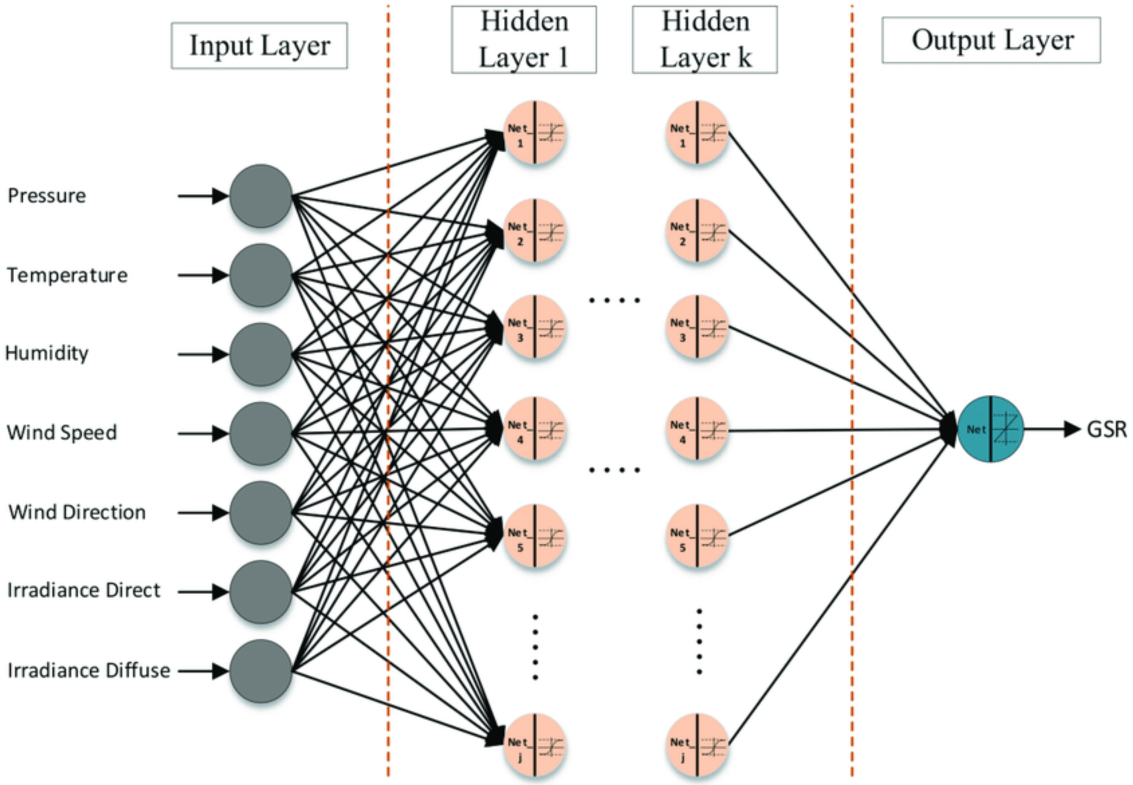
Supervised Non-Linear + Linear Regression

Penalized Regression (LASSO, LARS, Elastic Nets), Neural Networks, Deep Learning

Key Components and Difference from Linear

- Key difference is that the equation can include polynomial terms, interaction effects, and variable transformations
- LASSO (L1), although traditionally linear, can be used with polynomial terms to become non-linear
- LARS can also be fitted with non-linear models, targeting highly correlated predictors in a set
- Elastic Nets (L1 + L2) attempts to combine ridge and LASSO to regularize statistical models, also consider linear
- Neural Networks uses a structure of nodes to predict an outcome using various layer complexities
- Deep Learning is the more complex, higher-level version of neural networks

Neural Network Structure



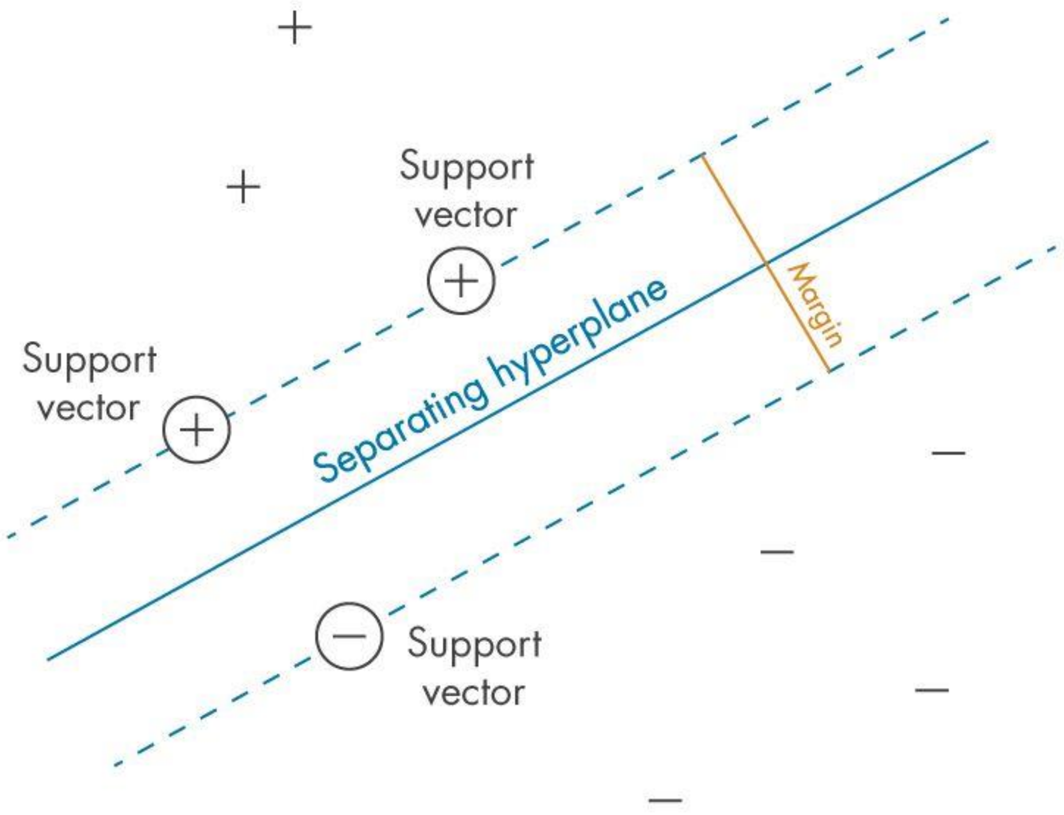
Supervised Linear Classification

Support Vector Machines

Key Features and Differences from Regression

- Classification groups into buckets whereas regression looks for a specific output value
- Support Vector Machines (SVMs) is an algorithm that splits data into groups and classifies data
  - Can be considered for regression as well and excels specifically in binary classification settings
- Theoretically, any unsupervised algorithm has its supervised counterpart (although not always recommended
  - e.g., K-means can be implemented with SVM to create a supervised algorithm

Support Vector Machine Diagram



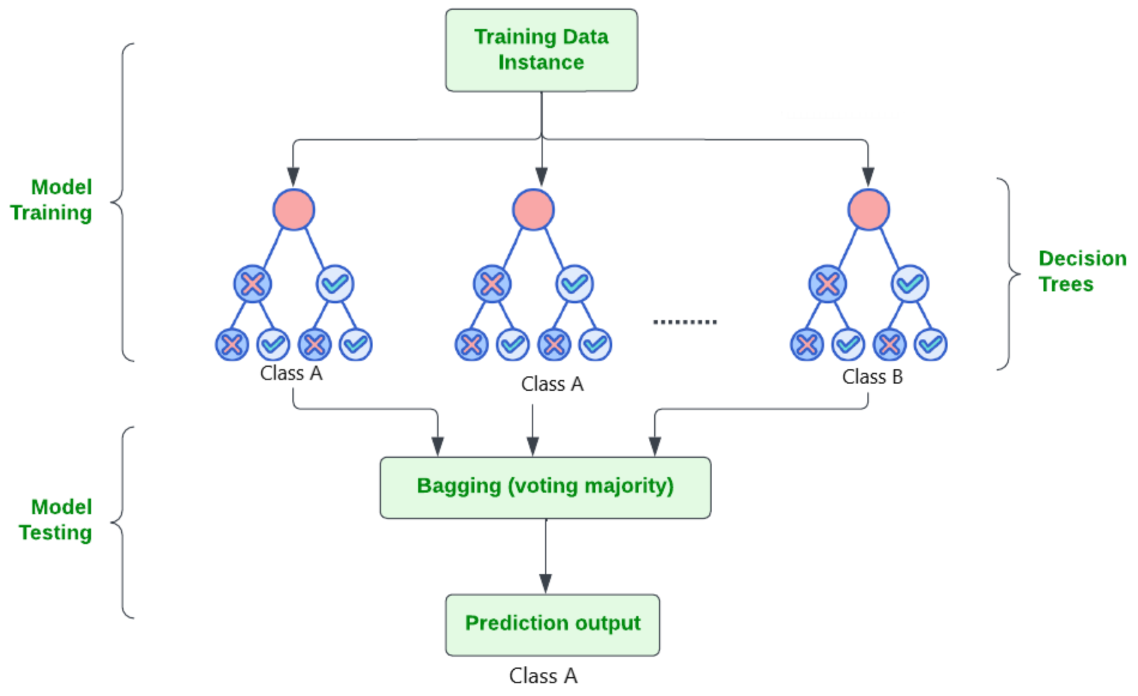
Supervised Non-Linear Classification

Decision Trees (Classification, Regression, Random Forest), Support Vector Machines, Deep Learning

Efficiencies with Non-Linear Classification

- Non-linear classification allows for more efficient, correct division of data
- Decision Trees are models that follow a tree like structure to determine various classifications
  - ID3: Iterative Dichotomiser 3 – utilizes entropy
  - C4.5: v2 of ID3 – uses information gain and gain ratios to evaluate split points within decision trees
  - CART: Classification and Regression Trees – utilizes Gini impurity to identify the best attributes to split itself on
  - Random Forest – utilizes a set of trees
- SVMs can be modified to become non-linear
- Deep Learning is typically non-linear since it can take into account multiple features and classify with weights

Random Forest Graph



1. https://hunch.net/~coms-4771/quinlan.pdf

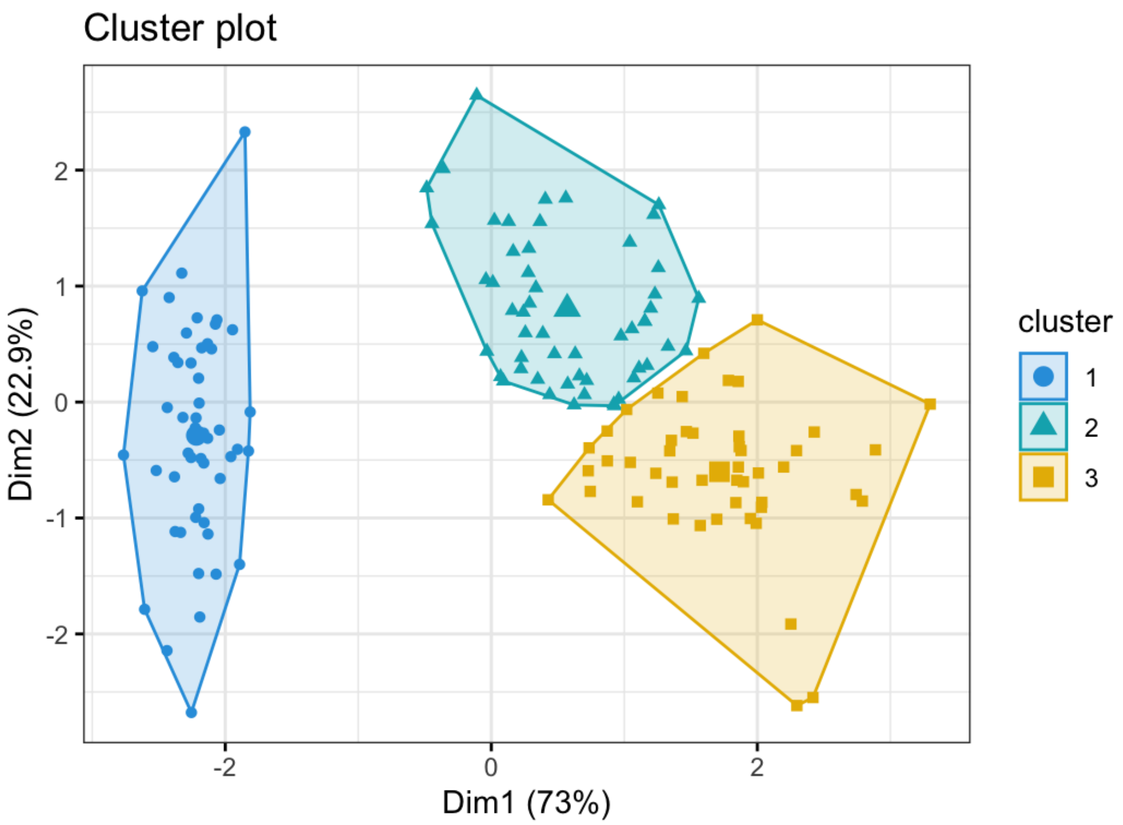
Unsupervised Clustering

Clustering Methods (K-, X-means, hierarchical), PCA, Deep Learning

Key Points and Advantages

- Unsupervised utilizes unlabeled data, which is significantly more prevalent
- K-means partitions n data points into k clusters with areas known as Voronoi cells, minimizing intra-cluster variance – mean squared error (MSE)
- X-means is an improved version of K-means with an improved local decision maker
  - Bayes Information Criterion
  - Akaike Information Criterion
- Hierarchical Agglomerative Clustering (HAC) uses a tree, a dendrogram, for group objects with general strategies including single-linkage and complete-linkage clustering (SLINK and CLINK, respectively)

K-means Graph



1. https://www.sciencedirect.com/science/article/pii/S1110866522000639#:~:text=X%2Dmeans%20%5B29%5D%20is,to%20obtain%20a%20better%20cluster.

Struggle with Complex Financial Instruments

Traditional Methods of Risk Analysis Fall Short on Capturing the True Essence of Instruments like Derivatives

Current Inefficiencies

- ↳ Data selection is slow and cross-correlation among explanatory variables is common
- ↳ Many key assumptions in models like Monte Carlo may not hold up in the real world
- ↳ Large portfolios have many cross-correlated assets and including these considerations is hard for non-Machine Learning algorithms
- ↳ Extremely complex results that can be hard to read, understand, and implement
- ↳ VaR often establishes a 99% confidence interval, leaving a 1% chance for a huge loss (inaccurate weighting)
- ↳ GIGO
- ↳ Many financial crises since the inception of VaR, the most widely used financial risk model, despite advances in its implementation

Machine Learning Solution

- ↳ Manages garbage in a little more efficiently and can exclude certain useless data
  - Void if the whole dataset is garbage
- ↳ Can perform more complex cross-correlation analyses through algorithm structure
  - Hidden layers in neural networks
  - Complex, adaptive categorization algorithms
  - Entropy-based decision making in tree structure
- ↳ Does not rely on previous assumptions
- ↳ Creates easier-to-read solutions

1. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8219786/

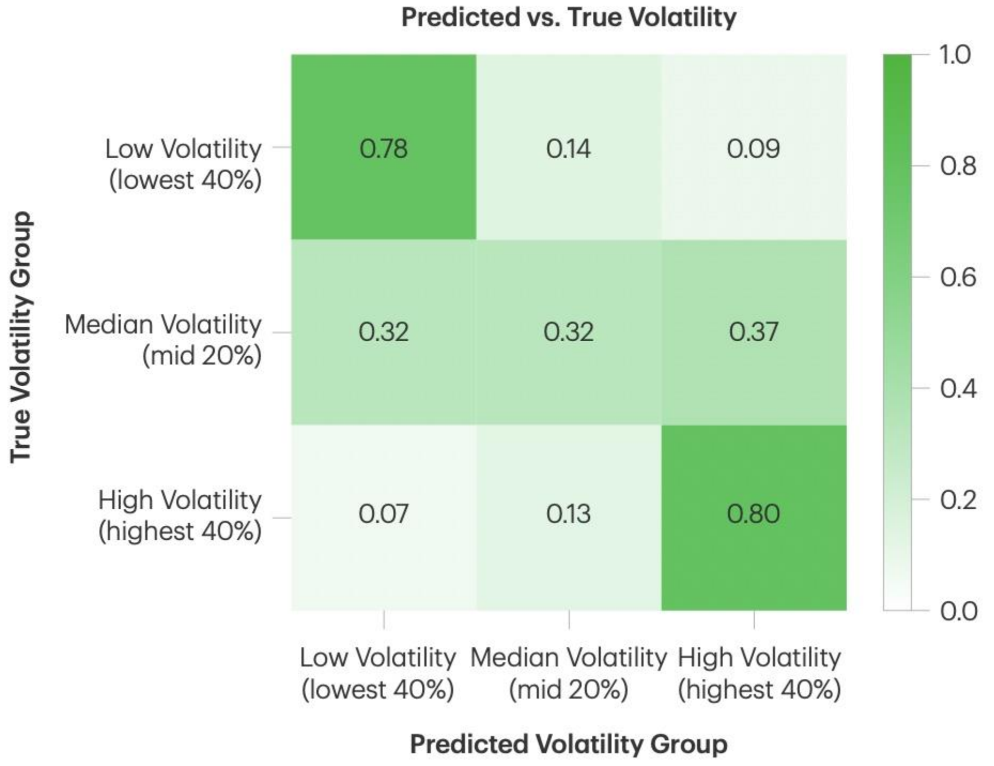
Case Study: TD Bank

TD Bank's Machine Learning Implementation in Risk Forecasting

Key Findings

- S&P standard deviation is ~14%
- After volatility reduction model, SD is ~11%
- After additional ML reduction, SD is ~10%
- ML models face significant success in picking reduced volatility assets for low volatility funds
- Idiosyncratic risk is independent and uncorrelated with risk model factors, which can be identified by ML models
- Limitation is that it reduces the size of the universe of investible stocks
  - More problematic in already smaller equity universes

Confusion Matrix from Machine Learning Model



Source: TDAM

1. https://www.td.com/us/en/asset-management/documents/pdf/TDAM-USA-Low-Vol-Machine-Learning.pdf

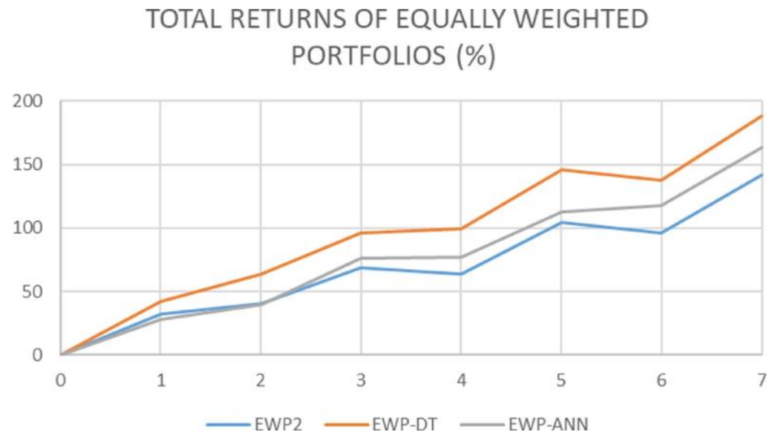
Case Study: AIRMS

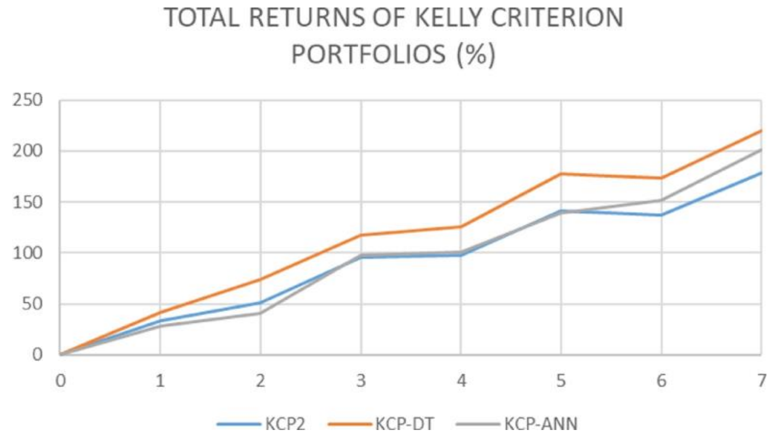
Artificial Intelligence Risk Management System (AIRMS)

Implementation and Findings

- Utilized artificial neural networks and decision trees on dynamic sliding windows in 5 major FOREX pairs to determine a breakout strategy from 2010-2016
  - Outperformed a SVM, genetic algorithm combination
  - Globally-optimal classification tree analysis (GO-CTA)
  - Limited at 20 features
- Applied to the enhanced equally weighted portfolio (EWP2) and enhanced Kelly criterion portfolio (KCP2), strategies that relate to SD
- Results showed a 50% increase in profit when using the suggestion from the machine learning models compared to regular strategies
- Further work could be done in other markets, implementing SVMs, more complex NNs (RNNs, LSTM, GRU)

Results





1. https://www.sciencedirect.com/science/article/pii/S0957417418301908?via%3Dihub

Questions?