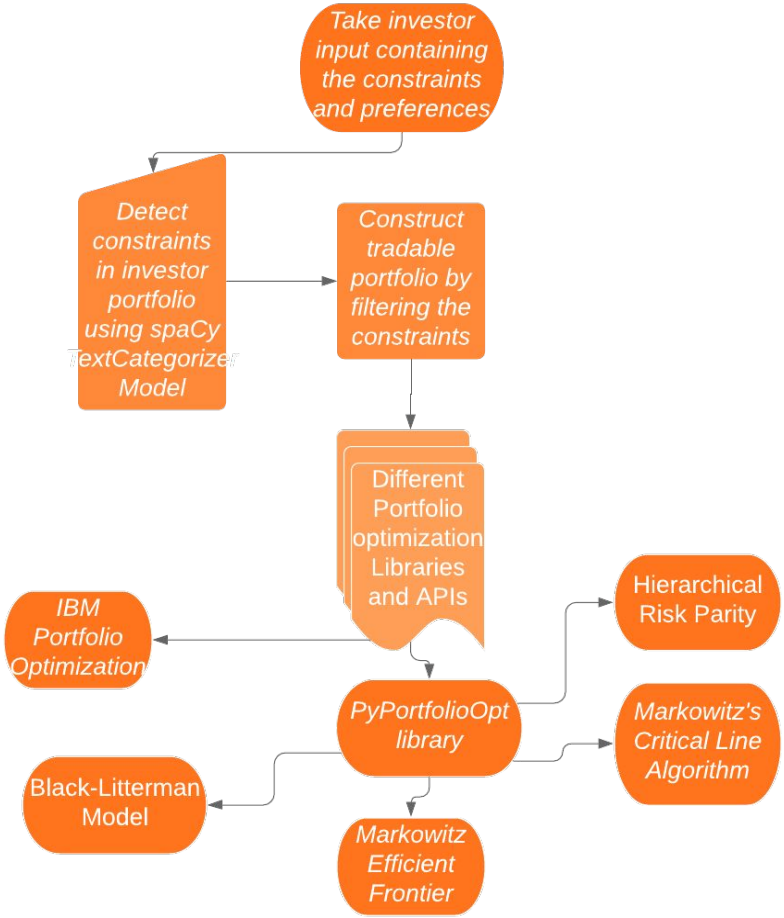
PORTFOLIO OPTIMIZATION USING NLP

MITACS RESEARCH PROJECT

KARTIK SETHI
3rd Year Undergraduate
Indian Institute of Technology, Guwahati

WORKFLOW


Take investor input containing the constraints and preferences
Detect constraints in investor portfolio using spaCy TextCategorizer Model
Construct tradable portfolio by filtering the constraints
Different Portfolio optimization Libraries and APIs
IBM Portfolio Optimization
Hierarchical Risk Parity
PyPortfolioOpt library
Markowitz's Critical Line Algorithm
Black-Litterman Model
Markowitz Efficient Frontier

Text Classifier

- Programmatically generated synthetics dataset of investor input
- Labelled the dataset as different constraints such as “Has Tobacco”,”Has military” etc.
- Used nltk library to generate different word synonyms
- Created a corpus of ~ 30,000 investor inputs
- Used a pre-trained spaCy text-classification model to predict different constraints
- Text classifier predicts the final constraints to be applied to the investor’s tradable portfolio

Portfolio Optimization using time series data

- Extracted time series data of the given stocks from Yahoo Finance
- Extracted exponentially weighted mean returns of the stocks from the time series data of stock prices
- Covariance matrix of the stocks constructed using time series analysis
- Generated Markowitz Portfolio maximising Sharpe Ratio

Different portfolio optimization models

- Efficient frontier optimisation via quadratic programming
- Hierarchical Risk Parity, using clustering algorithms to choose uncorrelated assets
  - From a universe of assets, form a distance matrix based on the correlation of the assets.
  - Using this distance matrix, cluster the assets into a tree via hierarchical clustering
  - Within each branch of the tree, form the minimum variance portfolio (normally between just two assets.
  - Iterates over each level, optimally combining the mini-portfolios at each node.
- Markowitz Critical Line Algorithm
  - Robust alternative to the quadratic solver used to find mean-variance optimal portfolios
  - Unlike generic quadratic optimisers, the CLA is specially designed for portfolio optimisation
  - Guaranteed to converge after a certain number of iterations, and can efficiently derive the entire efficient frontier

- Black-Litterman Allocation
  - Takes a Bayesian approach to asset allocation
  - It combines a prior estimate of returns (canonically, the market-implied returns) with views on certain assets, to produce a posterior estimate of expected returns.
  - Can provide views on only a subset of assets and BL will meaningfully propagate it, taking into account the covariance with other assets.
  - Using Black-Litterman posterior returns results in much more stable portfolios than using mean-historical return