Project Summary: Feature and Portfolio Bagging

a) Problem Background and Statement

Ensemble methods such as bagging are widely used in machine learning to improve prediction accuracy and robustness. In finance, similar ideas can be applied to both predictive modeling (feature bagging) and portfolio optimization (portfolio bagging). The project aims to investigate how aggregating models or portfolios constructed from random subsets of features or assets can enhance out-of-sample performance, reduce overfitting, and provide more stable solutions. The mathematical relationship between these two bagging approaches will be explored, with a focus on their application to financial data.

In the portfolio setting, bagging can be done with equal inclusion probabilities for all traded markets, with the purpose of enhancing robustness, but it can also be done with unequal inclusion probabilities. Such unequal probabilities can be set to reflect different degrees of liquidity in markets, and a bagged portfolio would then accordingly assign different levels of risk in different markets.

b) Expected Results

If successful, the project is expected to:

- Provide a theoretical and empirical understanding of how feature and portfolio bagging affect prediction accuracy, risk, and robustness in financial applications.
- Develop and analyze new ensemble estimators for precision/covariance matrices and portfolio weights.
- Identify optimal choices of bagging parameters (e.g., subset size k, number of bags B) and selection strategies.
- Compare the proposed methods to standard regularization and ensemble techniques.

c) Methods, Software, and Tools

The project will likely involve:

- Statistical and mathematical analysis (e.g., linear regression, Markowitz optimization, random matrix theory).
- Simulation studies and empirical evaluation on financial datasets.
- Implementation in Python (NumPy, pandas, scikit-learn, CVXPY) or MATLAB; use of Jupyter notebooks for reproducible research.
- Visualization and reporting using LaTeX and plotting libraries (matplotlib, seaborn).

d) Preferable Profile of Student(s)

- Strong background in mathematics, statistics, or applied mathematics.
- Familiarity with probability, linear algebra, and optimization.
- Experience with programming in Python or MATLAB.
- Interest in financial modeling, machine learning, or quantitative finance.
- Ability to work independently and communicate results clearly in writing.