# Sustainable ETF Performance and Volatility Predictor

**INTRODUCTION:**

An ETF (Exchange-Traded Fund) is a type of investment fund that is traded on stock exchanges like individual stocks. ETFs are made up of a collection of assets, such as stocks, bonds, or commodities, and they are designed to track the performance of a particular index or sector. For example, an ETF may track the S&P 500 index, which represents the performance of the 500 largest publicly traded companies in the US.

Machine learning techniques can predict and forecast stock returns. These predictions have the power to heavily influence the decisions of financial traders. However, because of the non-linearity of the stock market, the wide range of features and variables that influence the price of any stock or ETF, which could be political, economic, or natural, and the complex relationships between all of these factors, predicting any value regarding individual stocks, bonds, or ETFs is complex.

Through this project, we aim to build a machine-learning model that predicts ETF performance with high accuracy on the basis of past data.

**DATASET:** Stock prices from 2018 to the current fiscal quarter of multiple holdings along with iShares MSCI Global Impact ETF extracted from Yahoo  Finance

**Observations**: 7000
**Columns**: 6
- Open - Opening price of the ETF
- Close - Closing price of the ETF
- High - Highest price the ETF reached during that particular day
- Low - Lowest price the ETF reached during that particular day
- Adj close - closing price after adjustments for all applicable splits and dividend distributions
- Volume - Amount of ETF exchanged on that particular day
- Date

**Data Type:** Numeric
**Index:** Date

**PROGRAMMING LANGUAGE:** Python

**ROLES OF MEMBERS:**

Arshiya: Data extraction and pre-processing

Swapnil: Exploratory data analysis using both statistical and non-statistical methods

Raj: Feature selection and training data on different models

Rashi: Model optimization by tuning hyperparameters and evaluating test accuracy

Daisy: Prediction and comparisons of different models used

**SCHEDULE AND DELIVERABLES:**

03/30/23: Data pre-processing, weekly report

04/06/23: Exploratory data analysis, weekly report

04/13/23: Feature selection and model training, weekly report

04/20/23: Model optimization, testing, and editing the final draft of the final report

04/27/23: Presentation and demo

04/28/23: Final report submission