# Heilmeier catechism rubric answered:

## **1. What are you trying to do? Articulate your objectives using absolutely no jargon.**

The goal of this project is to predict whether the S&P 500, a major U.S. stock market index, will go up or down each day or week. To do this, we’ll look at different things that might affect the stock market, like the price of gold and oil, how much people are spending on credit cards, the value of the U.S. dollar, and currency exchange rates. By studying data from the past few years, we want to see if these factors can help us guess what will happen to the stock market in the future.

## **2. How is it done today, and what are the limits of current practice?**

**Current Practice**-

Analysts use broader economic data like inflation rates, interest rates, unemployment, and global events to assess market direction. Many professionals use a combination of these factors to predict market trends. Using combinations of all factors might mislead as some factors should affect more than others. So, our hybrid (Considering the factors making more affects) features should increase the accuracy of predictions. As well as these practices need to stay reliable on vast dataset which can make data noisy (Inconsistent data, unexpected spikes in data frequently).

**Identify and understand the current stakeholders: -**

* Investors
* Financial Analysts
* Models focusing on vast dataset and high frequency to make predictions
* Trading Platforms which analyze the data but lack deeper insight into data

**Limitations of current practices: -**

* *Narrow Focus*- Todays model are like conventional models relying mostly on macroeconomic factors but neglect some important factors such as Credit card usage, currency rates in comprehensive way

(What we did is considered non-conventional factors like credit card, crude oil movements as well as currency rates but only top five impacting on currency)

* *Data Complexity*: Financial analysts are using complex models, But they don’t have any access to easy-to-use models that integrate multiple data sources like gold, oil, and forex together.
* *Limited Integration*: Many current approaches are not fully integrating diverse macroeconomic factors such as consumer spending trends (e.g., credit card expenditure) into their predictions. They are tending to focus on simpler stock-related factors like price and volume.

By understanding these limitations, the value of our project lies in integrating a wider range of macroeconomic factors (e.g., gold, oil, forex rates, and credit card spending) into a simpler, more accessible predictive model. **This addresses gaps in current practices by offering a broader and more holistic approach to forecasting market movements**.

# **3. What is new in your approach and why do you think it will be successful?**

**What’s New in Our Approach:**

Our approach is different from traditional methods because we *combine a wide range of macroeconomic factors* that are often analyzed separately. Instead of focusing solely on stock prices, trading volume, or basic financial metrics, we integrate *gold prices, crude oil prices, forex rates, credit card expenditure data, and the strength of the U.S. dollar*which affect towards movement of USD, but it is neglected in most of the cases into a single model to predict the S&P 500’s daily or weekly movements. This broader view allows us to capture a more comprehensive understanding of what impacts the stock market.

**Innovations:**

* *Incorporating Consumer Spending Data*: We bring in credit card expenditure data, which is a useful reflection of consumer behavior. Underused concept.

Why It’s Innovative: Most existing models don’t have access to or don’t use this kind of granular, real-time consumer behavior data. By including it, we can better capture shifts in consumer sentiment and spending patterns, which are key drivers of stock market performance.

* *Multi-Source Macroeconomic Integration:* By simultaneously tracking commodity prices (gold, oil) and forex rates alongside traditional stock market indicators, we are better positioned to capture how global economic shifts affect the U.S. market. So, we are not resisting the predictive power of model.

Why It’s Innovative: Traditional models often give equal or fixed importance to various factors, but markets are dynamic. By allowing the model to adjust to real-world changes (Focusing more on factors with more importance), we can create a more responsive and accurate forecasting tool.

* *Predictive Risk Modelling*: We integrate risk modeling into our forecasting system to predict not just the direction of the S&P 500 but also the probability of high volatility or market risks. By tracking economic indicators that are linked to risk (such as large fluctuations in forex or commodity prices), our model can signal potential periods of instability or calm.

Why it’s Innovative: Most models focus on forecasting price movements without assessing potential volatility or risk. This additional layer of prediction can provide more useful insights, particularly for investors who want to manage risk, not just predict stock direction.

**Why you believe it will succeed:**

*Full Market Understanding:* By integrating factors from multiple sources (commodities, forex, consumer spending, and the dollar index), we are painting a more rounded picture of market movements. This broader approach allows us to capture trends that other models might overlook by focusing only on stock-related data.

*Adaptable to Market Conditions*: The ability to adjust weighting dynamically based on changing economic conditions ensures that the model remains relevant even as the macroeconomic landscape shifts. This gives it a significant edge over static models. Considering pandemic covid situation, this model will also be adaptable to shocking market situations.

*Risk-Aware Predictions*: By incorporating volatility and risk assessments, we offer predictions that aren’t just focused on price direction but also on the likelihood of market turbulence, which can help investors make better-informed decisions. Just like managing portfolio but focused on reducing losses in future than predicting profit in future.

*Proven Track Record of Machine Learning in Finance*: The success of machine learning in financial modeling has been well-documented in quantitative hedge funds and large institutions. But our approach democratizes access to this kind of advanced forecasting by building on proven techniques but making them more accessible and interpretable.

# **4. Who cares? If you are successful, what difference will it make?**

**Direst stakeholders:**

* Retail Investor: If our model is successful, it will provide investors with more accurate, allowing them to make better decisions about their Portfolio management. No need to be dependent on basic flows and charts.
* Financial Analysts: Analysts and portfolio managers can use the model to incorporate macroeconomic factors (e.g., gold prices, forex rates, and consumer spending data) so to enhance their market forecasts, allowing them to better manage portfolios.
* Financial Companies: FinTech platforms can integrate the model to enhance their existing predictive tools for investors by offering more robust forecasting features that go beyond stock price movements and include macroeconomic factors.

**Indirect Stakeholders:**

* Corporate Leaders: Corporate leaders might indirectly benefit from a better understanding of how macroeconomic trends affect the stock market so that their organization. This can help them in decision-making in terms of managing risk, capital allocation, and business strategy.

**Potential positive impact:**

* Improved Investment Strategies
* More Accurate Risk Management
* Greater Accessibility to Advanced Market Insights
* Comprehensive Economic Awareness

**Potential negative impact:**

* Over-reliance on the Model
* Market Manipulation/ Overfitting
* Disadvantaging Less experienced Investors

# 5. What are the risks and how will you address them?

**Data related risks:**

* For data, we have found the data, and, in our opinion, it is not incomplete or has any integration issues.

**Technical risks:**

* Given the variety and size of the datasets (multiple macroeconomic factors over a long time), there may be performance bottlenecks, particularly when handling large volumes of data when using complex machine learning models.

Solution: Use models like xGB which is a scalable library.

* As the data we have collected is from different API, websites. To integrate datasets from different sources can require more time. Or might lead to ineffienciency.

Solution: Use pipeline which can standardize or align the datasets.

**Methodological risks:**

* Overfitting: This was an identified problem when we were looking for datasets. Because if the model fails to generalize the unseen data, it relies too much on historical data, then overfitting may cause.

Solution: Use cross-validation techniques.

* Model Bias: We know that model is going to give less importance to forex trade and gold. As the historical influence of these factors is not considerable.

Solution: Need to discuss. It is not like this is confirmed that model is going to get biassed. This is just an assumption.

**Ethical Concerns**:

* Only if the model relies on the dataset which favors some sectors or regions. So, this could lead to unintentional misleading. As companies in S&P 500 are not identified top 500. They keep in changing. So, unintentional misleading may or may not happen.

Solution: Keep on auditing the model for different timestamps. This would make model make precise predictions.

# 6. How much would it cost, if implemented in real world?

If taken data around 20 years or 30 years it might need to use AWS or Google Cloud services.

As scope we have identified a problem that can we use Satellite Image data to get the details about the ports and then we can get to know about the exports and imports of the commodities. But the data for this is not freely available as open source. Api’s are expensive and we need around 5$ per picture.

We were able to find these problems regarding implementation in real world.

# 7. How long will it take?

We took one week for proposal, problem defining, collection of datasets.

And the further plan is:

Week 2,3: Data Preprocessing

Week 4,5: Model development

Week 6: Model evaluation, refining if required, Visualization

Week 7: Report writing, Presentation preparation.

As of the current plan we have decided among our group mates, this should take around more 6 weeks.

# **8. What is the mid-term and final "exams" to check for success?**

By the end of this month: The data should be cleaned, noisy data should not be present. Has same time for all the datasets. Minimal data after preprocessing.

Understand the correlation between feature and target factors. Still, build such model that will not be biased. Identify the trends, correlations between features (just for research).

Model should reach till its baseline performance. It should not be a weak model with limitations which can be identified.

Generate early hypotheses on which factors are most relevant for predicting S&P 500 movements, backed by model explainability tools like feature importance (just for research).

Final:

A fully trained, optimized model capable of predicting daily/weekly S&P 500 movements with documented performance metrics.

Final report.

A series of visualizations (e.g., line charts, scatter plots, feature importance charts) that explain the model’s predictions and insights.