

## **DSCI Final Project Proposal**

Aditya Venkataramani

### **1. Name of Project & Team Members**

- Name of Project: **Data Underminer**
- Team Members: **Aditya Venkataramani**

### **2. What problem are you trying to solve?**

The problem I am trying to solve is: The primary objective of this project is to evaluate the effectiveness of linear regression and LSTM neural network models in predicting daily 10-year Treasury yields using macroeconomic and financial data. A secondary goal is to directly compare the predictive accuracy of these two models to identify which approach provides more reliable insights for forecasting daily yield trends.

### **3. How will you collect data and from where?**

#### Data Collection Sources:

- **U.S. Treasury Website** – [Link](#)
  - o Daily 10-Year Treasury Yield: Web scraped using Beautiful Soup to capture daily yield data.
- **FRED API** – [Link](#)
  - o Macroeconomic Indicators: CPI (inflation), GDP growth rate, unemployment rate, federal funds rate.
- **Yahoo Finance API** – [Link](#)
  - o Financial Market Data: Daily S&P 500, crude oil prices for broader market context.
- **U.S. Bureau of Economic Analysis (BEA)** – [Link](#)
  - o Additional Economic Indicators: Personal consumption expenditures (PCE), real income data.
- **BLS (Bureau of Labor Statistics)** – [Link](#)
  - o Labor Market Data: Monthly labor force participation rate, employment data.

### **4. What analysis will you do and what visualizations will you create?**

- **Exploratory Data Analysis (EDA):** Summary statistics, correlation analysis, and trend decomposition to understand relationships between yield and economic indicators.
- **Modeling:** Train and compare linear regression and LSTM models to predict daily 10-year Treasury yields, evaluating each model's performance on accuracy and directional trends.

#### Visualizations:

- **Time Series Plots:** Actual vs. predicted yields for both models to compare accuracy.
- **Correlation Heatmap:** Relationships between macroeconomic indicators and yield.
- **Residual and Error Plots:** Model residuals and error distributions to assess prediction quality.
- **Rolling Error Plot:** Track model errors over time to highlight performance stability.