**Project Title**

**“Illuminating Solar Markets: A/B Testing and Predictive Models for Renewable Energy Insights”**

**Project Description and Justification**

**This project provides an in-depth analysis of the solar energy sector by leveraging advanced data analytics, statistical methods, and machine learning. Key components include:**

* **Exploratory Data Analysis (EDA): Uncovering trends in stock performance, ESG scores, and macroeconomic indicators.**
* **A/B Testing: Comparing the performance of two major solar companies, First Solar (FSLR) and Enphase Energy (ENPH), to evaluate strategies and market dynamics.**
* **Predictive Models: Employing time series, regression, and machine learning models to forecast stock trends and industry-wide performance.**
* **Risk Analysis: Conducting financial risk evaluation using Value at Risk (VaR), Conditional VaR, and scenario analysis.**
* **Portfolio Optimization: Applying Modern Portfolio Theory (MPT) to recommend optimal investment allocations.**
* **Sentiment and ESG Analysis: Analyzing sentiment data from news and social media, combined with ESG scores, to assess correlations with stock performance.**

**Planning**

**Task Descriptions**

1. **Data Collection:**
   * **Fetch historical stock data for 10–15 major solar companies.**
   * **Collect ESG metrics and macroeconomic indicators.**
   * **Scrape news and social media data for sentiment analysis.**
2. **Data Preprocessing and Storage:**
   * **Clean and preprocess all datasets.**
   * **Store structured data in PostgreSQL using an organized schema.**
3. **EDA:**
   * **Perform a detailed analysis of stock trends, ESG metrics, and sentiment correlations.**
   * **Use Plotly for visualizing key findings.**
4. **A/B Testing:**
   * **Conduct statistical tests to compare the performance of FSLR and ENPH.**
   * **Visualize results to highlight key insights.**
5. **Predictive Modeling:**
   * **Implement time series models, regression analysis, and ensemble machine learning techniques.**
   * **Forecast stock performance for the entire solar energy sector.**
6. **Risk and Portfolio Analysis:**
   * **Evaluate financial risks using VaR, Conditional VaR, and scenario analysis.**
   * **Optimize portfolios with MPT to recommend investment allocations.**
7. **Visualization and Reporting:**
   * **Design dashboards in Tableau for an executive summary.**
   * **Document findings in a comprehensive report.**

**Minimum Viable Product Features**

* **Clean and structured datasets stored in PostgreSQL.**
* **EDA of solar companies’ stock performance and trends.**
* **Results from A/B testing for FSLR and ENPH.**
* **A time series model for solar stock performance forecasting.**
* **An executive dashboard summarizing key metrics in Tableau.**

**Additional Features**

* **Risk Analysis: Including VaR, Conditional VaR, and scenario analysis.**
* **Portfolio Optimization: Applying Modern Portfolio Theory (MPT).**
* **Sentiment Analysis: Scraping and analyzing news and social media data.**
* **Comprehensive ESG Analysis: Assessing sustainability performance and trends.**
* **Advanced Predictive Models: Utilizing ensemble machine learning for better accuracy.**

**Tools and Frameworks**

|  |  |
| --- | --- |
| **Tool/Framework** | **Included** |
| **Database** | **Yes** |
| **SQL** | **Yes** |
| **Pandas** | **Yes** |
| **Machine Learning** | **Yes** |
| **Data Visualization** | **Yes** |
| **Tableau** | **Yes** |
| **Power BI** | **Maybe** |
| **Google Colab** | **Yes** |

**Deployment Platform**

* **Google Colab: For analysis and visualization preparation.**
* **Tableau: For creating executive summary dashboards.**

**Here is the updated glossary:**

**Glossary for "Illuminating Solar Markets: A/B Testing and Predictive Models for Renewable Energy Insights"**

**A. Key Concepts**

1. **Exploratory Data Analysis (EDA):**
   * **The process of analyzing datasets to summarize their main characteristics, often using visualizations.**
2. **A/B Testing:**
   * **A statistical method to compare two groups (e.g., companies) by testing metrics such as returns or ESG scores to determine significant differences.**
3. **Time Series Analysis:**
   * **Analytical techniques used to study datasets that are time-ordered, such as stock prices.**
4. **Regression Models:**
   * **Statistical models that examine the relationship between a dependent variable and one or more independent variables.**
5. **Machine Learning Models:**
   * **Algorithms that learn from data to make predictions or identify patterns (e.g., LSTM, XGBoost).**
6. **Value at Risk (VaR):**
   * **A risk management metric that estimates the potential loss in value of an asset or portfolio over a defined period at a certain confidence level.**
7. **Modern Portfolio Theory (MPT):**
   * **A framework for constructing an optimal portfolio that minimizes risk for a given return or maximizes return for a given level of risk.**
8. **Sentiment Analysis:**
   * **The use of Natural Language Processing (NLP) to determine the sentiment (positive, neutral, or negative) expressed in text, such as news articles or social media posts.**
9. **Environmental, Social, and Governance (ESG):**
   * **Metrics that assess a company's sustainability and societal impact.**
10. **Scenario Analysis:**
    * **A method of analyzing possible future events by considering alternative outcomes or scenarios (e.g., energy price fluctuations).**

**B. Indicators and Assets**

1. **Core Companies:**
   * **FSLR (First Solar):**
     + **A U.S.-based company specializing in manufacturing thin-film solar panels and providing utility-scale solar solutions. Known for its focus on sustainability and efficiency.**
   * **ENPH (Enphase Energy):**
     + **A leading provider of microinverter technology and energy management solutions. Enphase is recognized for its innovation in solar energy storage and grid independence.**
   * **SEDG (SolarEdge Technologies):**
     + **Focused on smart energy solutions, including power optimizers and inverters. SolarEdge is a global leader in residential and commercial solar systems.**
   * **CSIQ (Canadian Solar):**
     + **A Canadian company that manufactures solar PV modules and develops large-scale solar projects globally. It is a major player in the renewable energy market.**
   * **RUN (Sunrun):**
     + **A U.S.-based company specializing in residential solar installations and battery storage solutions. Sunrun is a key player in the home solar market.**
   * **JKS (JinkoSolar):**
     + **One of the largest solar panel manufacturers in the world, headquartered in China. JinkoSolar supplies high-efficiency panels for residential, commercial, and utility-scale projects.**
   * **NEE (NextEra Energy):**
     + **A leading clean energy company in the U.S., heavily invested in wind and solar power generation. NextEra Energy operates through its subsidiary Florida Power & Light.**
   * **TPIC (TPI Composites):**
     + **A U.S.-based manufacturer of composite wind blades and a supplier to renewable energy industries, including solar projects.**
   * **ORA (Ormat Technologies):**
     + **A company specializing in geothermal energy and energy storage solutions. Ormat contributes to diversifying the renewable energy mix.**
   * **MAXN (Maxeon Solar):**
     + **A spin-off from SunPower, Maxeon Solar focuses on manufacturing high-efficiency solar panels and expanding its presence in international markets.**
2. **Renewable Energy ETFs:**
   * **ICLN: iShares Global Clean Energy ETF.**
   * **PBW: Invesco WilderHill Clean Energy ETF.**
   * **TAN: Invesco Solar ETF.**
3. **Energy Sector ETFs:**
   * **XLE: Energy Select Sector SPDR ETF.**
   * **VDE: Vanguard Energy ETF.**
4. **Broader Market and Risk Indicators:**
   * **SPY: SPDR S&P 500 ETF Trust (tracks the S&P 500 index).**
   * **QQQ: Invesco QQQ ETF (tracks the Nasdaq-100 index).**
   * **VIX: CBOE Volatility Index ("fear gauge" for market volatility).**
5. **Commodities and Macro Indicators:**
   * **CL=F: Crude Oil Futures.**
   * **NG=F: Natural Gas Futures.**
   * **DXY: US Dollar Index.**
   * **GC=F: Gold Futures.**
   * **SI=F: Silver Futures.**
6. **Other Indicators:**
   * **EEM: iShares MSCI Emerging Markets ETF.**
   * **LIT: Global X Lithium & Battery Tech ETF.**
   * **URA: Global X Uranium ETF.**
   * **HYG: iShares iBoxx $ High Yield Corporate Bond ETF.**
   * **EURUSD=X: Euro to US Dollar exchange rate.**
   * **GBPUSD=X: British Pound to US Dollar exchange rate.**

**C. Tools and Platforms**

1. **Google Colab:**
   * **A cloud-based platform for writing and executing Python code, providing GPU/TPU support.**
2. **Tableau:**
   * **A visualization tool used to create polished, executive-level dashboards.**
3. **PostgreSQL (pgAdmin):**
   * **A relational database management system used for storing and querying data efficiently.**
4. **Python Libraries:**
   * **Pandas: For data manipulation and analysis.**
   * **Plotly: For creating interactive visualizations.**
   * **Statsmodels: For statistical modeling.**
   * **scikit-learn: For machine learning models.**
   * **NLTK, VADER, TextBlob: For sentiment analysis.**

**D. Key Metrics and Calculations**

1. **Stock Metrics:**
   * **OHLC: Open, High, Low, Close prices.**
   * **Adjusted Close: Reflects stock splits/dividends.**
   * **Volatility: Measured by rolling standard deviation.**
2. **Sentiment Metrics:**
   * **Sentiment Score: A numerical value representing positivity, neutrality, or negativity.**
3. **ESG Metrics:**
   * **Environmental Score: Measures environmental impact.**
   * **Social Score: Evaluates social responsibility.**
   * **Governance Score: Assesses corporate governance practices.**
4. **Financial Metrics:**
   * **Beta: Sensitivity of a stock to market movements.**
   * **Sharpe Ratio: Measures return per unit of risk.**
   * **Alpha: Excess return compared to a benchmark index.**

**Refined 4-Day Plan**

**Day 1: Data Collection**

**Goal: Fetch and organize all datasets into separate Google Colab notebooks.**

1. **Tasks:**
   * **Fetching\_Solar\_Stock.ipynb:**
     + **Fetch historical stock data for the 10–15 solar companies.**
     + **Perform initial cleaning and validation.**
   * **Fetching\_ESG\_Data.ipynb:**
     + **Collect ESG metrics for the selected companies.**
   * **Fetching\_Macro\_Indicators.ipynb:**
     + **Retrieve macroeconomic indicators and broader market data.**
   * **Fetching\_Sentiment\_Data.ipynb:**
     + **Scrape news and social media data for sentiment analysis.**
2. **Deliverables:**
   * **Four Google Colab notebooks, each containing the respective data-fetching scripts.**
   * **Initial data outputs saved as CSV files for later processing.**

**Day 2: Data Preprocessing and Storage**

**Goal: Clean, preprocess, and store data in PostgreSQL.**

1. **Tasks:**
   * **Process datasets within their respective Google Colab notebooks:**
     + **Handle missing values, validate data, and standardize formats.**
     + **Save processed data to CSV files.**
   * **Database Setup:**
     + **Create a structured PostgreSQL schema.**
     + **Import the cleaned datasets into their respective tables:**
       - **solar\_stock\_data**
       - **esg\_metrics**
       - **macro\_indicators**
       - **sentiment\_data**
2. **Deliverables:**
   * **Fully processed datasets saved in PostgreSQL.**
   * **Google Colab notebooks updated with cleaning and database upload scripts.**

**Day 3: Exploratory Data Analysis (EDA)**

**Goal: Analyze data trends and uncover key insights.**

1. **Tasks:**
   * **EDA for Solar Stocks:**
     + **Analyze stock trends, volatility, and rolling averages.**
   * **EDA for ESG Metrics:**
     + **Compare ESG scores across companies.**
     + **Correlate ESG scores with stock performance.**
   * **EDA for Sentiment Data:**
     + **Assess sentiment trends over time and correlate with stock trends.**
   * **EDA for Macroeconomic Indicators:**
     + **Identify correlations between macro indicators and stock performance.**
   * **Visualize all findings using Plotly.**
2. **Deliverables:**
   * **One Google Colab notebook: Exploratory\_Data\_Analysis.ipynb.**
   * **Interactive visualizations summarizing key insights.**

**Day 4: Advanced Analysis and Modeling**

**Goal: Conduct A/B testing, build predictive models, and perform risk analysis.**

1. **Tasks:**
   * **A/B Testing:**
     + **Compare FSLR and ENPH performance using statistical tests.**
     + **Visualize differences in performance metrics (e.g., returns, volatility).**
   * **Predictive Models:**
     + **Implement time series models (e.g., ARIMA, Prophet) for solar stock forecasts.**
     + **Use regression and machine learning models to predict broader market trends.**
   * **Risk and Portfolio Analysis:**
     + **Perform VaR, Conditional VaR, and scenario analysis.**
     + **Optimize portfolios using MPT and create an efficient frontier plot.**
2. **Deliverables:**
   * **Two Google Colab notebooks:**
     + **AB\_Testing\_and\_Predictive\_Models.ipynb**
     + **Risk\_and\_Portfolio\_Analysis.ipynb**
   * **Initial results saved for use in dashboards and reports.**

**Summary of Google Colab Notebooks**

1. **Fetching\_Solar\_Stock.ipynb**
2. **Fetching\_ESG\_Data.ipynb**
3. **Fetching\_Macro\_Indicators.ipynb**
4. **Fetching\_Sentiment\_Data.ipynb**
5. **Exploratory\_Data\_Analysis.ipynb**
6. **AB\_Testing\_and\_Predictive\_Models.ipynb**
7. **Risk\_and\_Portfolio\_Analysis.ipynb**

Here’s a quick glossary of the engineered features for the solar company dataset:

**Glossary of Engineered Features:**

1. **Daily\_Range**:
   * **Definition**: The difference between the highest (High) and lowest (Low) prices on a given day.
   * **Formula**: High - Low.
   * **Insight**: Measures daily price volatility.
2. **Normalized\_Range**:
   * **Definition**: The daily range as a percentage of the closing price.
   * **Formula**: (High - Low) / Close.
   * **Insight**: Normalizes volatility relative to the stock's price.
3. **Daily\_Return**:
   * **Definition**: The percentage change between the opening and closing prices.
   * **Formula**: (Close - Open) / Open.
   * **Insight**: Indicates daily performance.
4. **MA\_7**:
   * **Definition**: 7-day moving average of the closing price.
   * **Formula**: Average(Close over last 7 days).
   * **Insight**: Highlights short-term trends.
5. **MA\_30**:
   * **Definition**: 30-day moving average of the closing price.
   * **Formula**: Average(Close over last 30 days).
   * **Insight**: Captures longer-term trends.
6. **Momentum**:
   * **Definition**: The change in closing price compared to the previous day.
   * **Formula**: Close - Close (previous day).
   * **Insight**: Measures price movement strength and direction.
7. **Volume\_Change**:
   * **Definition**: The percentage change in trading volume compared to the previous day.
   * **Formula**: (Volume - Volume (previous day)) / Volume (previous day).
   * **Insight**: Highlights changes in trading activity.
8. **Volume\_to\_Price**:
   * **Definition**: The ratio of trading volume to the closing price.
   * **Formula**: Volume / Close.
   * **Insight**: Indicates trading intensity relative to price.
9. **Lag\_Close\_1**:
   * **Definition**: The closing price from the previous day.
   * **Formula**: Close (previous day).
   * **Insight**: Provides historical context for current price.
10. **Lag\_Close\_7**:
    * **Definition**: The closing price from 7 days ago.
    * **Formula**: Close (7 days ago).
    * **Insight**: Highlights weekly trends.
11. **Weighted\_Avg\_Price**:
    * **Definition**: The volume-weighted average price for a stock on a given day.
    * **Formula**: Close \* Volume.
    * **Insight**: Combines price and trading activity.
12. **Market\_WAP**:
    * **Definition**: The overall market's volume-weighted average price for the day.
    * **Formula**: Sum(Weighted\_Avg\_Price) / Sum(Volume).
    * **Insight**: Reflects the market's average trading price.
13. **Market\_WAP\_MA\_30**:
    * **Definition**: 30-day moving average of the market's weighted average price.
    * **Formula**: Average(Market\_WAP over last 30 days).
    * **Insight**: Indicates long-term market trends.
14. **Market\_Trend**:
    * **Definition**: Identifies whether the market is in a bull or bear phase.
    * **Criteria**: Bull if Market\_WAP > Market\_WAP\_MA\_30, otherwise Bear.
    * **Insight**: Provides a high-level view of market sentiment.