Capstone Project Proposal

“Stocky” Project

<Aseel Meshal AlKhadaidi>

Business Goals

Project Overview and Goal What is the industry problem you are trying to solve? Why use ML/AI in solving this task? Be as specific as you can when describing how ML/AI can provide value. For example, if you’re labeling images, how will this help the business?	**The industry problem**: Many face a challenge to manage vast amounts of financial data and marking timely, accurate investment decisions. Investors face information overload, emotional biases, and inefficiencies that can impact their returns. **Why using ML/AI?** It can process large datasets, predict market trends, and automate investment decisions faster & more accurately than humans. This will reduce biases, improve response times, and adapt to market changes, which will lead to optimized investment strategies. **What Value Can ML/AI provide?** It’ll provide real-time data analysis, predictive insights, and automated trading, which will lead to better returns and Improved risk management. Since it’ll be a website I’ll also focus on enhancing user experience on the website, ensuring the interface is intuitive and responsive.
Business Case Why is this an important problem to solve? Make a case for building this product in terms of its impact on recurring revenue, market share, customer happiness and/or other drivers of business success.	**Importance:** Automating investment decisions with ML/AI addresses the critical need for faster, unbiased, and data-driven strategies in a competitive financial market. **Impact on Business:** 1. **Recurring Revenue:** An AI-driven investment model can attract a broader customer base, including those seeking efficient, automated solutions, leading to higher subscription or management fees.

	2. **Market Share**: Offering cutting-edge AI-driven investment services can differentiate the business, attracting tech-savvy investors and increasing market share. 3. **Customer Satisfaction:** By delivering consistent returns and minimizing risks, the AI model enhances customer trust and loyalty, boosting retention rates. 4. **Operational Efficiency:** Automating the investment process reduces costs associated with manual analysis, improving profit margins. 5. **Enhanced User Engagement:** A web-based platform can attract a wider audience by offering easy access, which can further drive recurring revenue and customer satisfaction. 6. **Accessibility:** Highlight how a web-based model can be accessed from anywhere, making it more attractive to users.
Application of ML/AI What precise task will you use ML/AI to accomplish? What business outcome or objective will you achieve?	**Precise Task:** The ML/AI model will analyze financial data, predict market trends, and automate investment decisions by buying and selling assets based on predefined strategies. **Business Outcome:** The objective is to achieve higher returns on investments, minimize risk, and provide a scalable, automated solution that enhances customer satisfaction and increases recurring revenue through consistent portfolio performance. **User Interaction:** The AI model must deliver real-time recommendations and updates efficiently through the website interface. **Scalability:** The model must be optimized for multiple concurrent users without significant delays.

Success Metrics

Success Metrics What business metrics will you apply to determine the success of your product? Good metrics are clearly defined and easily measurable. Specify how you will establish a baseline value to provide a point of comparison.	**Portfolio Return on Investment (ROI):** The percentage gain or loss on the portfolio managed by the AI model over a specific period. Baseline: Compare with historical returns of a traditional portfolio or market benchmark (e.g., S&P 500) over the same period. **Customer Retention Rate:** Definition: The percentage of customers who continue using the AI investment service over time. Baseline: Use the retention rate of current customers using traditional investment methods. **Under Management (AUM)** Growth: Definition: The increase in the total value of assets managed by the AI model. Baseline: Compare with the initial AUM at product launch and track quarterly growth. **Subscription/Service Adoption Rate:** Definition: The rate at which new customers sign up for the AI-driven investment service. Baseline: Measure against initial projections or industry averages for new financial service products. **Cost Reduction in Operations:** Definition: The percentage decrease in operational costs due to automation and reduced manual intervention. Baseline: Compare with pre-automation costs associated with traditional investment management. **User Engagement Metrics:** In addition to the financial performance metrics, I'd also track website-specific metrics like user retention, session duration, and conversion rates (e.g., how many users take the suggested actions). **Load Time and Uptime:** Ensuring the site loads quickly and is consistently available would also be crucial to success.

Data

Data Acquisition Where will you source your data from? What is the cost to acquire these data? Are there any personally identifying information (PII) or data sensitivity issues you will need to overcome? Will data become available on an ongoing basis, or will you acquire a large batch of data that will need to be refreshed?	**Data Sources:** • Market Data: Stock prices, trading volumes, financial statements from sources like Bloomberg, Yahoo Finance, or Alpha Vantage. • News & Sentiment: Financial news, social media feeds (e.g., Twitter), and sentiment analysis tools. • Economic Indicators: Macroeconomic data from government agencies or financial institutions. **Cost:** Data acquisition costs vary; free APIs provide basic data, while premium services like Bloomberg can be expensive, depending on the level of detail and frequency of updates. **PII and Data Sensitivity:** The model primarily uses non-PII data, such as market prices and news articles, minimizing privacy concerns. If customer data is used for personalization, strict data protection and anonymization protocols will be required. **Data Availability:** Data will be acquired on an ongoing basis, with continuous real-time updates needed for market data, sentiment analysis, and economic indicators to ensure timely and accurate decision-making. **Real-Time Data Streaming:** Data needs to be integrated seamlessly into the website, with minimal latency to ensure real-time updates **Privacy Concerns:** Emphasize the importance of securely handling user data on the website, especially if PII is involved.
Data Source Consider the size and source of your data; what biases are built into the data and how might the data be improved?	**Potential Biases:** 1. **Historical Bias:** Past performance may not predict future results accurately. The model might overfit historical data. 2. **News Bias:** Sentiment analysis tools may have biases based on the sources or algorithms used, leading to skewed sentiment scores.

	3. **Data Coverage Bias:** Market data might be missing or less accurate for certain assets or regions, impacting model performance. 4. **Real-Time Availability:** Ensure that the data is available in formats that can be easily fed into a web application. 5. **Web-Specific Biases:** Consider biases in user behavior data that could influence the model's recommendations. **Improvement Strategies:** • Diversify Data Sources: Use multiple sources to minimize biases and get a comprehensive view of the market. • Regular Updates: Continuously update data to reflect the most current market conditions and reduce reliance on outdated information. • Bias Correction: Implement techniques to detect and correct biases in sentiment analysis and historical data. • Validation: Regularly validate and backtest the model with out-of-sample data to ensure robustness and accuracy.
Choice of Data Labels What labels did you decide to add to your data? And why did you decide on these labels versus any other option?	1. **Price Movement (Up/Down/Stable):** • Reason: These labels indicate the direction of asset prices, which is crucial for predicting future movements and making buy/sell decisions. 2. **Trade Signals (Buy/Sell/Hold):** • Reason: These labels help in defining actionable trading decisions based on the model's predictions. 3. **Sentiment Score (Positive/Negative/Neutral):** • Reason: Sentiment labels from news or social media reflect market sentiment, influencing asset prices and trading decisions. 4. **Economic Indicators** (High/Medium/Low Impact): • **Reason:** Labeling the impact level of economic indicators helps the model assess their influence on asset prices. **Reasons for These Labels:**

	• **Relevance:** These labels are directly relevant to making informed investment decisions and aligning with common trading strategies. • **Actionability:** They facilitate clear, actionable insights (e.g., whether to buy or sell an asset). • **Data Integration:** They integrate well with other data types (price, sentiment, and economic indicators) for comprehensive analysis.

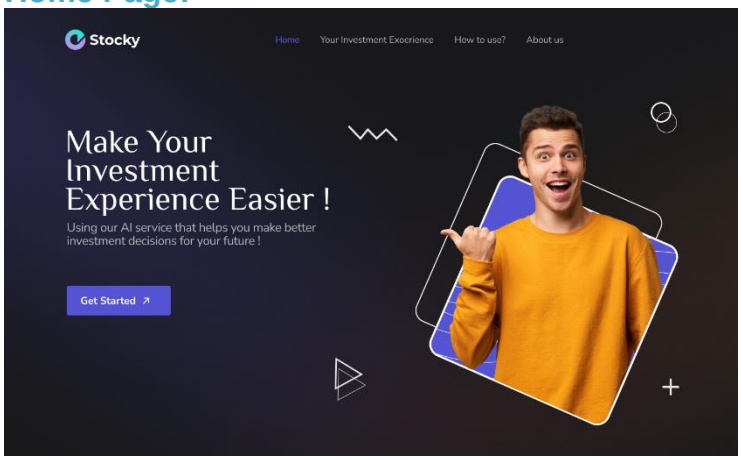
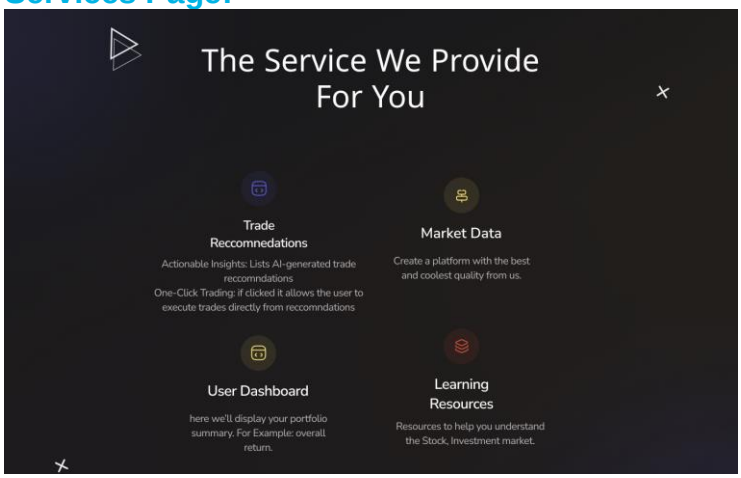
Model

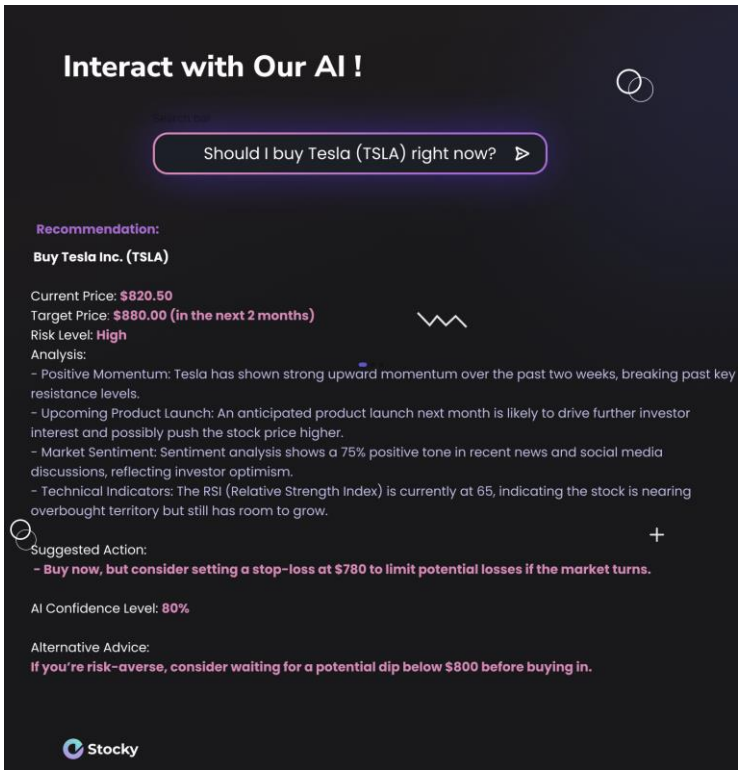
Model Building How will you resource building the model that you need? Will you outsource model training and/or hosting to an external platform, or will you build the model using an in-house team, and why?	• **In-House Development:** The model will be built using an in-house team of data scientists, ML engineers, and financial analysts. This approach ensures full control over the model's design, customization, and alignment with specific business goals. **Outsourcing vs. In-House:** • **In-House Advantage:** Developing the model in-house allows for deep integration with existing systems, better alignment with proprietary strategies, and faster iteration cycles. It also ensures data security and the ability to fine-tune the model based on specific business needs. • **Outsourcing:** While outsourcing to platforms like AWS Sagemaker or Google AI could speed up initial deployment, it may limit customization

	and control, making it less ideal for a specialized financial model. **Decision:** An in-house approach is preferred to maintain control, ensure alignment with strategic goals, and protect proprietary data. **Web Integration:** While an in-house team might still build the model, there may be additional focus on how the model integrates with web technologies (e.g., APIs, cloud hosting). **Scalability**: consider cloud-based solutions for hosting the model to handle the load from web traffic.
Evaluating Results Which model performance metrics are appropriate to measure the success of your model? What level of performance is required?	Performance Metrics: 1. **Sharpe Ratio:** • Definition: Measures the risk-adjusted return of the portfolio. • Target Level: A Sharpe Ratio above 1.0 is typically considered good; above 2.0 is excellent. 2. **Annualized Return:** • Definition: The average return generated by the model on an annual basis. • Target Level: Outperform the market benchmark (e.g., S&P 500). 3. **Drawdown:** • Definition: Measures the largest peak-to-trough decline in the portfolio. • Target Level: Minimize drawdowns to less than 10% to control risk. 4. **Hit Rate:** • Definition: The percentage of successful trades relative to the total number of trades. • Target Level: Aim for a hit rate above 60%. 5. **Prediction Accuracy:** • Definition: Accuracy of price movement predictions (up, down, or stable). • Target Level: Greater than 70% accuracy. **Required Performance:** The model must outperform traditional benchmarks with a strong Sharpe Ratio, high annualized return, and low drawdown while maintaining a high prediction accuracy

	and hit rate to be considered successful. **Web-Specific Metrics:** measure website performance (e.g., page load times, user interaction rates).

Minimum Viable Product (MVP)

Design What does your minimum viable product look like? Include sketches of your product.	**Home Page:**  **Services Page:**  **The Trade Recommendation Page:** Here is an example of how the user interacts with the AI model and how the model should response.



Use Cases

What persona are you designing for? Can you describe the major epic-level use cases your product addresses? How will users access this product?

Persona:

- Retail Investor: A tech-savvy individual, typically between 25-45 years old, interested in managing their own investments while seeking advanced tools for optimized decision-making.
- Financial Advisor: Professionals managing client portfolios, looking for AI-driven insights to support their investment strategies.

Major Epic-Level Use Cases:

- Portfolio Management: Users can monitor their investment portfolios, receive AI-driven insights, and make informed decisions to buy, sell, or hold assets.
- Market Analysis: Access real-time market data, predictive analytics, and sentiment analysis to identify investment opportunities.
- Risk Management: Receive personalized alerts and recommendations to mitigate risk based on market volatility and individual risk tolerance.
- Education and Support: Users can learn about investment strategies and complex financial concepts through interactive

	tutorials and FAQs. **Access:** • Web Platform: Accessible via any web browser, with a responsive design for both desktop and mobile devices..
Roll-out How will this be adopted? What does the go-to-market plan look like?	**Adoption Strategy:** • Phased Roll-Out: Begin with a beta version targeting a small group of early adopters and tech-savvy investors to gather feedback and refine the product. • Marketing Campaign: Leverage digital marketing channels (social media, targeted ads, influencer partnerships) to build awareness and attract users. • Partnerships: Collaborate with financial institutions and fintech platforms to integrate the product as a value-added service. • Incentives: Provide discounts, free trials, or additional features to encourage initial adoption and gather user feedback. **Go-to-Market Plan:** • Launch Phase: Focus on user acquisition through online marketing, webinars, and financial community engagement. • Growth Phase: Expand the user base with referral programs, continuous feature updates, and expanding partnerships. • Expansion Phase: Explore additional markets, such as international investors, and launch mobile apps.

Post-MVP-Deployment

Designing for Longevity How might you improve your product in the long-term? How might real-world data be different from the training data? How will your product learn from new data? How might you employ A/B testing to improve your product?	**Long-Term Improvement Strategies:** • **Feature Expansion:** Continuously add new features based on user feedback, such as advanced analytics tools, multi-currency support, or integration with new financial platforms. • **User Experience Enhancement:** Regularly update the user interface to enhance usability, making the product more intuitive and user-friendly. **Real-World vs. Training Data:** • **Variability:** Real-world data may include more noise, unexpected market events, or shifts in user behavior compared to training data. • **Adaptation:** Implement a continuous learning pipeline where the model is regularly retrained with new data to adapt to evolving market conditions and user patterns. **Learning from New Data:** • **Continuous Learning:** The product should automatically incorporate new market data, user interactions, and feedback to update the AI model and improve predictions. • **Feedback Loop:** Establish mechanisms for users to provide feedback on recommendations, which can be used to refine the model. **A/B Testing:** • **Implementation:** Regularly conduct A/B testing to compare different versions of the product (e.g., recommendation algorithms, UI layouts) to determine which performs better. • **Improvement:** Use the results to iterate on the design, focusing on enhancing user satisfaction and engagement.
Monitor Bias How do you plan to monitor or mitigate unwanted bias in your model?	**Monitoring and Mitigation Strategies:** • **Bias Detection:** Implement tools to regularly analyze model outputs for patterns that indicate bias, such as disproportionate recommendations for specific asset classes or market conditions.

	• **Diverse Training Data:** Ensure the training data covers a wide range of market conditions, asset types, and user profiles to minimize inherent biases. • **User Feedback:** Encourage users to report any perceived bias in recommendations and use this feedback to adjust the model. • **Regular Audits:** Perform periodic audits of the AI model to assess for biases and retrain the model as necessary to correct any identified issues. • **Transparency:** Provide users with insights into how the AI makes decisions, including the data and logic used, to build trust and allow users to identify potential biases themselves.