May 1, 2024

C964: Computer Science Capstone Template

Task 2 parts A, B, C and D

[Part A: Letter of Transmittal 1](#_Toc98598250)

[Letter of Transmittal Requirements 2](#_Toc1738085866)

[Letter Template 2](#_Toc1133266111)

[Part B: Project Proposal Plan 3](#_Toc1370766476)

[Project Summary 4](#_Toc904507251)

[Data Summary 4](#_Toc1736393957)

[Implementation 4](#_Toc1241988654)

[Timeline 4](#_Toc1357178365)

[Evaluation Plan 5](#_Toc623361460)

[Resources and Costs 5](#_Toc1538507987)

[Part C: Application 5](#_Toc1471073175)

[Part D: Post-implementation Report 6](#_Toc651895932)

[Solution Summary 7](#_Toc1134136520)

[Data Summary 7](#_Toc182221765)

[Machine Learning 7](#_Toc1505466430)

[Validation 7](#_Toc391434166)

[Visualizations 7](#_Toc201059345)

[User Guide 7](#_Toc1365484010)

[Reference Page 8](#_Toc1702353417)

# Part A: Letter of Transmittal

[10/25/2024]

[Mark Pendleton]

[Pendleton Trading]

[556 Rehl St]

Dear [Mark Pendleton],

I am pleased to present a proposal for a cutting-edge project to improve our stock analysis capabilities through advanced machine learning models, specifically the fine-tuned LLaMA 3.2 model. This project addresses the challenges of accurately predicting stock trends and providing actionable insights to support investment decisions in today’s fast-paced market.

**Summary of the Problem**

The stock market is complex and rapidly changing, making accurate predictions difficult. Traditional tools often lack the adaptability needed to analyze large volumes of data while capturing subtle patterns in stock performance. Investors and analysts require more reliable forecasts and insights to make timely, data-driven decisions.

**Proposed Solution**

This project proposes to develop a stock analysis application that leverages the LLaMA 3.2 model fine-tuned specifically for predictive analysis in financial markets. By incorporating historical data and stock indicators from Yahoo Finance, this application will enhance our ability to forecast trends and assess market momentum, helping to predict future price movements.

**Benefits to the Organization**

Implementing this application will yield practical benefits, including:

* **Enhanced Forecasting Accuracy**: The model will analyze large volumes of stock data, generating precise predictions for future trends.
* **Data-Driven Decision Support**: With real-time analytics and interactive query functionality, users will receive insights to improve investment strategy formulation.
* **Competitive Advantage**: Using LLaMA 3.2 provides our organization with a unique, state-of-the-art tool in a competitive market.

**Summary of Costs, Timeline, Data, and Ethical Considerations**

] This project will require approximately 35 hours to complete, with an estimated completion date of October 29, 2024. The main costs involve cloud resources for model training and any fees for accessing premium datasets. Ethical considerations include ensuring data privacy, especially when handling sensitive financial information. Robust data encryption and access controls will be implemented to meet industry standards.

**Relevant Expertise**

With a strong background in machine learning, natural language processing, and financial data analysis, I am well-equipped to execute this project successfully. My expertise in model fine-tuning, combined with hands-on experience in data-driven solutions, will ensure that the application is developed efficiently and meets the high standards of accuracy and usability.

I appreciate your consideration of this proposal and look forward to the opportunity to enhance our organization’s analytical capabilities through this innovative solution.

Sincerely,

[Trayvonious Pendleton]

[Trayvonious Pendleton, Software Engineer]

# Part B: Project Proposal Plan

## Project Summary

## • Problem Description:

## Accurately predicting stock market trends is a complex challenge due to the volume and volatility of financial data. Existing tools often need more adaptability and can process large volumes of data in real time, limiting their effectiveness for investors seeking reliable forecasts. The project addresses the need for a machine learning application that can enhance accuracy in stock prediction and provide meaningful insights for investment decisions.

## • Client Needs:

## The financial advisory firm's client requires a tool that supports accurate, data-driven decision-making. Their fundamental needs include real-time trend analysis, historical data forecasting, and insights into stock price movements. This project will deliver an advanced predictive tool that integrates LLaMA 3.2 with financial data, enabling the firm’s analysts to make informed investment decisions based on enhanced prediction accuracy.

## • Deliverables:

## The final deliverables will include:

## The Stock Prediction Application: A web-based interface with interactive queries, data visualizations, and predictive capabilities.

## User Guide: A comprehensive guide detailing application setup, usage, and troubleshooting.

## Technical Documentation: Documentation covering system architecture, data processing, model training, and maintenance procedures.

## • Project Justification:

## The application will enhance the firm’s ability to make timely, informed decisions by providing real-time stock forecasts and interactive insights. Leveraging LLaMA 3.2 for NLP processing allows for greater predictive accuracy, empowering the client to stay competitive in a rapidly evolving financial environment.

## Data Summary

## Data Source:

## Historical and real-time stock data will be collected from Yahoo Finance through its API. Additional datasets, including economic indicators or sector performance data, may be integrated to enhance prediction accuracy.

## Data Processing and Management:

## Before feature engineering, data will be preprocessed to handle outliers, missing values, and formatting issues. During development, data pipelines will clean and structure data for model training, while the application will use automation to refresh and update data for predictions. Data will also be securely stored in SQLite to maintain accuracy and accessibility over time.

## Data Relevance:

## Financial indicators and stock price histories are well-suited to the project’s goal of predicting market trends. Any anomalies, such as outliers or incomplete data points, will be addressed through statistical preprocessing, ensuring model reliability.

## Ethical and Legal Concerns:

## This project involves publicly available financial data; therefore, no specific ethical or legal concerns exist. However, data security protocols will be applied to ensure safe handling and prevent unauthorized access.

## Implementation

## Methodology:

## The project will follow the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology, which consists of business understanding, data understanding, data preparation, modeling, evaluation, and deployment. This ensures a structured approach that addresses each phase of the machine-learning pipeline.

## Implementation Plan:

## Business Understanding: Define objectives, scope, and user requirements.

## Data Understanding: Acquire and explore Yahoo Finance data, identifying trends and relevant features.

## Data Preparation: Clean and preprocess data, engineer features such as moving averages, and prepare data for LLaMA 3.2 fine-tuning.

## Modeling: Fine-tune LLaMA 3.2 and develop predictive models (e.g., regression, time series).

## Evaluation: Test and validate model predictions and refine based on performance metrics.

## Deployment: Implement a user-friendly interface and set up automated data updates for real-time analysis.

## Timeline

|  |  |  |  |
| --- | --- | --- | --- |
| Milestone or deliverable | Duration  (hours or days) | Projected start date | Anticipated end date |
| Business Understanding | 5 hours | 10/25/2024 | 10/25/2024 |
| Data Collection and Understanding | 5 hours | 10/25/2024 | 10/25/2024 |
| Data Preparation | 5 hours | 10/26/2024 | 10/26/2024 |
| Model Fine-Tuning and Development | 5 hours | 10/27/2024 | 10/27/2024 |
| Model Evaluation | 5 hours | 10/28/2024 | 10/28/2024 |
| Deployment and Documentation | 5 hours | 10/28/2024 | 10/28/2024 |
| Testing and Final Revisions | 5 hours | 10/29/2024 | 10/29/2024 |

## Evaluation Plan

## Verification Methods:

## Unit tests and data validation checks will be used at each development stage to ensure data integrity and accuracy.

## Model performance will be verified through metrics such as Mean Absolute Error (MAE) and R-squared to track accuracy throughout development.

## Validation Method:

## The application will be validated upon completion through test cases that simulate real-world usage scenarios. This ensures that predictions align with expected outcomes and meet user requirements.

## Resources and Costs

**Hardware and Software Costs**:

* **Google Colab Pro**: Approximately $10/month (used for model training and testing).
* **Python Libraries**: Free, open-source libraries (e.g., sci-kit-learn, transformers).

**Labor Costs**:

* Estimated labor time: 35 hours.
* Estimated cost at $50/hour: $1,750.

**Environment Costs**:

* Deployment on a cloud server (e.g., AWS or Heroku) for real-time access and maintenance: $20–50/month.

This plan provides a comprehensive framework to deliver a powerful stock prediction tool that addresses the client’s needs, offering technical rigor and strategic value.

# Part C: Application

**Stock app link:**

[**Stock Interface.zip**](https://westerngovernorsuniversity-my.sharepoint.com/:u:/r/personal/tpend32_wgu_edu/Documents/Stock%20Interface.zip?csf=1&web=1&e=VyArhS)

**Google Colab for model training:**

**https://colab.research.google.com/drive/1HT3PfeLOpLQYKS87PyVMCeQgnm1plgdA?usp=sharing**

# Part D: Post-implementation Report

## Solution Summary

**Problem and Solution**: The stock market’s volatility and complexity pose significant challenges to predicting trends and providing timely, data-driven insights. Traditional analysis tools cannot often process and interpret large amounts of available data effectively. This project developed a web-based application to improve stock trend predictions using advanced machine learning models, explicitly fine-tuning the LLaMA 3.2 model to deliver actionable insights for investment decisions.

**How the Application Provides a Solution**: The application integrates historical data, machine learning predictions, and interactive visualizations to offer investors an intuitive tool for stock analysis. By leveraging the LLaMA 3.2 model, fine-tuned on historical stock data, the application generates forecasts for future stock trends. Additionally, interactive data visualizations such as price trends, moving averages, and volatility distribution enable users to make informed, data-driven investment decisions.

## Data Summary

## Source of Raw Data: The primary data source was the Yahoo Finance API, which provided real-time and historical stock price information. Additional financial indicators were derived through preprocessing techniques like moving average calculations to enhance model predictions.

## Data Processing and Management: The data processing pipeline involved multiple stages:

## Design: Determined essential features such as stock closing prices, moving averages, and daily percentage changes.

## Development: Collected and cleaned raw stock data, handling missing values and outliers to ensure model accuracy.

## Maintenance: Data was stored in an SQLite database to facilitate continuous updates, ensuring historical data was accessible for new analyses. Periodic automated updates kept the application relevant to current data.

## Machine Learning

## Employed Method:

## Method: Fine-tuning the LLaMA 3.2 language model.

## Purpose: LLaMA 3.2, trained initially as a language model, was adapted to analyze and predict stock trends by recognizing patterns in historical stock data.

## Development: The model was fine-tuned using historical stock data to adapt its outputs to financial predictions. Additional feature engineering, including moving averages and daily changes, further enhanced its predictive capabilities.

## Justification: LLaMA 3.2’s capacity for pattern recognition and predictive analysis made it ideal for the financial context. By fine-tuning it on historical data, we leveraged its deep learning capabilities to provide more nuanced, data-driven predictions.

## Validation

## Validation Method:

## We employed Mean Absolute Error (MAE) and R-squared (R²) metrics to validate model performance. MAE quantified the average absolute errors in predictions, while R² measured the model's explanatory power about actual stock trends.

## Validation Results: The model’s MAE remained within acceptable limits (under $5 for most stocks tested), and R² values consistently demonstrated high explanatory power, confirming the model’s reliability for short-term stock trend predictions.

## Visualizations

## The following visualizations were integral to the application’s insights:

## Closing Price Trend: Displayed the historical closing price of the stock over a specified period, allowing users to observe long-term trends.

## Moving Average Comparison (SMA vs. EMA): Visualized 20-day SMA and 12-day EMA to highlight short-term and long-term market momentum.

## Daily Percentage Change Distribution: Showed the distribution of daily percentage price changes, offering insights into stock volatility.

## These visualizations are available within the application and are attached as static images for further reference.

## User Guide

# Steps to Execute and Use the Application:

# Set up Python Environment:

# Download and install Python 3.10+ from python.org.

# Clone the project repository to your local machine.

# 2. Install Required Libraries:

# Open a terminal or command prompt in the project directory and run:

# pip install -r requirements.txt

# 3. Prepare the Data:

# Run the application to download data automatically from Yahoo Finance, or use the provided dataset file to pre-load data into the SQLite database.

# 4. Start the Application:

# In the project directory, run:

# python app.py

# Access the application in a web browser at http://127.0.0.1:5000.

# 5. Using the Application:

# Enter Stock Ticker: Type a stock ticker (e.g., AAPL for Apple) and submit.

# View Results: The application will generate predictions for the closing price and display relevant visualizations (closing price trend, SMA vs. EMA, and daily price change distribution).

# Analysis Insight: The model also summarizes the stock based on its predicted trend.

# 6. Viewing Analysis History:

# Click the Analysis History button to view previously analyzed stocks, predictions, and insights.

# 7. Troubleshooting:

# Model Errors: Ensure your system meets the model's requirements. For any model compatibility issues, refer to the Hugging Face documentation.

# Data Fetching Errors: Verify the internet connection and check for Yahoo Finance API restrictions if data fails to load.