Stock Price Forecasting System

Thejes Raj Gangadhar[1], Yashwanth Reddy Katipallly[2]

Department of Applied Data Science, San Jose State University
1 Washington Sq, San Jose, CA 95192
[1]thejesraj.gangadhar@sjsu.edu
[2]yashwanthreddy.katipally@sjsu.edu

Abstract**—**This project outlines the development of a stock price forecasting application system where real-time stock data pipeline is used to fetch stock and run them through the machine learning models then store the results in Snowflake. Using a data pipeline shaped by Airflow, the system daily extracts stock data from Alpha Vantage, forecasts the future stock prices with ARIMA, and stores the forecasted data in Snowflake for further utilization. The present report covers such aspects as the problem statement, solution expected, functional description, and direct technical characteristics, such as tables and columns, the Python script, and the SQL queries. The full implementation is residing on GitHub, same for all Python codes and the queries. This system is useful for financial analysts or institutions that would like to make predictions of stock fluctuations for a 7-day time horizon from past information.

Keywords**—** Stock price, data pipeline, ARIMA, Alpha vantage, snowflake

I. INTRODUCTION

Stock price prediction is one of the most important requisites for trading financial assets, being helpful for buyers and sellers to make decisions in the stock market. Large datasets, as well as situations requiring real-time operations, are not always easily solved with traditional methods. To that end, in this project, we suggest a modern data pipeline and database solution for collecting, predicting, and storing stock prices. This system is meant to be versatile, dependable and user friendly to satisfy the needs of users who need stock prices on a daily basis.

A. Problem Statement

The first problem is to forecast the price of certain stock of any particular company given its previous performance. Traditional approaches or manual techniques for price modeling that usually involved only legacy systems also prove to be highly impractical and do not yield reliable short-term forecasts. Moreover, the approach deals with great volumes of real-time data; therefore, it needs a powerful database and data processing system. To solve these problems, we introduced an AQM system that would download stock data on a daily basis, use the ARIMA model to make forecasts, and store the results in Snowflake for either analysis or visualization

II. SOLUTION REQUIREMETS

A. Functional Overview

The proposed system must:

1. Collect real-time stock data for selected companies (MCD, AAPL) from the Alpha Vantage API.
2. Process this data using machine learning (ARIMA) to predict future stock prices for the next 7 days.
3. Store the historical and forecasted data in a Snowflake database.
4. Provide a scalable, automated solution that runs daily using Airflow to manage data collection and processing.

B. Functional Components

1. Data Collection: Retrieve the 90-day Moving Average Stock prices for MCD and Apple Company every day from Alpha Vantage.
2. Data Processing: The last step is the usage of the ARIMA model for the next 7 days forecast by using historical data.
3. Data Storage: From there, save both the historical raw data as well as the forecasted data in Snowflake for further work.
4. Data Pipeline: With Airflow, it is now easy to schedule and automate the daily

operations of data acquisition, processing, and archiving.

C. *Users of the System*

This system is particularly developed for financial analysts, traders, or institutions that use stock price prediction in their operation. The system will perform the data acquisition and analysis task thereby enabling users to obtain updated forecast daily without additional intervention.

III. FUNCTIONAL ANALYSIS

A. *Data Collection and API Interaction*

The system communicates with the Alpha Vantage API so as to pull the stock prices for the last 90 days for MCD and AAPL among others. Data is pulled in JSON format, and then loaded into a Pandas DataFrame for analysis.

B. *Data Processing (Forecasting with ARIMA)*

After that, the historical data is used to help the system generate and apply the ARIMA model to predict the stock price of the next seven days.

C. *Data Storage (Snowflake)*

The historical and the forecasted stock prices are written into a Snowflake table with the data processed as the input. Historical data record and forecast information contain the stock symbol, date, open, close, high and low for the historical data section and forecasted close for the forecast information.

D. *Data Pipeline (Airflow)*

Airflow is for the operational part of the pipeline, that means it's used for the execution of the daily tasks. The DAG is configured to run every 10minutes (when in testing), it parses stock data, and loads it into Snowflake.

IV. TABLE STRUCTURE, PYTHON CODE, SQL QUERIES

A. *Table Structure*

Column Name	Data Type	Constraints
symbol	VARCHAR(10)	PRIMARY KEY (date, symbol)
date	DATE	
open	FLOAT	
high	FLOAT	
low	FLOAT	
close	FLOAT	
Volume	INTEGER	

B. *Python Code Repository*

The Python code for this project, including all Airflow DAGs scripts, is available on GitHub: GitHub Repository Link

V. SCREENSHOTS

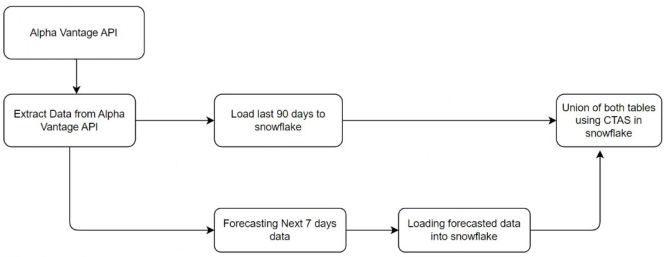

Fig 1 System design

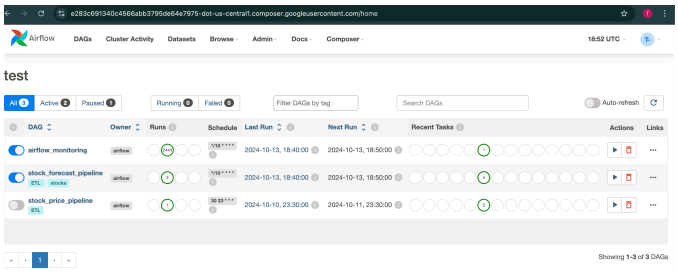

Fig 2 Airflow DAG execution

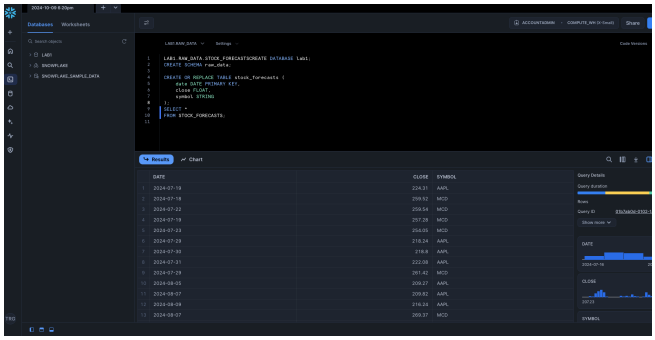

Fig 3 Snowflake query execution

VI. CONCLUSION

This project successfully automates stock price forecasting using data pipelines and a Snowflake database. The system solves the challenge of accurately predicting stock prices while handling large amounts of data in real-time. The integration of Airflow, Snowflake, and machine learning (ARIMA) models ensures a scalable, reliable, and efficient solution. This approach can be extended to other stock symbols or financial instruments, providing a valuable tool for traders and financial analysts.

VII. ACKNOWLEDGEMENT

We would like to express our deepest gratitude to our professors and teaching assistant's for their valuable guidance throughout the course of this project. Their expertise and support have been instrumental in shaping our understanding and approach. We also appreciate the resources and platform provided by Snowflake, Alpha Vantage, and other tools that made this project possible.

REFERENCES

[1] "Stock Market Prediction Using Machine Learning Techniques" by K. R. Chandar, S. Sumathi, and R. Santhi (2020). https://ieeexplore.ieee.org/document/10142862

[2] "A Machine Learning Approach to Stock Price Prediction" by Kumar, M., & Shaikh, A. (2022) https://ieeexplore.ieee.org/document/9634910

[3] "Stock Price Prediction Based on SVM, LSTM, ARIMA" by Y. Zhao, L. Wang, and J. Li (2022). https://www.researchgate.net/publication/369427623_Stock_price_prediction_based_on_SVM_LSTM_ARIMA