MLops Final Project Design

Title:

Stock prices prediction System using Neural Network Architecture

MLops Zoomcamp Final Project

Introduction

This final project aims to address key aspects learned of MLOps through the implementation of an LSTM-based stock market price prediction system. By following the modules outlined in zoomcamp syllabus, various components of MLOps were explored, so we are able to deploy and manage our LSTM model effectively.

Starting with experiment tracking and model management, MLflow was used. This is an open-source platform to track and manage our experiments, while we can log important metrics, parameters, and artifacts associated with different iterations of the LSTM model. The model registry functionality provided by MLflow also allows us to store and version trained models, ensuring reproducibility and facilitating collaboration among team members.

Moving forward orchestration and ML pipelines, Prefect 2.0 was employed, which is a robust workflow orchestration framework. This platform allows us to transform the LSTM model from a notebook prototype into a production-ready pipeline and then, we are able to automate an end-to-end process, including data preprocessing, model training, and inference. By leveraging Prefect, we can ensure fault tolerance making our LSTM model more accurate in real-world scenarios.

The LSTM model was deployed as a batch workload in the Google Cloud Platform (GCP). Leveraging GCP's cloud services such as Cloud Scheduler Cloud Storage, Cloud functions and Pub/Sub topics. As a result, a serverless architecture was designed, and it’s capable of executing periodic batch predictions on stock market data.

To ensure the performance and reliability of the deployed LSTM model a monitoring system was implemented by using Google Looker Studio (before Data Studio) and BigQuery. By utilizing tools *Evidently*, we will continuously monitor key metrics, detect anomalies, and visualize model performance after the batch process is finished.

Throughout the project, we will also adhere to industry best practices, emphasizing testing, code quality (linting and formatting). By incorporating makefiles, we will enhance the reliability, maintainability, and scalability of our MLOps solution.

Finally, with the implementation of this project, we hope to gain hands-on experience in applying MLOps principles and techniques, specifically tailored to LSTM-based stock market price prediction.

Objectives:

General objective:

- Develop and deploy a stock price prediction system using an LSTM model, which predicts stock prices with a 10-day lag based on OHLC prices and technical indicators.

Specific objectives:

- Deploy LSTM model on GCP as end-to-end batch process service.
- Integrate MLflow for experimental tracking and model registry.
- Orchestrate model using Prefect framework for processing, schedule and manage workflow.
- Deploy LSTM model in batch workload on GCP using services like Cloud Storage and Cloud Dataflow.

Motivation (Justification)

Given the significant advancements in the Machine Learning (ML) field in recent years, we are able to develop sophisticated models for various domains, including financial markets. The accurate prediction of stock market prices is of uppermost importance for investors, traders, and financial institutions. However, deploying ML models in production environments and ensuring their reliability and scalability has posed challenges. This is where MLOps (Machine Learning Operations) succeeds.

MLOps is a field that integrates the principles and techniques of DevOps with ML workflows to simplify the process of deploying, managing, and monitoring ML or Deep Learning (DL) models. It offers a structured approach to address the challenges involved in making ML models production-ready, facilitating smooth collaboration between data scientists, ML engineers, and operations teams (Alla, S. et al. 2021).

This final project aims to apply MLOps principles and techniques to a LSTM (Long Short-Term Memory) model for stock market price prediction. LSTM models have demonstrated remarkable capabilities in capturing sequential dependencies, making them suitable for time series forecasting tasks.

By undertaking this final project, we will not only gain practical experience in applying MLOps principles but also develop a robust and scalable LSTM-based stock market price prediction system. We will also deep dive into Google Cloud Platform (GCP) services. Furthermore, the knowledge and skills acquired throughout this project will allow us to contribute to the field of MLOps and address future challenges in the deployment and management of ML models.

Machine Learning Modeling approach

Our approach is mainly based on LSTM architecture. The model is initialized indicating a sequential stack of layers. In this case, a LSTM layer is added with 200 units (or neurons) ensuring that the output is returned for each timestep in the input sequence
A dropout layer with a dropout rate of 0.2 was added to prevent overfitting. Lastly, a dense layer with a single neuron is added, and the activation function *'relu'*

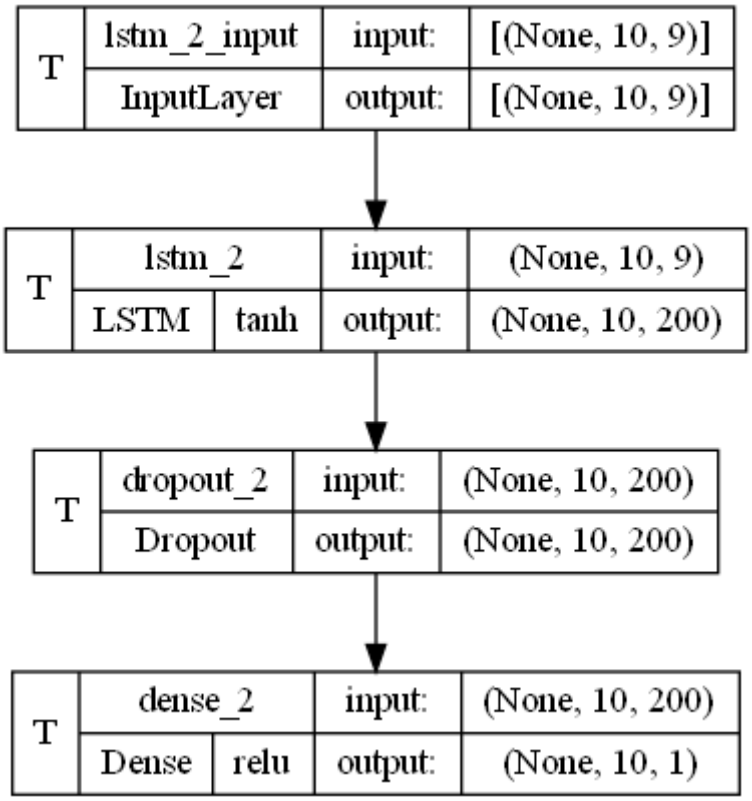

Figure 1. LSTM model used in this project.

Additionally, the model was compiled using the Adam optimizer. Loss function was set as mean squared error (MSE). In the model's performance assessment, MSE and mean absolute error (MAE) were used.

Planning Scheme

Start Date: *2023-06-10* **End Date:** *2023-07-30*

Task number	Task Name	Activities	Weeks
1	Solution design planning	➢ Theoretical research ➢ Related works ➢ Feasibility analysis ➢ Pre-requisites and technology requirements. ➢ Architecture design ➢ Development planning	1
2	Setting up Environment	➢ Cloud platform selection ➢ Preparing VM instance ➢ Installing dependencies and libraries ➢ Running VM ➢ SSH Connection with VS Code and PuttyGen (Windows 11) ➢ Setting up Storage system ➢ Running and Instancing other related cloud services (Scheduler, Functions, Topics, etc.)	1
3	ETL + Model Design	➢ Data Mining (obtaining data) ➢ Processing data (ETL) ➢ ML solution design ➢ Tuning Hyperparameters ➢ Running on premise	2
4	Experimental tracking and Orchestration	➢ Tracking experiments in the ML model with MLFlow ➢ Model Registry system ➢ Orchestrate pipeline with Prefect	2
5	Deployment	➢ Deploying solution in Cloud platform ➢ Running deployment as batch workload (neither streaming nor web service) ➢ Testing results	2

6	Monitoring	➢ Building a customizable dashboard with actual time series and predicted data. ➢ Data Drift Score in Google Data Studio	1
7	Documentation	➢ User Guide Installation ➢ Examples	1

Materials and Methods

Data preparation

Origin, cleaning, debugging, formating

Technical Requirements

Software	**Hardware** (on premise)	**Paquetes Desarrollo**	**Platforms and services**
• MLFlow • Prefect 2.0 • Docker • Docker compose • VSCode • OpenSSH • PuttyGen • Windows 11 • Evidently	• ASUS X515E Laptop • Procesador Intel ® i7-1165G7 2.80 GHz, 11th Gen • RAM 8GB • SSD 512 GB •	• Numpy • Pandas • SciPy • ScikitLearn • Keras • Matplotlib • mlpFinance • TA-LIB • Tensorflow	• Google cloud platform (GCP) • Google VM • BigQuery • Cloud Storage • Pub/Sub • Cloud functions • Cloud Scheduler

References

- Alla, S., Adari, S.K. (2021). What Is MLOps?. In: Beginning MLOps with MLFlow. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-6549-9_3
- Bhandari, H. N., Rimal, B., Pokhrel, N. R., Rimal, R., Dahal, K. R., & Khatri, R. K. (2022). Predicting stock market index using LSTM. Machine Learning with Applications, 9, 100320. https://doi.org/10.1016/j.mlwa.2022.100320
- Moghar, A., & Hamiche, M. (2020). Stock market prediction using LSTM recurrent neural network. Procedia Computer Science, 170, 1168-1173. https://doi.org/10.1016/j.procs.2020.03.049
- Ghosh, A., Bose, S., Maji, G., Debnath, N., & Sen, S. (2019, September). Stock price prediction using LSTM on the Indian share market. In Proceedings of 32nd international conference on (Vol. 63, pp. 101-110). https://doi.org/10.29007/qgcz
- Machine Learning to Predict Stock Prices. Utilizing a Keras LSTM model to forecast stock trends (2019). ARTIFICIAL INTELLIGENCE IN FINANCE. https://towardsdatascience.com/predicting-stock-prices-using-a-keras-lstm-model-4225457f0233
- Stock Market Predictions with LSTM in Python (2020). Datacamp Tutorial. https://www.datacamp.com/tutorial/lstm-python-stock-market
- Run Prefect on Google Cloud Platform (2022). https://medium.com/@mariusz_kujawski/run-prefect-on-google-cloud-platform-7cc9f801d454
- Running a serverless batch workload on GCP with Cloud Scheduler, Cloud Functions, and Compute Engine. https://medium.com/google-cloud/running-a-serverless-batch-workload-on-gcp-with-cloud-scheduler-cloud-functions-and-compute-86c2bd573f25