



Intern: Elena Pan

Supervisor: Harry Gu

Company Name: VisionX

Internship Date: Jan 2022 – Aug 2022

Internship Report

Internship Description

My role at VisionX was to evaluate and improve the stock prediction model. I worked directly with the data scientist group in introducing a new data source, deploying novel GUI features, implementing alternative algorithms, and learning neural networks and deep learning models, and eventually uploaded the version 1.3.4 to GitHub.

Overview of Internship Experience

During my internship experience at VisionX, I was able to develop skills in Python, machine learning, artificial intelligence, active learning, collaboration, and time management. Throughout my internship, I was always on time for every week's meeting, actively discussed my progress, and helped other interns with questions. In the following, I will summarize my results.

Updated VNX-22Stock Prediction Report

After successfully running the AI Stock Prediction For S&P LIVE ver 1.3.2, I updated the predicted prices of 24 stocks in the VNX-22Stock Prediction Report and calculate the corresponding MAPEs and accuracies.

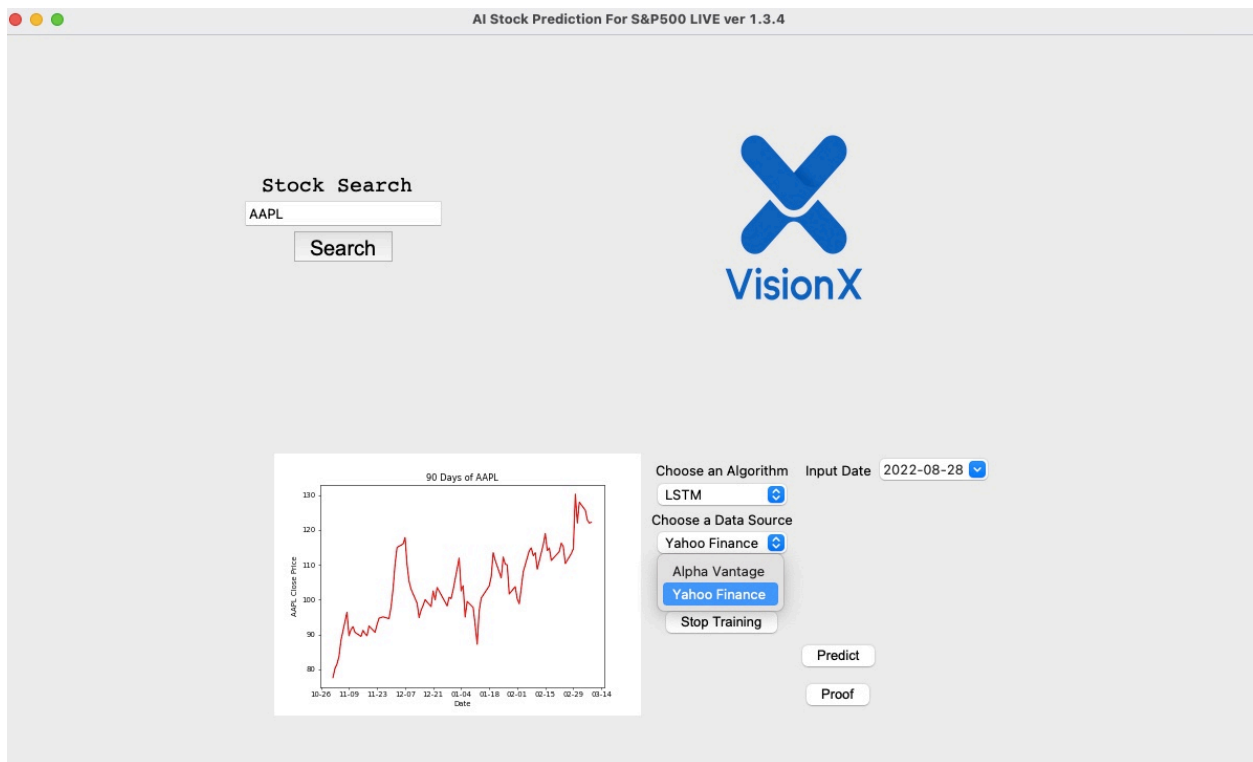
1	AAPL	real price		predict price	actual price	pred price diff	acc1
33		real price on the given day		prediction for the	actual price fo the next day		
34							
35	03/02/2022	166.56		186.23355	166.23	-20.00355	0.8796634182
36							
37	03/09/2022	162.95		158.33493	158.52	0.18507	0.9988325132



Introduced New Data Source – Yahoo Finance

After comparing and contrasting the pros, cons, and prices of over 10 APIs, I selected Yahoo Finance to replace the default Alpha Vantage as a new data source so that the stock prediction model can request more than 100 times data per request.

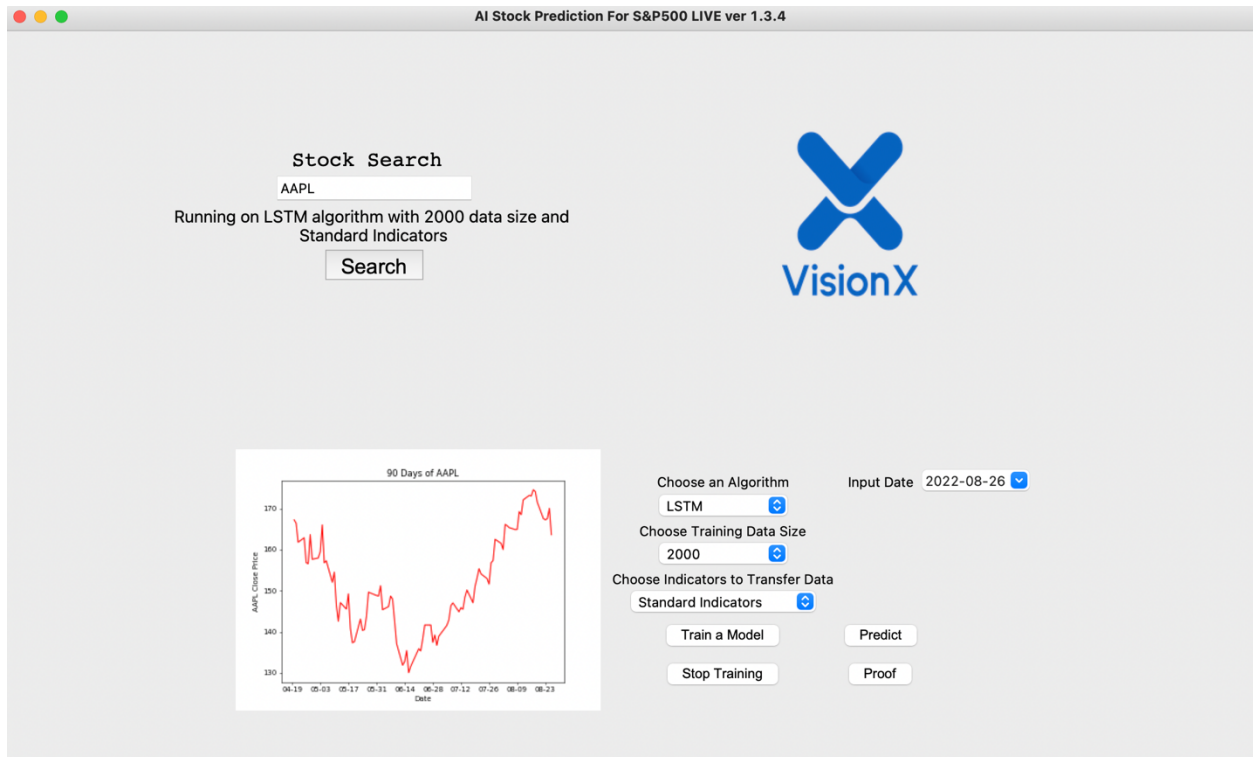
At first, I added Yahoo Finance as an alternative data source for users to choose from (refer to the picture below). After several meetings, we decided to completely request data from Yahoo Finance instead of Alpha Vantage. Yahoo Finance can provide closing prices from 1980 till now so that this new data source can help the team achieve the data sampling setting from 1k to 4k.





Deployed New GUI Features

I collaborated with my co-workers to add more features on GUI: users can select different algorithms (LSTM, BiLSTM, CNN), different training data sizes (from 1k to 4k), and SOD indicators (whether apply the SOD or not).



Implemented New Algorithm – Transformer + Time2Vec

I collaborated with a co-worker to learn and implement the Transformer combined with Time2Vec, a parallel training algorithm, as a new alternative algorithm. Although the Transformer algorithm was unable to be finalized in the 1.3.4 version because of the incompatible versions of several packages, I tested the Transformer with the latest Tensorflow and Keras packages and found a projected runtime reduction of 60% compared to the baseline LSTM model. This finding provided a potential improvement that can be applied to the stock movement model in the future.

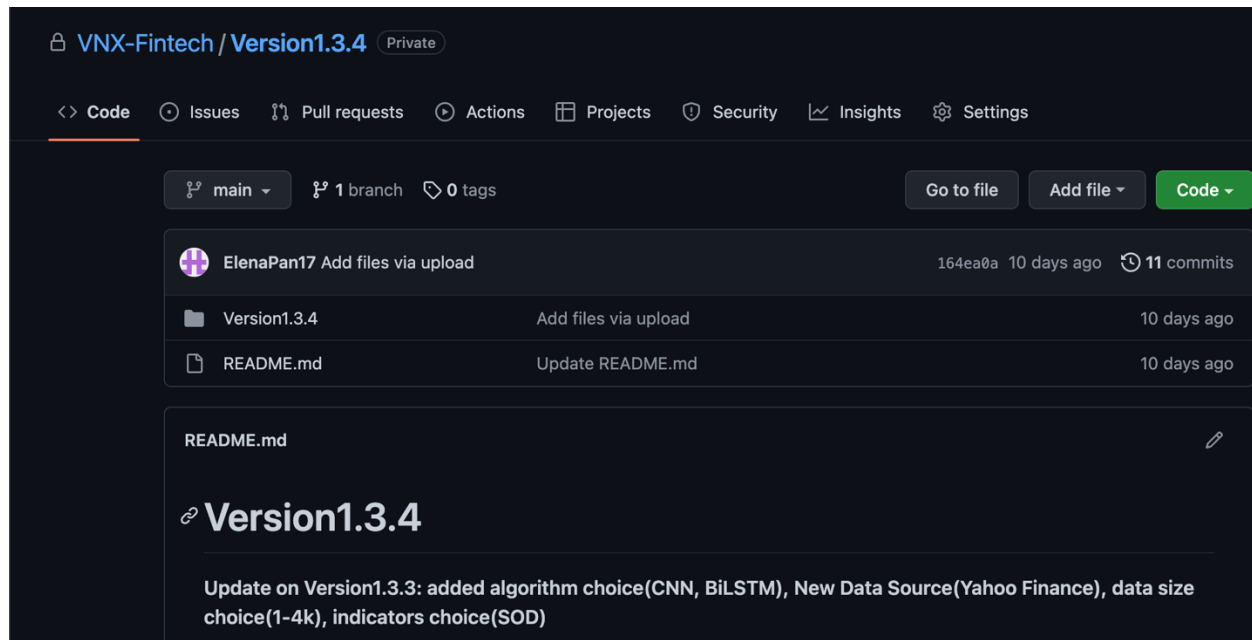


Learned Deep Learning and GAN Application in Deepfakes

Throughout my internship, the greatest achievement was my understanding of Neural Networks (LSTM, RNN) and Deep Learning (Generative Adversarial Networks). I really appreciated this opportunity to learn the advanced techniques and models nowadays.

Contributed the Version 1.3.4 on GitHub

Finally, I uploaded the latest version of 1.3.4 by adding new algorithms, a new data source, data sampling settings from 1k to 4k, and SOD indicator to GitHub.



Conclusion

My internship journey was from stressful to proud. At first, I encountered so many difficulties in setting the virtual environments and package version conflicts problems. Thanks to the enthusiastic data scientist team, a lot of co-workers gave me support and encouragement so that I successfully dealt with every problem one by one. This process also taught me not to be afraid of meeting errors and started improving the model from those errors. After I became familiar with the models and algorithms, I was also actively helping other co-workers to deal with their concerns.



First of all, I really appreciate that VisionX and Dr. Gu could offer me this opportunity to become a data scientist intern to learn advanced neural networks and deep learning models. Through this internship, I enhanced my Python skills, communication skills, active learning, collaboration, and time management. More importantly, I was intrigued by machine learning, and various algorithms and felt impressed by applying algorithms to the application. I also decided to apply for graduate school with the intended major in data science. In conclusion, my experience with VisionX was crucial in my development as a data scientist. I will take the lessons and skills I learned and apply them in the future.