Predictive Models on Tesla Stock

Masters Project

S. Sekarski
Supervisor: M. Gholam

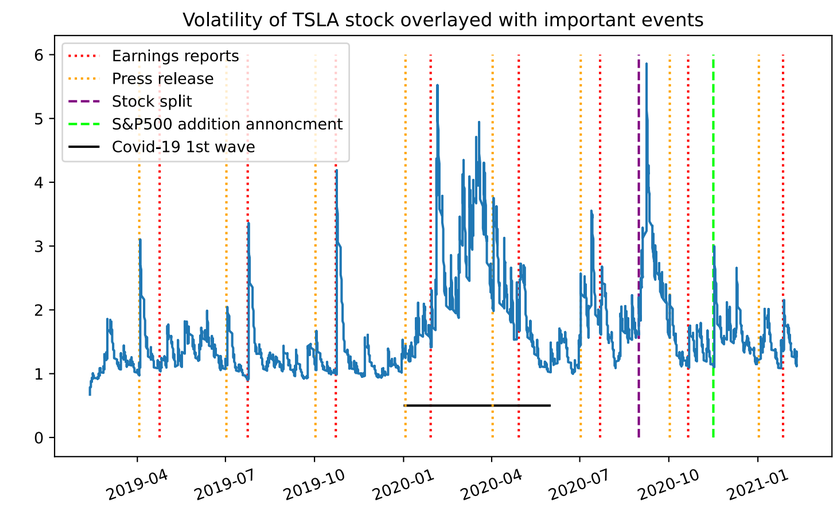
INTRODUCTION

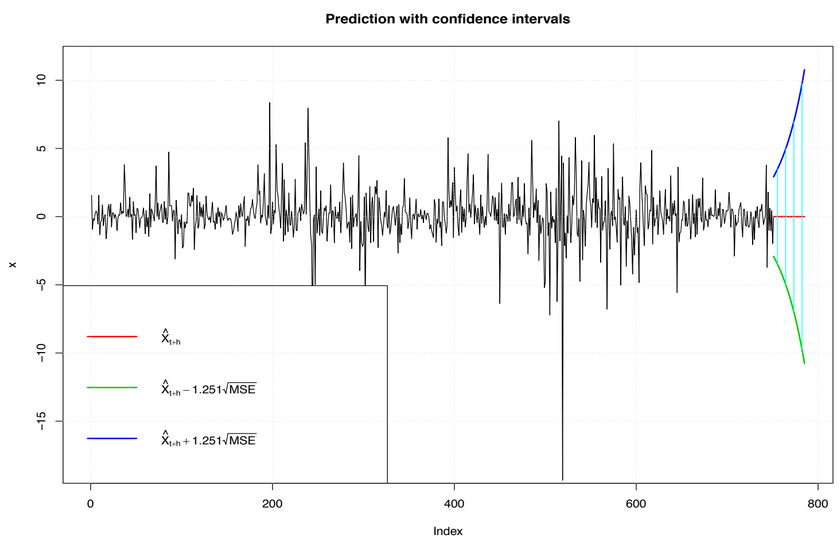
In this master project, we will explore how various Time Series and Machine Learning methods work for modeling and predicting Tesla stock price. We shall use data from Twitter and Google search trends to see if these covariates can improve accuracy. The training set spanned 2019 and 2020, while the test set was the first month of 2021.

METHOD

Before fitting the models, we preprocessed the data by considering the log returns of the stock price, fixing timestamps, aggregating data, scaling data and other corrections.

For the Time Series approach, we fit a GRACH(1,1) model (1). For the Machine Learning methods, we fit a linear model, K-Nearest Neighbors, a decision tree, a random forest (2), an extremely randomized tree (3) and a neural network.





CONCLUSION

Time Series methods give interesting insight to explain why the series behaves as it does, but is not very interesting for prediction. It has however the advantage of confidence intervals and measure of variable significance.

Machine learning methods don't give much information about the behavior of the series, but have more interesting predictions. They lack however easy access to confidence intervals and variable significance.

REFERENCES

(1) T. Bollerslev, Generalized Autoregressive Conditional Heteroskedasticity, Journal of Economics, 31 (3), 307-327, 1986L.

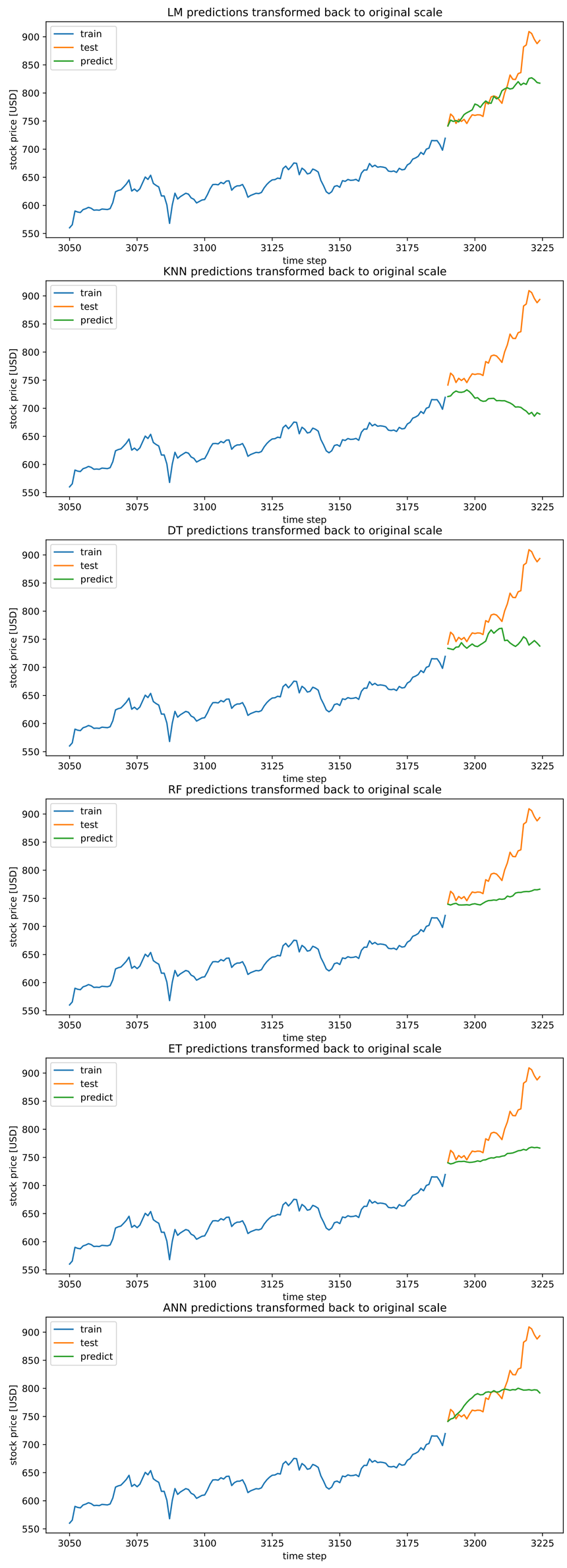
(2) L. Breiman, Random Forests, Machine Learning, 45(1),5-32, 2001

(3) P. Geurts, D. Ernst., and L.Wehenkel, Extremely randomized trees, Machine Learning, 63(1),3-42,2006

RESULTS

For the time series analysis, we see (left) the volatility extracted by the GARCH(1,1) model (with t distributed residuals) can largely be explained by key events. The prediction from the time series analysis on the returns (left) shows that the uncertainty greatly increases with every future timestep.

As for the Machine Learning methods, predictions results (below) show that linear models and neural networks seem to preform the best, while basic stratification methods such as KNN and a simple decision tree preform the worst.



SMA