Investigation of pandemics impact on stock prices.

Aitore Issadykova
Department of Computer Science
Nazarbayev University
Nur-Sultan, Kazakhstan
aitore.issadykova@nu.edu.kz

Assem Kussainova
Department of Computer Science
Nazarbayev University
Nur-Sultan, Kazakhstan
assem.kussainova@nu.edu.kz

Selim Temizer
Department of Computer Science
Nazarbayev University
Nur-Sultan, Kazakhstan
selim.temizer@nu.edu.kz

***Abstract* — The COVID-19 pandemic has pushed the world into a recession and caused economic damage is mounting across all countries, tracking the sharp rise in new infections and containment measures put in place by governments. There is an opinion that the behavior of stock prices during a pandemic can be studied and predicted using data from previous historical data on epidemics, but existing studies are not aimed at studying this theory. This paper proposes estimation of economical effects of previous pandemics which took place around the world in different periods of time using prediction models built by data mining and statistical methods and possibility of results application on current pandemic COVID-19. Research focuses on different sectors of economy that are expected to be particularly severely influenced. The results are compared using accuracy metrics like MSE and MAPE. Basic conclusion is that, the effect of COVID-19 is currently similar to the previous pandemics and although it takes a huge toll in human suffering, it would most likely not be a severe threat to the companies stock prices in long-term period.**

Keywords* — *Pandemic, Stock, Data Mining, Autoregressive Integrated Moving Average, Linear Regression, Long Short-Term Memory, Support Vector Machine, Data Analysis

I. INTRODUCTION

The COVID-19 coronavirus pandemic brought down the usual lifestyle of millions of people and has made an impact on all areas (see figure 1), but most of all people care about how the current situation will affect the world's economy and financial markets. The market was faced with the spread of Ebola, SARS, MERS, bird and swine flu and a number of other terrible diseases that had an impact on stock prices, and, in general, in such situations, the market reacted in similar ways. Currently, the spread of the COVID-19 pandemic continues, but although the relative pace has slowed, the negative impact of measures taken to limit the epidemic on the economy of all countries of the world is already a fact.

The most affected by the spread of coronavirus may be the shares of companies whose business is associated with a massive concentration of people. It is understandable that many will avoid traveling, visiting shopping centers, cinemas. Consumer spending accounts for 70% of the global economy [1], so any suspension of activities can cause damage, in particular, interfere with its growth or lead to a reduction. A number of economic sectors have already been affected by coronavirus. There are fewer forecasts regarding the state in which the global economy will crawl out of the pandemic. Good news is not expected. A quarantine month is a loss of about a quarter of quarterly GDP for developed countries. Obviously, states consciously sacrifice economic growth to save lives. However, quarantine measures have a limit when national economies begin to crumble, and the victims of devastation and famine will be disproportionately more than those who died from the virus. In search of this balance today are authorities around the world.

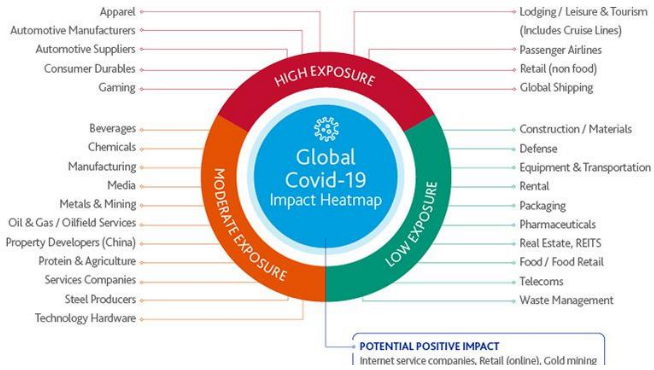

Figure 1. Spheres exposed to effects of coronavirus COVID-19

It is not known how long and strongly the current situation will continue. Regarding this situation, stock market analysts are trying to analyze at least the approximate effect of the coronavirus epidemic on the income of global corporations. However, since this is the first appearance of coronavirus COVID-19, a very small number of studies in this direction have been carried out in a short time. In addition, there are no official published studies to predict market behavior using stock data during past pandemics.

For this reason, this project aims to study the impact of previous epidemics on stock prices and the possibility of predicting their future behavior during COVID-19 using historical data and various data mining algorithms. In addition, the work is aimed at studying and comparing the effectiveness of the methods used with the help of statistical metrics. The Section II discusses the history of the impact of viruses on markets over the past 20 years and their main trends. In Section III, we describe used data, analyzed the proposed algorithms and built models. Section IV illustrates the results of the entire work and following part contains discussion and list of all the observations of our project. In section VI, we provide the conclusion of our final project and some recommendation for future research and improvements.

II. Market History Of Viruses Impact

Judging by previous outbreaks of SARS in 2002-2003, swine flu virus in 2009 and Ebola in 2014, markets may fall in the short term due to increased volatility, but they can also bounce back relatively quickly.

According to the World Health Organization, from November 2002 to July 2003, as a result of SARS, 774 people died worldwide, most cases occurred in mainland China and Hong Kong [2]. During this time, economic growth in the region slowed down; Hong Kong fell into recession. Stock indices experienced double-digit percentage falls, but then recovered. According to Timothy Moe of Goldman Sachs, markets tended to adjust to the number of infections reported daily. The worst drop occurred when the number of patients began to increase sharply. All three markets have fully recovered by the time the World Health Organization announced the end of the outbreak.

Other outbreaks of avian and swine flu, Ebola, respiratory syndrome in the Middle East and Zika virus have had less impact on markets than SARS [4]. The devastating pandemic of the Spanish flu, which claimed millions of lives, was accompanied even by the growth of quotations. The market grew due to inflation and recovery after the First World War. The 1968–1969 Hong Kong flu was met by a drop in indices, which soon gave way to more significant growth [3]. Since the 1980s, the HIV virus has appeared in the world, the global spread of which continues to this day. Nevertheless, we can say that the spread of this virus does not affect global markets at all.

In general, the reaction of financial markets is logical: assets at risk experience increased pressure with increasing demand for protective instruments. There are two reasons for the negative reaction of the markets. Firstly, the overbought factor - before this, stocks and oil rose significantly. Secondly, there was a risk of worsening economic performance in the Asia-Pacific region. Among the main trends, it can be observed that during pandemics, stocks of insurance companies, air carriers and tour operators are declining. Meanwhile, during periods of epidemic, medical companies are always getting more expensive.

III. Methods

A. Pandemics Data

As the data in this study, we used the stock prices of seven companies representing different areas to observe the full picture of the impact of pandemics on all sectors of the market, instead of one specific direction to avoid prejudice and overfitting of results. The list of companies used to obtain the data includes Disney, Amazon, Southwest Airlines Company, IBM, Marriott, Samsung, Twitter. In addition, all data was collected during a certain period including the last 3 major epidemics such as SARS, MERS and Ebola, as they may have a closer relationship with the current coronavirus COVID-19.

The data was received by downloading stock data from the Yahoo Finance website. The Yahoo Finance is one of the most popular sites for monitoring changes in the value of shares of global companies, which also stores all the historical data related to them. The data was processed using the Python programming language and its libraries in the Jupyter Notebook environment.

B. Algorithms

- AutoRegressive Integrated Moving Average (ARIMA)

AutoRegressive Integrated Moving Average (ARIMA) is the most popular forecasting model. This is an important class of parametric models to describe non-stationary series. The purpose of applying this method is to identify the ARIMA model with the minimum necessary order of parameters that adequately reflects the behavior of stock prices, based on which reliable long-term or short-term forecasts can be made [5].

The model ARIMA (p, q, d) means that the differences of the temporal order obey the ARMA (p, q) model.

$$\Delta^d X_t = c + \varepsilon_t + \sum_{i=1}^{p} \alpha_i \Delta^d X_{t-i} + \sum_{i=1}^{q} \beta_i \varepsilon_{t-i} \qquad (1)$$

where Δ^d - time difference operator of order d, and c, α_i, β_i – parameters of the model. This reseacrh uses non-seasonal ARIMA model of random walk type with parameters p, d, q equal to 0, 1 and 0 respectively.

- Linear Regression

Linear regression is a regression analysis method, which represents a statistical study of the influence of one or more independent variables $X_1, X_2, \ldots, X_n$ on the dependent variable Y [6]. Independent variables are otherwise called regressors or predictors, and dependent variables are called criteria variables.

The model of linear regression represents the linear equation:

$$y = ax + b \qquad (2)$$

where a is the angular coefficient or gradient of the estimated line, which represents the amount by which Y increases on average if we increase X by one unit, and b is a free member of the estimation line; this is the value of Y when X = 0.

In this paper, we consider an example of linear regression with one predictor and several predicted values, which tries to find the coefficients a and b minimizing the magnitude of the error.

- Support Vector Machine

Support Vector Machine (SVM) is a supervised learning algorithm, they infer a function or relationship from the given training data, this algorithm learns by analyzing data and recognizing patterns, and are frequently applied in the field of pattern regression analysis and classification [7]. When provided with a set of training set, belonging to two different classes, an SVM algorithm design a model which can then efficiently assigns a new example point to one of the two classes. This algorithm was used in current paper for regression goals.

The model is built using such parameters as cost, gamma and kernel. The implemented model was of linear SVM type, and the cost function is calculated using following formula:

$$Cost(h_\theta(x), y) = \begin{cases} \max(0, 1 - \theta^T x) & if\ y = 1 \\ \max(0, 1 + \theta^T x) & if\ y = 0 \end{cases} \quad (3)$$

where $h_\theta(x)$ is SVM hypothesis, and $\theta^T x$ is raw model output.

For this model were chosen the values of parameters as kernel = linear, gamma = 0.1 and cost = 1e3.

- Long Short-Term Memory

Long Short-Term Memory (LSTM) is a type of recurrent neural network that can learn long-term dependencies [8]. Recursive neural networks based on this approach have a more advanced and more complex way of calculating the values of hidden layers. This method, in addition to the input values and the previous state of the network, also uses filters (gates) that determine how the information will be used to calculate both the output values on the current layer and the values of the hidden layer in the next step.

Our LSTM model has four hidden layers: two LSTM layers and two dense layers (see figure 2). First hidden layer (LSTM) has 50 neurons, and the same is the second LSTM layer. Third hidden layer is a dense layer with value of 25, and final dense layer has value of 1.

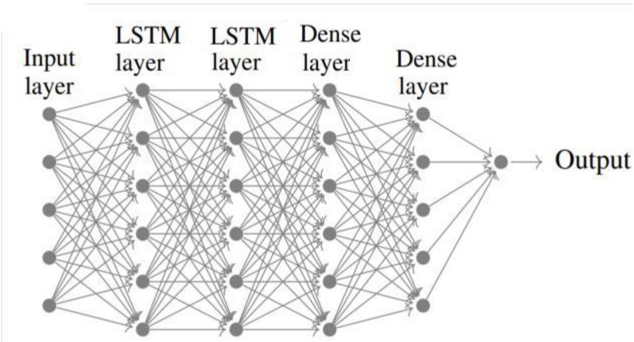

Figure 2. Implemented LSTM model

The proposed forecasting LSTM Model are trained with 10 epochs with batch size equals to 1. Moreover, the ”adam” optimizer parameter and MSE loss functions was applied.

C. Experimental Setup

The forecasting price models were constructed by using pairs of pandemics and their data as implemented methods have a single input for one time series and due to the fact that results will be more precise if experiments conducted separately without the mixture of data. Therefore, three train/test pairs created on pandemics data:

- train – SARS data, test – MERS data
- train – SARS data, test – Ebola data
- train – MERS data, test – Ebola data

For validation of the performance of each model and two compare the models between each other, two metrics were used: MSE and MAPE. Mean Squared Error or MSE is an estimator which measures the average of sum of squares of the errors, i.e. the error between testing data and predicted values using following formula:

$$MSE = \frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y_i})^2 \quad (4)$$

Mean Absolute Percent Error (MAPE) assessment is used for time series, the actual values of which are significantly greater than 1.

$$MAPE = \frac{100\%}{N} \sum_{i=1}^{N} \left|\frac{y_i - \hat{y_i}}{y_i}\right| \quad (5)$$

where y_i is an actual value of stock price for testing dataset and $\hat{y_i}$ is a predicted value of stock price by forecasting model.

All models were trained on CPU with 8 GB RAM and Windows system with Python Anaconda packages.

IV. Results

All data from companies was visualized on graph and main trends were identified. As previous pandemics state, all companies tend to decrease their stock prices when the peak of epidemia occurs and after some time the prices tend to increase gradually till the reinstatement. On the figures 3-5, can be seen example of the trends of Disney company stock prices variation during SARS, MERS and Ebola periods.

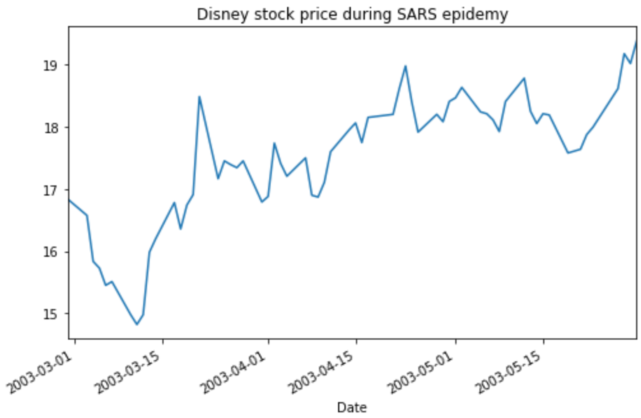

Figure 3. Disney company stock price variation during SARS period

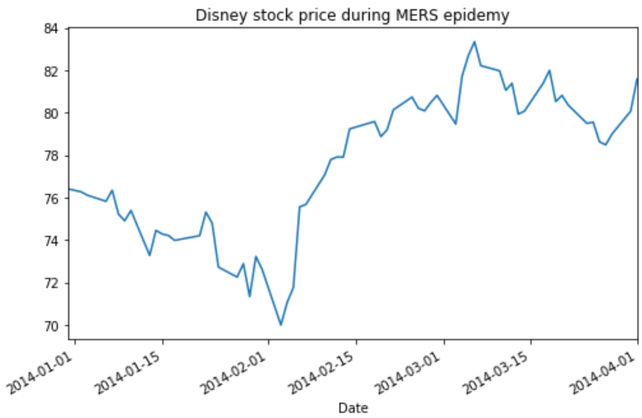

Figure 4. Disney company stock price variation during MERS period

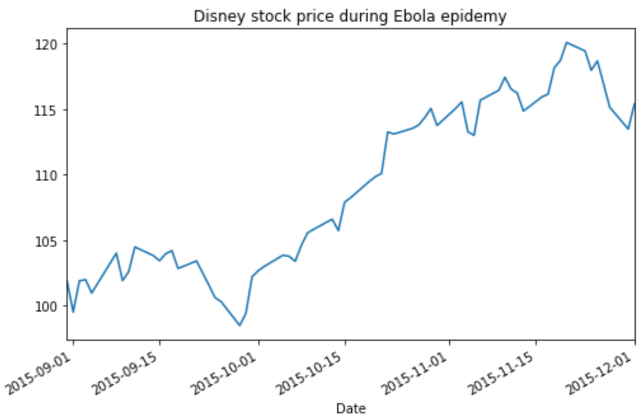

Figure 5. Disney company stock price variation during Ebola period

Implemented models were trained and tested on listed pandemics data pairs for all companies, except Twitter since data is not present for that period. Following are represented results for each company, models used and the accuracy metrics values in the form of tables.

TABLE I.
AVERAGE MODELS RESULTS ON DISNEY COMPANY STOCKS

Metrics	ARIMA	LSTM	Linear Regression	SVM
MSE	3.055	107.510	1.940	1.771
MAPE	6.963	1.014	1.413	1.354

TABLE II.
AVERAGE MODELS RESULTS ON AMAZON COMPANY STOCKS

Metrics	ARIMA	LSTM	Linear Regression	SVM
MSE	24.518	567.619	22.487	20.028
MAPE	12.345	2.011	3.352	2.877

TABLE III.
AVERAGE MODELS RESULTS ON MARRIOTT COMPANY STOCKS

Metrics	ARIMA	LSTM	Linear Regression	SVM
MSE	2.425	74.162	1.263	2.146
MAPE	5.034	1.605	5.030	2.579

TABLE IV.
AVERAGE MODELS RESULTS ON SAMSUNG COMPANY STOCKS

Metrics	ARIMA	LSTM	Linear Regression	SVM
MSE	880.022	24375.698	520.309	827.670
MAPE	8.040	1.735	1.653	2.685

TABLE V.
AVERAGE MODELS RESULTS ON SOUTHWEST AIRLINES COMPANY STOCKS

Metrics	ARIMA	LSTM	Linear Regression	SVM
MSE	1.860	40.650	1.332	1.301
MAPE	10.611	1.369	2.759	2.680

TABLE VI.
AVERAGE MODELS RESULTS ON IBM COMPANY STOCKS

Metrics	ARIMA	LSTM	Linear Regression	SVM
MSE	6.965	140.171	3.292	3.259
MAPE	4.916	1.429	1.853	1.990

For almost all of the six companies LSTM model forecasts stock prices more accurate than other methods for one week forward and therefore it was chosen as the best model from all implemented. It was expected that statistical method ARIMA will have better prediction score, however its performance was the lowest.

V. DISCUSSION

Almost all of four proposed methods: Linear Regression, SVM, LSTM and ARIMA performance achieved our expectations. Some methods perform better than others for different stocks of companies. As the prediction period was comparatively chosen looking to available capacity of data, there was no issue for machine learning methods performance of chosen models. The following table shows the average obtained results of MAPE for all four methods.

TABLE VII.
AVERAGE MODELS RESULTS ON ALL COMPANY STOCKS

Metrics	ARIMA	LSTM	Linear Regression	SVM
MAPE	7.984	1.527	2.676	2.360

As it can be seen the LSTM is one of the best models to predict changes in behavior of stock prices. Moreover, it was checked that all stocks have similar behavior during different pandemics and that it repeats from pandemic to pandemic. For more accurate results it is suggested to train these models on higher amount of pandemic cases to get more information about stock changes.

Since COVID-19 is the most recent pandemic which happening now in almost all countries, the LSTM model can be used for observing the behavior of stock prices to predict the future price fluctuations, which can help to prepare companies for sudden changes and take appropriate measures for reducing the risks. Moreover, for the future works effect of other factors in combination wth close price of stocks during pandemics can be studied in order to increase prediction accuracy of LSTM model.

VI. CONCLUSION

Any pandemics resulted in the major loss in the world economy and the most of business sectors suffers during quarantine times. The current COVID19 pandemic caused the significant drops in the financial stock market. This study conducted to predict the long-term behavior of stock for companies from different sectors during pandemic times such that SARS, MERS and Ebola. The pattern of the stock's behavior is similar from pandemic to another pandemic. The proposed models can be applied for the COVID19 case in order to detect the recover speed of the stock market and predict the future stock prices for second wave of COVID19 pandemic which is predicted to be in the fall and winter period in the year 2020.

Among four proposed models, the machine learning methods outperform the statistical ARIIMA model. The LSTM model provide the outstanding results with 98.57% accuracy in average. This result indicated that LSTM model can almost perfectly imitate the long-term price changing patterns of companies from different sector.

In summary, the LSTM model can be used to predict long-term pattern of the stock prices during pandemic times. Due to the fact that the pattern of stock prices changes is similar during pandemics the current findings can be useful for companies to forecast the long-term situation in the financial market. The limitation of this paper is ambiguity of the starting and ending time of the pandemics. Further researches can be conducted by comparing the performance of hybrid models such as CNN-LSTM or RF-SVM with LSTM Model results. Also, the performance of proposed models can be verified through second wave of the COVID19 pandemic.

REFERENCES

[1] U.S. Bureau of Economic Analysis, Shares of gross domestic product: Personal consumption expenditures [DPCERE1Q156NBEA], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/DPCERE1Q156NBEA, May 5, 2020.

[2] World Health Organization, Summary of probable SARS cases with onset of illness from 1 November 2002 to 31 July 2003, retrieved from: https://www.who.int/csr/sars/country/table2004_04_21/en/

[3] Henderson D., Courtney B., Inglesby T., Toner E., Nuzzo J. Public health and medical responses to the 1957–58 influenza pandemic. Biosecur. Bioterror. 2009; 7:265–273. doi: 10.1089/bsp.2009.0729.

[4] Zheng J. (2020). SARS-CoV-2: an Emerging Coronavirus that Causes a Global Threat. International journal of biological sciences, 16(10), 1678–1685. https://doi.org/10.7150/ijbs.45053

[5] Din, Marilena. (2015). ARIMA by Box Jenkins Methodology for Estimation and Forecasting Models in Higher Education. 10.13140/RG.2.1.1259.6888.

[6] Kumari, Khushbu & Yadav, Suniti. (2018). Linear regression analysis study. Journal of the Practice of Cardiovascular Sciences. 4. 33. 10.4103/jpcs.jpcs_8_18.

[7] Evgeniou, Theodoros & Pontil, Massimiliano. (2001). Support Vector Machines: Theory and Applications. 2049. 249-257. 10.1007/3-540-44673-7_12.

[8] Sherstinsky, A. (2020). Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. Physica D: Nonlinear Phenomena, 404, 132306.