Objective and Purposes

The main objective and reason of this research is to analyze how to predict the price of the stock using machine learning. First, there will be found a stock, which can be presented as a non-stationary time series: $y = f(n) + z$, where *f(n)* is a trend, that could be linear, exponential, power, polynomial or logarithmic, and z is a residue. In this final project the following types of trends were considered:

- $f(n) = an + b$ - linear
- $f(n) = ae^{bn}$ - exponential
- $f(n) = a + bln(n)$ - logarithmic
- $f(n) = a + n^b$ - power
- $f(n) = an^k + bn^{k-1} + \cdots$ - polynomial, with *k=2*

The trends were highlighted using the least square method: $\sum_{i=1}^{N}(y(n) - f(n))^2 \rightarrow min$. Then for further finding the optimal coefficients of the trend's equation, there were constructed and solved the corresponding systems. For example, following for the linear trend:

$$\begin{cases} a\sum_{i=1}^{N} n_i^2 + b\sum_{i=1}^{N} n_i = \sum_{i=1}^{N} n_i y_i \\ a\sum_{i=1}^{N} n_i + bN = \sum_{i=1}^{N} y_i \end{cases}$$

The appropriate coefficients a and b can be found as: $a = \frac{n\sum_{n=1}^{N} ny(n) - \sum_{n=1}^{N} n\sum_{n=1}^{N} y(n)}{n\sum_{n=1}^{N} n^2 - \sum_{n=1}^{N} n\sum_{n=1}^{N} n}$, $b = \frac{\sum_{n=1}^{N} n^2 \sum_{n=1}^{N} y(n) - \sum_{n=1}^{N} ny(n)\sum_{n=1}^{N} n}{n\sum_{n=1}^{N} n^2 - \sum_{n=1}^{N} n\sum_{n=1}^{N} n}$.

Substituting the calculated coefficients into the trend function, the linear trend equation can be obtained. Further, using the formula $y = f(n) + z$, a residue was found that was analyzed for stationarity and whose model should be identified. To verify stationarity, the autocorrelation matrix method was used. In the case of stationarity of the studied data, it should be positively determined, which was obtained, that is, the remainder can be considered stationary. And the data model was identified by the behavior of autocorrelation and private autocorrelation functions. For fitting the data were used following models:

- $X_n = \varphi_1 X_{n-1} + \ldots + \varphi_p X_{n-p} + \varepsilon_n$ - AR(p), p=1,2
- $X_n = \varepsilon_n - \theta_1 \varepsilon_{n-1} - \ldots - \theta_q \varepsilon_{n-q}$ - MA(q), q=1,2
- $X_n = \varphi_1 X_{n-1} + \ldots + \varphi_p X_{n-p} + \varepsilon_n - \theta_1 \varepsilon_{n-1} - \ldots - \theta_q \varepsilon_{n-q}$ - ARMA(p,q). p=1, q=1

where $\varphi_1, \varphi_2, \Theta_1, \Theta_2$ - are the parameters, and ε_n is the white noise dispersion. They were evaluated by the method of moments. Now fitting the residuals according to the model can be obtained - $\widetilde{y_n}$. For this step of the research according to the formula:

$$X_n = f(n) + \widetilde{y_n}, \text{ where}$$

f(n) is the trend equation, a fitting for the source data was obtained.

Moreover in this research was considered an alternative method foe removing the linear trend. $\xi_n = y_n - y_{n-1}$, where ξ_n is a alternatively obtained trend and y_n , y_{n-1} are the initial data. The final fitting was got by the inverse sum method: $\widetilde{X_n} = \widetilde{y_n} + Y_n$, where Y_n is initial sample.

One of the methods for comparing the received, fitted data with the source data is the standard error: $S_e = \frac{1}{N}\sum(X_n - \widetilde{X_n})^2$. By its value it is possible to conclude that the analyzed data are sufficiently close to the initial.

Then using the supervised machine learning possible to forecast if an investor has gains or losses. The final expected result would be useful for the investors in term to understand what shares are worth paying attention to and foresee how much the price on an investment will change.