

# Predicting the Saudi Stock Market Index - TASI

## Abstract

The project aims to predict the future values of the stock market index for the Saudi Stock Exchange (TASI) using machine learning. Experiments are performed in 1–10, 15 and 30 days in advance. The project proposed approach involving Support Vector Regression (SVR). The project is based on a paper by Patel et.al. (2015) where he used machine learning techniques to predict two stock market indices CNX Nifty and S&P Bombay stock exchange. The project aims to predict the future values of the stock market index for the Saudi Stock Exchange (TASI) using machine learning. Experiments are performed in 1–10, 15 and 30 days in advance. The project proposed approach involving Support Vector Regression (SVR). The project is based on a paper by Patel et.al. (2015) where he used machine learning techniques to predict two stock market indices CNX Nifty and S&P Bombay stock exchange.

## Data Description:

The dataset retrieved from Bloomberg. The dataset contains 3113 row with 40 features for. Initially, the data will be daily from 2010 until 2021 of TASI index since the Saudi stock market has changed in structure in the past 10 years. Predictor variables are taken from Patel et.al (2015) which are the following:

Indicator	Description
Weighted 10 days moving average (MA(10C)s)	This indicator calculate the closing prices of TASI for t last 10 days and divide the result by the number of days(10)
Weighted 10 days moving average (MA(10C)w)	Is a technical indicator that traders use to generate trade direction and make a buy or sell decision. It assigns greater weighting to recent data points and less weighting on past data points. The WMA is calculated by multiplying each observation in the data set by a predetermined weighting factor.
Momentum Stochastic K% (%K(141))	Based on the lowest low and highest high of a given period, the %K line displays the last closing price as a percentage of this range.
Stochastic D% (%D(3))	Based on the lowest low and highest high of a given period, the %K line displays the last closing price as a percentage of this range.
Relative strength index (RSI) (RSI(14C))	RSI is a momentum indicator commonly used in technical analysis to determine whether a stock or other asset is overbought or oversold
Moving average convergence divergence (MACD) (MACD(12269C))	MACD is an indicator of trend-following momentum that shows the relationship between two price moving averages. The MACD is calculated by subtracting the 26-period exponential moving average (EMA) from the 12-period EMA.
A/D (Accumulation/Distribution) oscillator (AccDist)	Using volume and price, the accumulation/distribution indicator (A/D) calculates whether a stock is being accumulated or distributed.

Indicator	Description
<b>CCI (commodity channel index) (CCI(20C))</b>	This indicator measures whether current prices are above or below the average over a certain period of time.
<b>Average True Range (ATR) (ATR(14))</b>	The indicator measures market volatility by decomposing the entire range of an asset price over a certain period of time.

### Algorithms:

- Cleaning the dataset: I used dropna() to drop any row with null value.
- Correlations: Try to find the best correlations. First I chose the variables (MA(10C)s ,MA(10C)w ,%K(141) ,%D(3),RSI(14C) ,MACD(12269C) ,AccDist ,CCI(20C) ,ATR(14), however, I found Ma(10C)s and MA(10C)W doing the same effect no new information. Therefore, I change my selection to (MA(10C)s ,%K(141) ,%D(3),RSI(14C) ,MACD(12269C) ,AccDist ,CCI(20C) ,ATR(14) which show no correlation.
- Models:
  - I first applied the SVM model, however to get better result should be done in two stages, by applying also ANN and RF in the second stage.
  - By looking at the problem and we couldn't use SVM I used the Linear Regression which return better result R Square= 99.59.

### Tools:

- Support Vector Regression (SVR)
- Linear Regression
- Numpy
- Pandas
- Matplotlib
- Seaborn
- Datetime
- sklearn.linear\_model

### Recommendation:

To get better result by using SVM algorithm should be done in two stages, by applying ANN and RF in the second stage