

AI in the Stock Marketing Prediction Using Data Science Implementation

Mahammad Ehtaramullah
University Institute of
Computing
Chandigarh University
Kharar, Punjab, India
21MCA2538@cuchd.in

Abstract— The stock includes the location where individuals often used acquire their stocks. People can buy or sell their shares only under the publically listed companies. Any purchaser or retailer's primary goal would be to forecast a share market's price rise in order to maximise profit while minimizing risk. To make the prediction much more effective and accurate we can use the concept of data science to implement AI in stock marketing prediction. Effectively projecting investment performance is highly problematic due to the unpredictability and perhaps quasi understanding of the banking equity markets. Since neural network and better process technology have indeed been developed, programmed techniques of forecasting were and still are increasingly effective in forecasting stock prices.

To implement the above concept we will use python programming language. While we begin, we must collect information from the prior year and in order to accurately estimate stock values. We'll utilize Finance.yahoo.com to get investment data from the past yr. Following that, we'll utilize streamlit to construct one web application, and then we'll use Fb Prophets to better forecasting stock values. The research would show how advanced data analytics may be used to forecast market prices.

Keywords: Artificial Intelligence, Data Science, Stock Market.

I. INTRODUCTION

Investors have strived to predict the stock market since its inception. Every day, the exchange trades billions of dollars, and behind each dollar is an investor wanting to profit in some way. It holds seductive promises of wealth and power for an investor who can perfectly predict market changes. It's no surprise, then, that the Stock Market and its accompanying difficulties enter the public consciousness whenever it misbehaves. The share price comprises a set of people that buy as well as sell shares. Usually referred to as stock, this term

refers to an individual's or a group's claim to ownership of a corporation. The stock market estimate is a method for determining the future value of stock market prices. Strong, accurate, and effective performance is expected. The system should be adequately related to real-life scenarios and work in accordance with them [5]. These include political events such as ministerial or government officials' statements, statements from government bodies such as the RBI and SEBI, scandal news, and so on. It can also be the global happening such as rapid movements in currencies prices and commodities prices [6]. All of this has an impact on company earnings, which in turn has an impact on stock market investors' sentiment. To measure these effectively and consistently is out of reach for practically everyone. This strategy usually necessitates the gathering of diverse social media data, news that affects stock market investor emotion, and individual feelings. Other information, such as stock prices from the previous year, is also taken into account. The connection between various data points. Is taken into account, and an estimate is made based on the many data points. Some works have been proposed which use Random Forest (RF) for forecasting purposes. RF is an ensemble technique. It is normally capable of performing both regression and classification tasks. It operates by constructing multiple decision trees at training time which outputs mean regression of individual decision trees [19]. Stock prices are now predicted using advanced intelligent approaches based on either technical or fundamental analysis. The data size is particularly large and non-linear for stock market analysis. To cope with such a wide range of data, an efficient model is required that can uncover hidden patterns and intricate relationships in this massive dataset. Machine learning techniques in this area have proved to improve efficiencies by 60-86 percent as compared to the past methods [20].

A. Aim

To keep share market investing simple, time apparently, and tiresome, and also to keep share

market trading simple, time apparently, and tedious. This method will help novice investors grasp the stock market quickly and effortlessly. Investors must anticipate the stock price trend and future range to execute a profitable trade in the share market, which needs regular market upgrades and going to keep an eye on current market information. Due to the daily difficulties in the stock market, this entire method makes it difficult for shareholders or those with financial restraints to maintain pace with such data and information.

B. Objectives

- Improve overall accuracy of pricing predictions.
- Making share price prediction relatively easy as possible.
- Give useful knowledge to prospective traders to assist people study business market better quickly.
- By merging many information studies all into a, you may minimize the duration that requires to create a forecast.
- To keep trading within share market relatively straightforward as feasible.

C. Scope

Share pricing movement, fluctuation, risks, payback, movement, comparability of share with several other competitor shares, as well as comparing to world market would be predicted using market opens, peak, lower, closing, dataset, cost of many other products linked to share, and forex cost. To determine the general sentiment of stock market investors, sentiment analysis of news is undertaken. Combining all of these data and the results of several algorithms to improve forecast accuracy.

II. LITERATURE SURVEY

A. Related Work

Authors Deepak, Uday, and Malathi wrote the paper "Machine Learning Approach in Stock Market Prediction." According to the report, the essential factors to consider when projecting share prices are a high level of accuracy and precision. To create forecasts inside the sharemarket, several shareholders and organisations use panel data research, financial evaluation of companies, and advanced analytics. Nevertheless, since such techniques aren't completely reliable, further sharemarket forecasting techniques are

utilized. Throughout this project, a ML algorithm would be taught on publicly released individual stocks, obtain relevant data, as well as then apply the details to effectively predict. ANN has been discovered. more appropriate than any other method after significant research into numerous algorithms and their robustness in diverse problem domains. It is easier to implement the customised neural network model, which comprises a large number of features and parameters. An important method used in this paper is to obtain the results of predicting machine learning concepts and the results the data set for the Bombay Share market Exchange was put to the test. [1].

Jan Ivar Larsen wrote the paper "Stock Price Prediction Using Technology Analysis and Mechanical Learning." Historical stock prices are utilised to forecast future stock price changes, according to the report. The employed stock price model employs a two-tier consultation approach, with the first layer of thinking using background knowledge from technical analysis to lead the second layer of thinking, which is depending on ML. A corporate finance strategy reinforces the system by determining the sum of funds to spend in forthcoming forecasts depending just on model's previous accuracy in shall be forecast. Researchers find that the prediction system outperforms the OSEBX depending on multiple investment scenarios and trade algorithm created by the methodology OSEBX.

Authors Soliman, Hegazy, and Salam wrote the paper "A Machine Learning Model for Stock Market Prediction." Stock market rate prediction, as per the report, is an approach to estimate possible trends of share prices as well as other economic assets transacted on numerous fronts. By properly estimating future stock values, independent traders can improve their earnings. This research uses an artificial intelligence technique to determine the share prices. The proposed approach includes particle crowd efficiency with the randomized forest technique. The swarm optimizer approach is used to build a stochastic forest to also predict market prices on such a regular basis. The above framework is designed on a sharemarket & technology indication database from the past. This particle swarm optimal method selects a subset of the most essential arbitrary forest assimilation features, avoiding additional overcrowding and the issue with local minima, or as such boosting prediction. The offered model has been tested on a range of payment datasets and matched to Leuralberg-Marquardt neuromorphic approaches. The obtained findings suggested that the model provided here has a greater accuracy and a more durable pso algorithm procedure. [7]. Aparna, Manohara, and Radhika, wrote the paper "Prediction Models for Indian Stock Market". The paper claims that stock market rate information is generated in large numbers as well as changes on such a regular basis. The share

market is indeed a complex and challenging process wherein individuals might gain additional gain or loss all perhaps they just own. This is an effort to forecast the share price trend. One type predicts prices for the following day, while the other predicts prices for the following month. The models are constructed using the learning algorithms of the monitored machines. Historical values are linked to emotions as part of the daily forecasting model. Using standardised machine learning approaches in daily prediction models, daily forecast models can achieve up to 70% perceived accuracy. The monthly forecasting model aims to determine whether any two-month courses have any commonalities. Tests prove that the trend for at least one month is related to the trend of other month [2].

Authors Guangyu and Qin published the paper "Study on the prediction of stock price based on associated network model of LSTM." . According to this document, Individuals and institutions from all walks of life have been drawn to the stock market. It has always been a fantastic meeting point for individuals and investment firms, since the sharemarket and its trends 're constantly shifting There seem to be currently a variety of approaches for predicting upcoming stock values. There are at least two types of approaches for predicting future share prices: Artificial Intelligence Mathematical Methods and Techniques The systematic model, the ARCH model, and other mathematical methods are examples. The MLP, Backward Propagations Networks, Convolutional Neural Networks, Bayes Networks, SVM, Single-Layer Long Short-Term Memory, and other AI approaches are examples. To estimate numerous values as outputs in a model, it is important to construct a model that can produce multiple values simultaneously and take many inputs. For this goal, A complex convolutional neural network having various intakes and outlets depending upon the short memory system is suggested. This network can estimate the price, open, high, and low of shares at the same time. Networking model that is practical In comparison to the Short Long Storage model as well as the Deep Learning Based model. Results show that accuracy in estimating multiple values parallelly the associated model has greater accuracy than the other models and Its estimated accuracy is more than 95% [4].

Prof. S.P., Karia, Tushar, Satyamandand, and M. Khan published the article "Stock Market Forecasts Using Machine Learning". According to the study, the guesswork will be built using a range of factors such as input to anticipate yet if or not the market price would be favorable. Amongst parameters used in the algorithm are fuel, real conversion value, loan, precious metals rates, Headline, twitter headlines sources, and trend analysis. A variety of ML methods are used including Vector Machines and Recurrent Neural Network [8].

Authors Mariam, El-Hajj, and Jaber was published the report "Automated Stock Price Prediction Using Machine Learning". Investors traditionally evaluate share values and stock indices to predict market movements, as well as event announcements connected to these shares, according to the report. As a result, the importance of news in the stock market. The bulk of earlier models either looked at historical price fluctuations and future projections value variations, or classified news as good, negative, or balanced and assessed its effect on share values. They devised a new trade system can improve stock forecasts and investing profitability by combining statistical methodologies, machine learning, and other observable factors including such media mood. We intend to identify the appraisal or tendency of a certain stock price for just a future point in time with examining the actual market period. To do this, they used efficient machine learning methods and also built and trained many in-depth comprehension models that considered the importance of relevant issues. Various tests were performed, with the highest accuracy (82.91%) obtained using Apple Inc.'s Support Vector Machine stock. (AAPL) stock [9].

Author V. Reddy published the report "Stock Market Forecasts Using Machine Learning". According to the article, Amongst the most significant companies on the planet is share trading. Stockmarket forecasting is the effort to forecast future values of many other investment vehicles exchanged on a currency conversion. Learning Machines are used to forecast stock prices in this paper. When formulating stock forecasts, most stockbrokers employ technical, fundamental, or temporal analysis. The syntax of a computer that forecasts the share market using Python active training. We provide a Computational Modeling (ML) framework which will be developed using currently accessible marketplace data in order to develop expertise and accurately predict in this study. The research used a neural network approach known as the Svm Classifier to forecast the rates of big and small companies as well as three separate marketplaces, utilizing regular as well as minute waveform prices. [13].

B. Existing System

The various forecasting algorithms in the current system can be divided into queues with Artificial intelligence frameworks. Points are awarded based on stock prices.

According to current models, the stock market employs only one algorithm to anticipate many situations and variables, and it does not combine or analyze numerous algorithm findings to accurately predict. Traditional divisions cannot be used to compute the stock price, according to previous results. Because it does not focus on external events such as

news events and other factors affecting pricing such as the Forex and Commodity markets, the existing system does not function optimally when the operational environment changes. It is incredibly biased because it only uses one data source. Because the existing system necessitates some form of input translation, it must be measured. The current approach only employs historical data or media analysis at the same time, not both.

Significant information like total trade and acquisition cost in trading activity, and also the proportion of such money that should be given and the proportion of distribution which indicates an acquisition or purchase by a significant financial adviser or big investors in a specific company, are ignored by the existing system. This element is frequently disregarded by new investors and established algorithms that do not account for these factors when analyzing market movement.

The combination of these characteristics plus volume data can be a useful predictor parameter that isn't taken into account much in current systems.

The majority of existing systems only use one algorithm and one set of data at a time. The current approach also tries to anticipate share prices in all conditions and on all days, but in the actual world, share markets are not always predictable, thus some conditions must be reviewed before projecting share markets.

III. PROPOSED METHODOLOGY

3.1. Artificial and Neural Network:

An sophisticated data gathering tool for detecting and summarizing a core trend using information is the ANN. This ANN may generate and interpret sophisticated structures in large datasets better than most previous techniques. A Human Brain with numerous hidden neurons underpins the concept. The design is made up of three layers. It consists of three layers: input, concealment, and output. H-L, O-C, and 1 Week MA, 2 Week MA, 3 Week MA, 1 Week STD DEV, and Quantity [18] are additional parameters with in input neurons. Prior to actually had been sent to the axons, the levels of perceived weights were compounded and combined. The deep learning model, also known also as layer, contains many axons. This total calculation is done and afterwards passed to final gradient output nodes. This network output, that forecasts the market's total value, is made up of just one neurons.

3.2. Random Forest (RF):

RF is a ML technique which uses an ensemble of machines to learn. It can do both regression and classification. Its goal is to combine several tree structure to decide the conclusion, rather than relying on single decisions tree to decrease forecast variance. Additional variables being created within that study for every decision tree's training, that influences the choice at the nodes of the tree. Sharemarket information includes a bunch of disturbance due to its huge scale, that also could also enable trees to grow in unexpected manner. The purpose is to decrease prediction inaccuracy by treating sharemarket monitoring as a clustering algorithm and utilizing learning parameters to estimate the following day ending value of a share for just a given business.

3.3. The system is further divided into following sections:

- Collection of data:

Data is gathered in the form of.csv files from a variety of sources, including yahoo finance and google finance. The system gathers news data using the Google News API.

- Manipulation, Visualization and Analysis of data:

Data collected is cleaned and pre-processed to make it ready to use for mal algorithms and models. and along with this data visualization is also performed [12].

- Building Model:

The information is extracted and well before being utilized to construct, build, and teach different ML techniques that may be utilized to make forecasts[14].

- Predict Outcomes:

Just after algorithm has indeed been appropriately developed, the next stage is to forecast an outcome pattern for a certain stock and double-check the accuracy of the forecasts.

- Forecast the outcomes among all techniques whenever used together:

The software then incorporates the findings of the algorithms for better forecast and precision, and predicts the outcome depending on available information.

We want to use multiple machine learning approaches in this proposed system to anticipate future share prices. We were able to train and test the ML algorithms in this proposed system using historical data to estimate future share prices. We considered EOD (End OF the Day) data of the past

years to test and train the ML models [13]. We utilized a range of ML packages and technologies to accomplish our aim, namely Sklearn, Kears, Pandas, Tensorflow and, Numpy. The other dataset for sentiment analysis included historical news headlines. The main library was NumPy library, which is commonly used to manipulate and clean and datasets, to convert it into form which can be directly used for ML models [11]. The other was sklearn library, which is used for actual calculation, estimation and prediction. The data we used was of historical share markets, and was gathered from the various public and open sources available online, and 85 or more percentage of dataset was used to training of the ML models depending upon the need and the rest had been applied for the purposes of verification and validation[15]. The most typical technique for a controlled ML model to provide results for the testing information would be to examine and memorize the patterns and relationships inside the dataset from train and validation datasets[16].

System does feature extraction, various pre-processing of data to make it ready for training of model [16]. For pre-processing data, such as integrating several datasets into an unified dataframe, the pandas python library was utilized. This dataframe had been created and was now ready to be analyzed for features. The dataframe included day, opening, closing, lowest, peak, supply %, volumes, number of deals, turnovers, as well as other related variables. We utilized all characteristics to build a decision tree structure and forecast the objects variables, and we applied certain specified characteristics to forecast the amounts of ownership of a future days with the LSTM and SVM algorithms[17]. On the test dataset as well as real values, we evaluated prediction accuracy. The method given here incorporates various components of research, such as data manipulation and pre-processing. The system employs an Artificial Neural Network (ANN) in order to analyze the sentiments of data sets and a Google News API to retrieve real-time data on stock market news.

Output of our web application which will be used for stock market prediction:

Stock Forecast App

Select dataset for prediction

GOOG

Years of prediction:

1

Loading data... done!

Raw data

	Date	Open	High	Low	Close	Adj Close
1531	Feb 1, 2021	1,853.5699	1,922.3920	1,850.9301	1,901.3500	1,901.3500
1532	Feb 2, 2021	1,922.5601	1,955.7600	1,914.4900	1,927.5100	1,927.5100
1533	Feb 3, 2021	2073	2,116.5000	2,018.3800	2,070.0701	2,070.0701
1534	Feb 4, 2021	2,068.8899	2,078.5500	2,042.5900	2,062.3701	2,062.3701
1535	Feb 5, 2021	2070	2,102.5100	2,059.3301	2098	2098

Fig. 1. Selecting the stock which we want to predict, here in this case we selected Google stock. After that selected number of year of prediction. After selecting the number of years it will fetch the raw data of Google stocks from its website.

Time Series data with Rangeslider

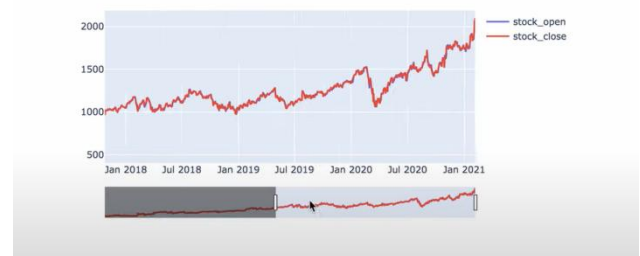


Fig. 2. Time series data with Rangeslider

Forecast data

	ds	trend	yhat_lower	yhat_upper	trend_lower	trend_upper
1896	Feb 1, 2022	2,164.4576	2,081.1744	2,346.6547	2,049.3826	2,277.075
1897	Feb 2, 2022	2,165.5988	2,082.5276	2,348.7877	2,050.1566	2,278.726
1898	Feb 3, 2022	2,166.7401	2,077.4739	2,342.5634	2,050.9306	2,280.377
1899	Feb 4, 2022	2,167.8813	2,072.9813	2,348.0629	2,051.7046	2,282.029
1900	Feb 5, 2022	2,169.0225	2,101.5272	2,366.8780	2,052.4786	2,283.680

Forecast plot for 1 years

Fig. 3. Forecast data



Fig. 4. Forecast plot

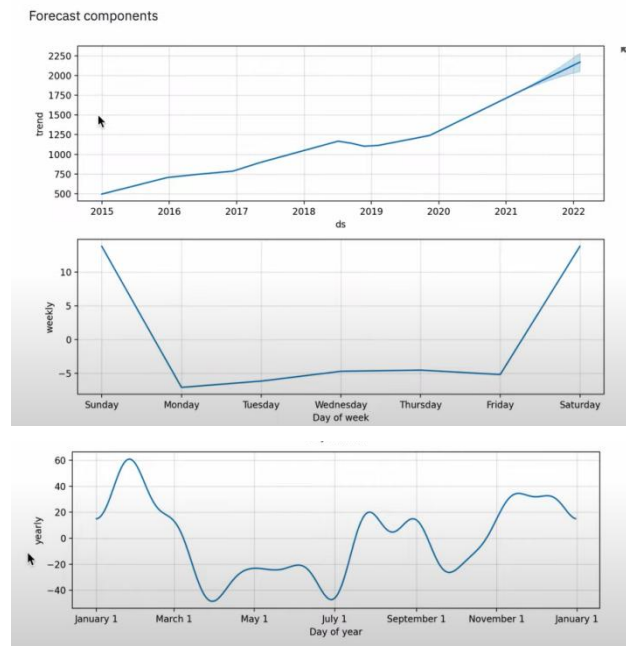


Fig. 5. Forecast components trend, weekly trend, yearly trend

IV. CONCLUSION AND WORKING PLAN FOR THE FUTURE

A. Conclusion

Using multiple algorithms such as ANN, Random Forest, SVM, and LSTM and combining their results for predictions, This was determined that using each of aspects impacting the share values, which including media emotions, currencies, asset costs, as well as various global share trading market information, is most accurate approach for share market predictions, while also considering all of the factors affecting the stock market price. All of this resulted in a high level of precision. The accuracy of stock price predictions improves when other factors affecting the price are taken into account. Predicting the stock market only when it is more predictable and avoiding it when it is more uncertain yields higher earnings. For example, predicting only when the standard deviation was within the threshold values enhanced the accuracy of our algorithm by 10% or more. The techniques are a useful resources for individuals and investment firms wishing to participate in these share markets since they are developed on such a big amount of past data and afterwards picked to also be evaluated on observations.

B. Future scope

More granular data, such as data collected every minute, second, or tick by tick, can enhance forecast accuracy since it contains more useful information than one-day data. Future work will include making forecasts and recommendations for individuals based on their risk and reward appetites, as well as portfolio management to reduce the risk of participating in the stock market. The process of placing orders and making trades can be automated, with AI and ML algorithms providing ideas. This eliminates human errors and elements like fear and greed, resulting in faster and more successful trading.

REFERENCES

- [1] RautSushrut Deepak, ShindeIshaUday, Dr. D. Malathi, "Machine Learning Approach In Stock Market. Prediction" <https://www.acadpubl.eu/jsi/2017-115-6-7/articles/8/12.pdf>
- [2] Aparna Nayak, and Radhika M. Pai. M. M. Manohara Pai "Prediction Models for Indian Stock Market" <https://www.researchgate.net/publication/306362748-Prediction-Models-for-Indian-Stock-Market>
- [3] Jan Ivar Larsen, Norwegian University of Science and Technology "Predicting Stock Prices Using Technical Analysis and Machine Learning" <https://pdfcoffee.com/predicting-stock-prices-using-technical-analysis-and-machine-learningpdf-pdf-free.html>
- [4] Liangxi Qin, Guangyu Ding "Study on the prediction of stock price based on the associated networkmodel of LSTM" <https://link.springer.com/article/10.1007/s13042-019-01041-1>
- [5] Xi Zhang, Siyu Qu, Jieyun Huang, Binxing Fang, Philip Yu "Stock Market Prediction via Multi-Source Multiple Instance Learning" <https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8464661>
- [6] Pei-Yuan Zhou , Keith C.C. Chan, Member, IEEE, and Carol XiaojuanOu, "Corporate Communication Network and Stock Price Movements: Insights From Data Mining" <https://www.semanticscholar.org/paper/Corporate-Communication-Network-and-Stock-Price-Zhou-Chan/12fdef1c6522cfe1d15de37eeeb739063bd362d9>

[7] Osman Hegazy, Omar S.Soliman and Mustafa Abdul Salam "A Machine Learning Model for Stock Market Prediction"
https://www.researchgate.net/publication/260393534_LSSVM-ABC_Algorithm_for_Stock_Price_prediction

[8] Prof. Kailash Patidar, Asst. Prof. Megha Jain, SachinSampatPatil "A Survey on Stock Market Prediction Using SVM"
<https://www.semanticscholar.org/paper/A-Survey-on-Stock-Market-Prediction-Using-SVM-Patil-Patidar/230644662d8d581840be8630f6df43202b43d5>
 51

[9] Mohamad Jaber, MariamMoukalled, Wassim El-Hajj "Automated Stock Price Prediction Using Machine Learning" <https://aclanthology.org/W19-6403.pdf>

[10]] V Kranthi Sai Reddy "Stock Market Prediction Using Machine Learning"
https://www.researchgate.net/publication/328930285_Stock_Market_Prediction_Using_Machine_Learning

[11] Hakob Grigoryan, "A Stock Market Prediction Method Based on Support Vector Machines (SVM) and Independent Component Analysis (ICA)"
https://www.dbjournal.ro/archive/23/23_2.pdf

[12] Loke.K.S. "Impact Of Financial Ratios And Technical Analysis On Stock Price Prediction Using Random Forests"
<https://ieeexplore.ieee.org/document/8270396>

[13] Kranthi Sai Reddy, "Stock Market Prediction Using Machine Learning"
https://www.researchgate.net/publication/328930285_Stock_Market_Prediction_Using_Machine_Learning

[14] Ashish Sharma, Dinesh Bhuriya, Upendra Singh. "Survey of Stock Market Prediction Using Machine Learning Approach"
<https://ieeexplore.ieee.org/document/8212715>

[15] VivekKanade, BhausahDevikar, Sayali Phadtare, Pranali Mande, ShubhangiSonone. "Stock Market Prediction: Using Historical Data Analysis"
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3871022

[16] Prof. S .P. Pimpalkar, Jenish Karia, Muskaan Khan, Satyam Anand, Tushar Mukherjee "Stock Market Prediction using Machine Learning"
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3871022

[17] MariamMoukalled, Wassim El-Hajj, Mohamad Jaber "Automated Stock Price Prediction Using Machine Learning" <https://aclanthology.org/W19-6403.pdf>

[18] Business Finance Stock Market News- Yahoo Finance , <<https://in.finance.yahoo.com/>> [Accessed on August 16,2018]

[19] Kumar, Manish, and M. Thenmozhi. (2006) "Forecasting stock index movement: A comparison of support vector machines and random forest" In Indian institute of capital markets 9th capital markets conference paper

[20] Li, Lei, Yabin Wu, Yihang Ou, Qi Li, Yanquan Zhou, and Daoxin Chen. (2017) "Research on machine learning algorithms and feature extraction for time series." IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC): 1-5.