JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

PREDICTION OF TSUNAMIS USING MACHINE LEARNING ALGORITHMS: A SURVEY PAPER

[1]Neha Pappachan, [2]Sai Dhanush V, [3]Jayashree M Kudari
[1]UG Scholar, [2]UG Scholar, [3]Associate Professor
[1]Department Of BCA
[1]Jain (Deemed-to-be-University), Bangalore, India

Abstract : This study has been undertaken to investigate the determinants of stock returns in Karachi Stock Exchange (KSE) using two assets pricing models the classical Capital Asset Pricing Model and Arbitrage Pricing Theory model. To test the CAPM market return is used and macroeconomic variables are used to test the APT. The macroeconomic variables include inflation, oil prices, interest rate and exchange rate. For the very purpose monthly time series data has been arranged from Jan 2010 to Dec 2014. The analytical framework contains**.**

I. INTRODUCTION

Tsunamis are produced when the ocean receives energy from an earthquake. This energy is produced by relative motion between the plates. When all of the earthquake's energy is released in a thrust motion, a massive tsunami is generated. Understanding this mechanism will therefore aid in the development of countermeasures, resulting in the saving of lives. For many countries to stay intact, there are multiple attributes such as Natural Disasters, Catastrophes and much more. Securing a country from Natural Disasters is critical as it ensures Financial and Economic well-being of a country, and saving countless lives. Predicting upcoming disasters is extremely crucial for multiple countries failing to which, there will be loss of lives, economical loss, infrastructural loss, etc., Let us now see the various ML approaches and compare them with each other to analyze which algorithm trumps in providing the highest accuracy in predicting the upcoming Tsunamis.

II. Objectives

To comprehend the mechanism of tsunamis and investigate various machine learning methodologies, compare them, and classify algorithms into static and real-time categories. To decipher the best algorithm(s) to employ to forecast impending calamities.

III. BACKGROUND

When earthquakes strike the areas and in the vicinity of the magma chambers, causing a crake in the seabed leading to these chambers, water will flow into the chamber, interact with the highly hot magma, resulting in the transformation of water into steam, its volume increased by 1,700 times, creating the force of pressure that will push water and steam out of the chamber, starting the tsunami wave, approaching beaches with foams, which is a condensed lava [1].

The following are the five major traits have been regularly observed:

a) Earthquakes are always the cause of tsunamis [1].
b) Immediately following earthquakes, a drawback of coastal water began, which might continue over hundreds of meters [1].
c) The withdrawal is preferred with a sucking sound [1].
d) Water flow (1-5 m) increased in return [1].

The excessive velocities of the overland flows brought on heavy casualties and severe harm for the duration of the tsunami attack[16].

Thus, understanding this mechanism will help in the creation of counter-measures, which will result in the saving of lives, the prevention of the phenomenon's destructive force, the decrease of the consequences of its impact on local societies [1].

3.2 Process of Predictive Analytics

Predictive analytics, a subset of advanced analytics, makes predictions about future occurrences. Predictive analytics models have dominated the field of decision-support systems as interest in them has grown. It analyses the data and then makes predictions. After learning about the client's needs, the analyst will collect the data needed to create the predictive model, which might come from a variety of sources. To build a predictive model, it is necessary to understand what the prediction goal is. [2].

The predictive analytics process consists of following steps:-

a) Requirement Collection :- The analyst realizes the expectations of the client [2].
b) Data Collection :- The required data is gathered [2].
c) Data Analysis :- The data is converted to structured form [2].
d) Statistics, Machine Learning :- The required techniques are applied[2].
e) Predictive Modeling :- The model is developed [2].
f) Prediction and Monitoring:- The analyst makes the predictions and monitors the model [2].

Statistical models were formerly thought to be predictive models. Novel ways have been created over time as a result of breakthroughs in computer science and the expansion of computer technology [18].

3.3 Evaluation of predictive algorithm

Algorithms and models for automated data analysis are created with several promises in mind, such as the precision, completeness, up-to-dateness, optimization, and predictive capability acquired by enterprises that use these technologies. Many of these promises and expectations will, of course, be broken in practise, and algorithms may even produce results that contradict the goals that motivated the regulation decision in the first place. PredPol is an algorithm that promises to be capable of more than just geographical analysis and locating "hot spots," claiming to be able to determine locations to patrol to prevent crime with precision based on current data. An algorithm designed to predict seismic aftershocks served as motivation. The quality of the forecasts is determined by the notion of contagion rather than the completeness of the parameters and data put into the system. A quality prediction evaluation might be performed. Cross-city experience exchanges are another method for assessing the usefulness of algorithms [3].

3.4 Real-time tsunami detection

Real-time detection of a tsunami on instrumental sea-stage facts is pretty an essential assignment for a Tsunami Warning System (TWS). The end result is that TEDA is capable of come across fast the bulk of the tsunami indicators and consequently proves to have the capacity for being a legitimate device withinside the operational TWS practice [20].

The Tsunami Early Warning System algorithm will be linked into software that will identify tsunamis based on historical records. The disruptions induced by long tidal waves are investigated using a 'quasi' real-time method. The signals are detected by the Deep-ocean Assessment and Reporting of Tsunamis program's dynamometers. The identification of tsunami may be successfully created based on the given data. The method developed by Beltrami and Di Risio (2011) aids in presenting the waveform of a tsunami in real-time. After identifying the tsunami, an algorithm is utilized to determine the waveform. A temporal interval will be generated between two waveforms, and this observation will be noticed by the real-time method. The generated error is shown as a function of delay t, with varied wave periods. The shorter the magnitude, the greater the time delay [4].

Focus is at the amplitude-discriminating algorithm, in particular primarily based totally on an endless impulse reaction time area filter, proposed with the aid of using the authors in a preceding paper. In particular, the troubles which can stand up in filtering out ‘disturbances’ together with lengthy tidal waves are addressed, and an authentic approach to carry out the entire characterization of the waveform of an simply detected tsunami, mechanically and in ‘quasi’ real-time is proposed[17].

The maximum perfect technique to estimate tsunami inundation as a result of a submarine earthquake is through undertaking a nonlinear tsunami simulation. However, this technique has the risks of a highly excessive computational value and the need for fast caution bulletins while a tsunami is imminent[19].

Chilean tsunami: February 27, 2010: Tsunamometers measure the depth of the water and the amplitude of the wave. The period of the wave is suggested by the measured tsunami frequency characteristics. As a result, the detection approach and characterization procedure should function admirably. It should be noted that the tsunami waveform is almost accurate in "quasi" real-time. The signals produced by the tidal-wave module before to and following the wind-wave module, on the other hand, are depicted in the middle and bottom panels, respectively [4].

Japanese tsunami: March 11, 2011: It was detected by nearly all of the DART network's dynamometers and recorded by nearly all of the coastline's tide gauges. The tsunamometers recordings offer a realistic record of the waves. As a consequence, a precise result is obtained. The current results have been compared to the prior records. After then, the signals are analysed using a real-time detection technique. Two sets of tests are run in order to evaluate the algorithm's performance. [4].

3. 5 Deep Predictive Coding Networks

Each network level is made up of four fundamental units: an input convolutional unit, a recurrent representation unit, a prediction unit, and an error representation unit. They aid in future prediction by learning the pattern of tsunami propagation during the training phase. Using a gradient descent, the model is trained by minimising the error units. This method was used to construct predictive coding networks, which were later recast in current deep learning approaches. [5].

Joint Data Assimilation-predictive coding: Because the predictive coding will instantly employ the DA's output to make future forecasts, the result of Data Assimilation (DA) is used for the tsunami warning system. Increasing the number of input frames may result in more accurate forecasts; nevertheless, it increases the calculation time of DA, causing the tsunami warning to be delayed. The sea surface displacement generated by an earthquake cannot be detected rapidly by shore bottom pressure sensors because the deformation wavelength is substantially greater than the ocean depth. Because the suggested method's learning process is based on a database, it is easier to add particular examples because the tsunami simulation is finished ahead of time. To address this problem, the DA should include a mechanism capable of causing sea surface deformation inside the source zone. As a result, the sea surface height immediately after the earthquake is roughly identical to the pre-event level. [5].

3.6 Batch Training Method

Based on the natural disaster data set for 2008-2019, we can predict the fore coming tsunami occurrence using the Batch Training Algorithm which produced 91% accuracy. Disasters can be caused in 2 ways- Natural catastrophes as well as disasters caused by human activities such as logging. Tsunamis occurred more often between 2008 and 2018, according to the National Emergency

Management Department of Republic Indonesia, but then began to decline. To predict this, one of the prominent algorithms that can be used is The Batch Training Algorithm. [6]

The Batch Training Algorithm is under ANN which is widely used for estimation, prediction, and pattern recognition.[11] To make this algorithm work accurately, it was first executed by Ginantra, which did not provide the results they expected. It was then trained by Wanto in Indonesia which provided 75% accuracy. [6]

Methodology- The BTA works on time series data. It requires 3 conditions to be fulfilled-

1. Variables such as Number of deceased, disappeared, broken houses, damages, and so on.
2. Results Of Training- Highest accuracy- 91% Overall accuracy - 75%
3. Prediction Results- Using the formula, predict the outcome-

$$- \mathrm{X_n} \frac{(x-0.1)*(b-a)}{0.8} + a \tag{1}$$

In formula (1),

X(n)= Prediction Result

x= Predicted Target

a= The smallest data from the dataset

b= The largest data set from the dataset [6]

3.7 Pitching Prediction Methods

To get real time data and work with it simultaneously, there are hardware requirements such as sustain antennas, good transfers, and many more. To increase the accuracy of the algorithm, short term pitch prediction will be required. Pitching models can be categorized into mathematical and non-mathematical forms.

One of the pitching methods is known as the Kalman Filter Algorithm. This algorithm assumes that pitching materializes randomly. Its advantage is that it be trained offline, on the other hand, its disadvantage is that is requires a lot of computing power, and is time consuming.[12]

To compare different pitching algorithms, a KF algorithm was compared with a NN algorithm. A pitch detection algorithm (PDA) is a program that calculates the pitch or fundamental frequency of an oscillating signal, which is commonly a digital recording of speech or a musical note or tone. After applying statistical formulae and KF algorithm, the standard deviation of noise σ=45deg, whereas NN based algorithm throws errors. In other cases, NN algorithm does not throw errors. To understand the dependence of results of both algorithms, 4 identical datasets were used on both and trained together .As our focus was to figure out the accuracy of 2 different pitch prediction methods the results completely varied depending on the dataset. KF method provides more accuracy and allows to obtain characteristics of prediction accuracy in real time unlike NN method. On the other hand, the NN method throws a great deal of errors and is very sensitive to the dataset provided.

3.8 Support Vector Machine, Denoising Autoencoder, Variation Autoencoder

Based on hypothetical short-time data from various observation locations, many machine learning techniques for forecasting tsunami amplitudes at a collection of forecast points were investigated. Training a support vector machine (SVM) to forecast the highest amplitude at the anticipated sites is one of the ML techniques examined. To predict the whole time series at the predicted sites, two deep convolutional neural networks were used.

A variational autoencoder (VAE) and a denoising autoencoder are two types of autoencoders (DAE). [13]

a) Support vector machines (SVMs):- They are supervised learning models that can perform nonlinear regression and classification on huge datasets.
b) Denosing Autoencoder:- Using a type of neural network known as the denoising autoencoder, we will predict the whole time series of the surface elevation in the forecast window, rather than just the highest surface elevation over the window.

Variation Autoencoder:- To set confidence boundaries on the output, we can train an ensemble of models. A VAE is an autoencoder that has been taught to have desired probability distributions for its latent random variables. Between the recognizing and producing models, the VAE may be split down into two models: a probabilistic encoder and a probabilistic decoder, with random latent variables in between.

The SVM uses time series data to provide results. We get faster and better results by using SVM than many existing methods and algorithms. We train the algorithm by providing hundreds of hypothetical attributes. To implement SVM, we first find a dataset, and pre-process it for it to work accurately. Once the data is prepared, SVM method can be implemented to perform non-linear regression and classification. It was observed that the longer SVM was being trained, the more accurate the results appeared. The observed data was compared to the scatter graphs for the three different methodologies, along with the projected maximum amplitude of the tsunami at the forecast gauges. All three methodologies may yield projections that are generally correct when compared to the data. As one might assume, forecasts based on a 60-minute observation window outperform predictions based on only 30 minutes of observable data in each case. When using the SVM technique on raw data, the results were typically poor, but when using the SVM on extracted features, the results were significantly better. In all of the testing, the DAE came out on top. The VAE and RFR for gauge 901 with 60-minute window and gauge 911 with 30-minute window, respectively, fared somewhat better than the VAE in terms of explained variance score.

3.9 Artificial Neural Network

Majority of the ML algorithms used for natural disaster prediction is the ANN. To understand the accuracy of ANN for Tsunami, testing in various locations and using real data is crucial. The ANN was trained using a variety of data, including fault displacement, fault length, fault breadth, fault slope, depth from the seabed, and more.

Non-linear shallow water wave equations were used to create the dataset. The ANN may be trained using the Levenberg-Marquardt technique and Bayesian regulation. ANNs have worked in a wide range of fields, including coastal engineering.[14] ANN has been used to analyzes the stability of rubble-mound breakwaters, predict wave impact forces, predict tides, forecast waves, and predict wave overtopping. The ANN approach works well for interpolation but not so well for extrapolation. The complete value of ANN can be extracted using real-time data to forecast the next disaster as not all algorithms support real-time data. To predict the upcoming disaster, let us take an example of earthquake. When an earthquake happens, data is collected in two minutes and analyzed to see if it has the potential to trigger a tsunami. There are 66 tsunami warning zones in Japan. Tsunami profiles were modelled using non-linear shallow water equations because there was a scarcity of data. Several settings were used in a total of 20 separate scenarios to change the relative intensity of the asperities and background sources.

There are 2 approaches that can be taken to handle overfitting-

1. Early stopping- A training set, a validation set, and a test set for early stopping are used to divide the data into three subgroups. When the validation error is at a minimum, the training is complete.
2. Bayesian Method- The performance function is modified to incorporate the total of the weights, resulting in a smoother network response.

The Bayesian method seemed to be useful and hence it was used along with the LM fast method.

Under various fault models, the nonlinear shallow water wave equations were employed to create both the training and test data sets for the ANN. The Levenberg-Marquardt technique with Bayesian regulation was employed for the ANN's training. These findings indicate that an ANN can provide rapid and accurate tsunami forecasts based on water heights recorded offshore. The ANN predicts both the preliminary and secondary tsunami waves. By combining the existing operational JMA tsunami database with the real-time warning system predicted by an ANN, it will be possible to instantaneously quantify a tsunami's potential and offer quantitative tsunami alerts.

3.10 DART Prediction Algorithm

DART or Deep-ocean Assessment and Reporting Tsunami is an early warning system that is being operated in certain areas by NOAA (National Oceanic and Atmosphere). We've already seen how the Kalman Filter Algorithm and the Artificial Neural Network (ANN) work. Let's compare the two techniques now. The device consists of a BPR that measures water levels with 0.1-millimeter resolution in the Tsunami frequency band. The Bottom Pressure Recorder (BPR) transmits information through acoustic telemetry. The quantity of samples required, accuracy, and computation time are all significant elements of Tsunami detection methods.

a) Dart Prediction Algorithm: The pressure is predicted using a 10-minute average measurement for 190 minutes using this tsunami detection algorithm created by mofjeld as part of the data protection initiative.
b) Kalman Filter: The Kalman filter is a recursive technique that uses observations over time to estimate an unknown quantity. There are two steps to the algorithm. It generates an estimate of the present state of variables, as well as their uncertainty, in the first phase. The parameters are then updated using a weighted average with a greater weight for more accurate measurement.
c) ANN Algorithm: The artificial neural network algorithm which was proposed by Beltrami is similar to that of a dart algorithm. Both the algorithms filters out the astronomical and meteorological surges recorded by BPR (Bayesian personalized ranking). [15]

All 3 methods were compared with each other with attributes such as RSME, Percentage error, correlation and coefficient, coefficient determination. All algorithms detect the event signal at the same time, and the Kalman algorithm takes longer to fall than the other two. The average percentage of errors is lower in ANN. The correlation coefficient in forward modelling has a higher strength, which is favorable

IV. Conclusion

A systematic effect is produced by an algorithm, and this effect is decided to answer the fundamental challenge of determining the correctness of the result set. There is an abundance of ML approaches one can take to predict Tsunamis. We noticed various approaches such PredPol algorithm, Kalman Filter, ANN, and numerous other approaches. The proposed system is capable of understanding the Tsunami propagation. KF algorithm provides Improved accuracy even in the face of measured mistake. Results of these experiments demonstrated that batch training method provided 91% accuracy.Overall, ANN is the root of any algorithm used for Tsunamis and is the go to method if one wants to predict imminent Tsunamis or any natural disasters in general. Both preliminary and secondary base can be predicted using ANN. These potential Tsunami sources detects future Tsunamigenic events and issue early warnings.

V. Acknowledgement

The successful completion of any task would be incomplete without thanking the people who made it possible whose constant guidance and encouragement motivated my effort with success. It is our great pleasure to thank our guide, Prof Jayashree M Kudari for her excellent guidance, constant encouragement, support, constructive suggestions. I, also thank all the faculties of the Computer Application Department for their suggestions enabled us to surpass many of the seemingly impossible hurdles. We would also like to thank all the staff members of the Department of Bachelor of Computer Applications-Data Analytics for their inspiration and kind cooperation in completing the paper. We are also grateful to all the university staff for their consistent support and assistance

References

[1]] David Demortain & Bilel Benbouzid ,Evaluating Predictive Algorithms ,Laboratoire Interdisciplinaire Sciences Innovations Sociétés [demortain@inra-ifris.org] Joint Workshop on Algorithmic Regulation, CARR/LSE and TELOS/King's College London, 3-4 July 20017

[2] M. Di Risioa,∗ , G.M. Beltramia,, P.le Pontieri Monteluco di Roio, L'Aquila,,12th International Conference on Computing and Control for the Water Industry, CCWI2013 Algorithms for automatic, real-time tsunami detection in wind-wave measurements: using strategies and practical aspects

[3] Ardiansyah Fauzi and Norimi Mizutani, Potential of deep predictive coding networks for spatiotemporal tsunami wavefeld prediction

[4] D. Antonov et al, Comparative Analysis of Pitching Prediction Algorithms 2022

[5] Christopher M. Liu, Donsub Rim, Robert Baraldi , Randall J. LeVeque, Comparison of Machine Learning Approaches for Tsunami Forecasting from Sparse Observations

[6] N L W S R Ginantra, M A Hanafiah, A Wanto, R Winanjaaya and H Okprana STMIK STIKOM , Politeknik Bisnis, PematangsiantarSTIKOM Tunas ,Utilization of the Batch Training Method for Predicting Natural Disasters and Their Impacts , Indonesia.

[7] Real-Time Prediction of Tsunami Magnitudes in Osaka Bay, Japan, Using an Artificial Neural Network Article in Journal of Waterway, Port, Coastal and Ocean Engineering · September 2011 DOI:10.1061/(asce)ww.1943-5460.0000092

[8] A Comparison of Forecasting Algorithms for Tsunami Detection with In-Situ Data

[9] Tata Sudhakar, C.D. Suriyakala, Estimation of Tsunami Direction and Velocity using Deep Sea Data, Arakkunnam, Ernakulam, Kerala, India

[10] Angel Bueno , Luz Garc´ıa , and Carmen Ben´ıtez,A Deep Neural Networks Approach to AutomaticRecognition Systems for Volcano-Seismic EventsManuel Titos

[11] P. Parulian et al., "Analysis of Sequential Order Incremental Methods in Predicting the Number of Tolls Affected by Disasters," Journal of Physics: Conference Series, vol. 1255, no. 1, pp. 1–6, 2019.

[12] S. Haykin and N. deFreitas, "Special Issue on Sequential State Estimation," in Proceedings of the IEEE, vol. 92, no. 3, pp. 399-400, March 2004, doi: 10.1109/JPROC.2003.823140.

[13] LeCun, Y. (1989). Generalization and network design strategies. Connectionism in perspective, 19(143-155), 18.

[14] Mase, Hajime, Masanobu Sakamoto, and Tetsuo Sakai. "Neural network for stability analysis of rubble-mound breakwaters." Journal of waterway, port, coastal, and ocean engineering 121, no. 6 (1995): 294-299.

[15] Beltrami, G. M., Di Risio, M., & De Girolamo, P. (2011). Algorithms for automatic, real-time tsunami detection in sea level measurements. In The Tsunami Threat-Research and Technology. IntechOpen.

[16] Prasetya G., J. Borrero, W. de Lange, K. Black and T. Healy, Modeling of inundation dynamics on Banda Aceh, Indonesia during the great Sumatra tsunamis December 26, 2004.

[17] Beltrami, G. M., Automatic, real-time detection and characterization of tsunamis in deep-sea level measurements. Ocean Engineering, 2011.

[18] Eric Siegel, "Predictive Analytics", John Willey and Sons Ltd.,2016.

[19] Fauzi A, Mizutani N, Machine learning algorithms for real-time tsunami inundation forecasting: a case study in Nankai region,2020.

[20] Bressan, L. and Tinti, S, Detecting the Tohoku tsunami arrival on sea-level records in the Pacific Ocean: application and performance of the Tsunami Early Detection , March 2011.