Maximizing Cloud Resource Utility: Region-Adaptive Optimization via Machine Learning-Informed Spot Price Predictions

Kavita Srivastava and Manisha Agarwal

January 2, 2024

Maximizing Cloud Resource Utility: Region-Adaptive Optimization via Machine Learning-Informed Spot Price Predictions

Kavita Srivastava[1[99719213382]] and Dr. Manisha Agarwal[2[9414543796]]

[1]Institute of Information Technology & Management, Delhi, India
[2]Banasthali Vidyapith, Rajasthan

kavitasrivastava@iitmipu.ac.in
manishaagarwal18@yahoo.co.in

Abstract. This research paper presents a comprehensive study on the use of machine learning models for price prediction of spot instances in various geographic regions in Amazon Web Services (AWS). The work focuses on forecasting prices across eleven unique locations using XGBoost and Random Forest regressors, with the goal of revealing significant insights into pricing dynamics and prediction accuracy. The research explores how well these models anticipate prices, finds factors that influence price fluctuation, and assesses the practical consequences of these predictions for enterprises.

The study employs a dataset containing pricing data from several places in a methodical manner. The study's findings reveal noteworthy trends and findings. The predicted performance of the models varies by region providing for region-specific insights into price forecast accuracy. To measure prediction performance, models are evaluated using Mean Squared Error (MSE) and Mean Absolute Error(MAE).

Significantly accurate forecasts show that the models can successfully capture pricing changes. A comparison of the XGBoost and Random Forest models also offers light on their relative performance, which will benefit in algorithm selection for future investigations.
.

Keywords: AWS, Spot Instances, Price Prediction, Random Forest Regressor, XGBoost.

1 Introduction

AWS Spot Instances were introduced by Amazon Web Services in December 2009. These instances allow users to bid for and use spare computing capacity in AWS's data centers at significantly lower costs compared to On-Demand or Reserved Instances. Spot Instances are ideal for workloads that are flexible in terms of timing and can tolerate interruptions, as AWS can terminate them if the capacity is needed by On-Demand or Reserved instances. Price forecast accuracy is a crucial aspect impacting decision-making across industries. As firms operate in more dynamic and competitive markets, the ability to estimate pricing precisely has strategic importance. Machine learning approaches have transformed price prediction, enabling data-driven models to capture complicated correlations between factors and provide insights into pricing dynamics.

This research study provides an in-depth examination of the use of machine learning models for price prediction in various geographic regions in AWS for AWS spot instances. The study use the XGBoost and Random Forest regressors to examine price trends in eleven unique locations. The study aims to uncover subtle insights into the accuracy of predictions, geographical factors impacting prices, and the practical consequences for businesses and decision-makers by training and testing models particular to each location. With the help of the experiments that we have conducted, we can find out how accurately machine learning models can predict prices for different regions. Therefore, we can get insight into the level of accuracy achieved by the models and whether these predictions are consistent across regions. By training separate models for different regions, we can identify which factors contribute the most to price variability in each region. This helps understand the key drivers of price fluctuations across different geographic areas. Furthermore, the results obtained can uncover whether there are region-specific trends or patterns in pricing. It helps determine if certain regions follow unique pricing dynamics due to local market conditions or consumer behaviors. Moreover, we can identify regions with the most predictable pricing trends (low prediction errors) and regions with less predictable pricing (high prediction errors). This information is valuable for risk assessment and decision-making.

The paper is organized as follows: Section 2 presents a literature overview. Section3 describes the methods used, including data collection, preprocessing, and the training and assessment processes. Section 4 presents the study's findings, which include predicted accuracy metrics such as Mean Squared Error (MSE) and Mean Absolute Error (MAE) for each location. Furthermore, we have used the trained model to predict spot instance price for a specified region using sample data. Also, a practical use case of dynamic resource provisioning is presented.

The study's findings reveal a variety of patterns and trends in pricing dynamics across areas. The models have varied degrees of prediction accuracy, giving insight on the areas where machine learning approaches are more efficient at catching price changes. Furthermore, a comparison of the XGBoost and Random Forest models reveals information about their relative performance features.

This study contributes to the broader understanding of how machine learning models can be harnessed to predict prices in a multitude of industries and contexts. The insights garnered from the research hold implications for businesses seeking to optimize pricing strategies, allocate resources efficiently, and gain a competitive edge through informed decision-making. Ultimately, the findings underscore the transformative potential of accurate price predictions in navigating the complexities of a rapidly evolving market landscape.

2 Literature Review

Since the introduction of spot instances, research in this area has evolved to address various challenges. Early research primarily focused on understanding the dynamics of Spot Instance pricing, bidding strategies, and optimization techniques to maximize cost savings while ensuring workload completion. Researchers have proposed algorithms to predict Spot Instance prices based on historical data and market trends to aid users in making informed bidding decisions. These predictive models aim to reduce the uncertainty associated with bidding and enhance the usability of Spot Instances for a wider range of workloads.

Furthermore, research efforts have explored adaptive strategies that dynamically adjust bidding strategies based on real-time changes in Spot Instance prices and user requirements. This approach helps users adapt to sudden price changes and maintain cost-efficiency. Utilizing the spot placement score (SPS) as a means of selecting spot instances that minimize interruptions results in cost-effective use of spot instances [1]. This approach aims to optimize the selection of computing resources for uninterrupted operations. DeepSpotCloud [2], a system designed to efficiently and reliably perform deep learning tasks using GPU-equipped spot instances on AWS EC2. During low demand periods, surplus cloud computing resources with GPUs are offered at reduced prices as spot instances. DeepSpotCloud leverages these spot instances across different regions globally to counter the price volatility of GPU spot instances. The impact of location on Spot Instance costs is explored [3], analyzing pricing data from various AWS regions over 60 days. It discovers that location greatly influences pricing, with substantial variations between regions. The impact of Amazon.com's pricing mechanism alteration for pre-emptible "spot instances" in cloud computing is investigated [4], aiming for smoother price changes. It assesses price variations before and after the change, comparing consecutive 90-day periods and the two most recent ones, upto October15, 2018. Results demonstrate that while the change achieved price smoothing, it also led to generally higher prices, an ongoing trend. To enhance spot instance utilization and attract more users, a framework is proposed [5]. This framework mitigates reclamation risk and reduces leasing costs by monitoring multiple markets and instance hopping. To illustrate the modern market's predictability, ARIMA models on both old and new data are employed [6], revealing a significantly improved accuracy for models trained on new data. This improved predictability makes spot instances highly suitable for cost-conscious cloud computing, A novel algorithm for spot price prediction is introduced [7], demonstrating its accuracy with 9.4% Mean Absolute Percent Error (MAPE) for short-term and less than 20% MAPE for long-term forecasting. While providers like AWS don't reveal pricing mechanisms, they share historical spot instance prices for the last 90 days. Utilizing a temporal convolution network (TCN), the price prediction challenge is transformed into a sequence data analysis task [8]. In comparative experiments, the TCN model demonstrates superior performance over traditional machine learning, CNN, and LSTM models in terms of MAE, MAPE, and RMSE metrics. A thorough examination of several machine learning methods for forecasting Amazon EC2 spot instance costs is provided [9]. A variety of techniques are examined, such as random forests, additional trees, K-nearest neighbors, multilayer perceptrons, ridge regression, lasso regression, and linear regression. The analysis shows significant differences in prediction accuracy between the various case categories, with high volatility instances having root mean squared errors that range from nearly zero to more than sixty. Notwithstanding these discrepancies, the majority of the time the overall predicting performance is judged to be exceptionally outstanding, indicating the need for more research and development in this developing sector. To improve cloud computing resource allocation dependability, accurate spot pricing forecast is essential. A modified gated recurrent unit (MGRU) model is built using Amazon EC2 as a testbed and its spot pricing history [10]. The study also presents a novel dropout technique to raise the prediction accuracy even higher. A machine learning-based method for analyzing and forecasting instance spot prices is provided [11]. It also covers implementation, a detailed exploration of the components, and results on many Amazon Elastic Compute Cloud (EC2) instances. This method lowers the amount of work and mistakes involved in pricing forecasting. Regression analysis is done on the spot instance price in order to forecast information that will be helpful to cloud customers and sellers who wish to begin selling spot instances [12]. In order to lower risk and improve pricing predictability for companies using the Amazon Web Services (AWS) elastic compute cloud (EC2) Spot instance pricing tier, methods are explored that use long short-term memory (LSTM) neural networks to accurately predict spot instance pricing over a given time-frame. The outcomes are then compared to traditional time-series Auto Regressive Integrated Moving Average (ARIMA) modeling [13]. The findings demonstrate that, in comparison to the baseline ARIMA model, the LSTM model Spot Instance price forecasts have an average mean absolute percent error (MAPE) reduction of almost 95%. Selecting virtual machines based on pricing trends can help save the expense of renting resources. Consequently, two markov regime-switching autoregressive models and a dynamic-ARIMA have been created as forecasting techniques [14]. For spot price prediction, the application of recurrent neural networks with long/short-term memory (LSTM) is investigated [15]. Through the use of historical spot pricing data, the model is evaluated against a baseline ARIMA model. It has been demonstrated that the LSTM method can cut training error by up to 95%.

3 Methodology

3.1 Data Collection

We have used AWS spot pricing data from Kaggle website[1]. The dataset has been organized in 11 csv files representing the data from 11 AWS regions namely ap-northeast-1, ap-northeast-2, ap-south-1, ap-southeast-1, ap-southeat-2, ca-central-1,eu-central-1, eu-west-1, sa-east-1, us-east-1, and us-west-1. The data has been collected in the month of April and May, 2017. The datasets in all the 11 CSV files have five attributes – price, datetime, instance_type, os, and region. The price attribute indicated the current price

[1]https://www.kaggle.com/datasets/noqcks/aws-spot-pricing-market?resource=download

of spot instance, whereas datetime represents the current date and time. Similarly, os refers to the operating system used and the instance_type refers to the spot instance type. Likewise, region attribute represents the region, and availability zone (AZ) for the spot instance. Now, after collecting the data, we wish to find out which of the region or availability zone has the least price for a particular instance_type. For this purpose, we have used two machine learning models – Random Forest Regression, and XGBoost Regression. We split the data contained in each CSV into training dataset and test dataset. Then we have trained both the models using the training dataset and evaluated the model using the test dataset. Finally, we have compared the performance of prediction for both the models.

3.2 Random Forest Regression

Random Forest Regression is a machine learning technique, which can handle non-linear correlations and high-dimensional data, mixes several decision trees to predict outcomes .Due to its ability to capture complex price determinants, handle outliers and noise, offer insights into influencing factors, and create robust predictions amidst the volatility of spot prices, Random Forest Regression is significant in predicting AWS Spot Prices. This helps users optimize their cost-effective use of AWS Spot Instances.

3.3 XGBoost Regression

XGBoost Regression is an advanced machine learning for forecasting AWS spot prices. Through gradient boosting and regularization, it excels at modeling intricate relationships and data patterns. This technique is essential for precise spot price prediction since it takes into account a number of variables, including instance type, region, and temporal patterns. By examining historical spot price data, XGBoost Regression successfully captures the dynamic nature of spot prices, empowering users to make knowledgeable choices when bidding on cost-effective spot instances in Amazon Web Services and optimizing resource allocation on the cloud infrastructure.

4 Results and Discussion

4.1 Price Prediction

Fig 1 demonstrates the price prediction of spot instances for each region using the Random Forest Regression. Whereas, Fig 2representsthe price prediction using XGBoost. Table 1 and table 2 show the corresponding values of Mean Squared Error (MSE), and Mean Absolute Error (MAE) in price prediction for each region using the two methods respectively.

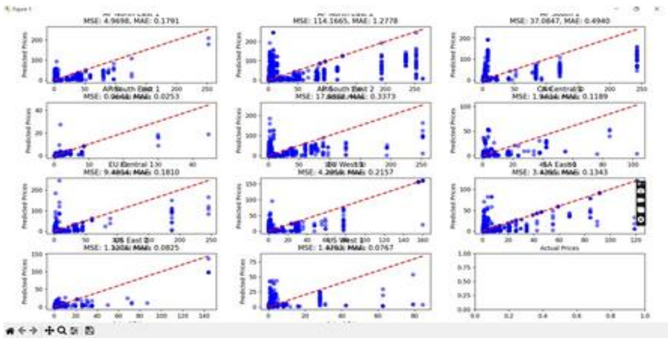

Fig. 1. Region-wise spot instance price prediction using Random Forest Regressor

Table 1. Region-wise Error in Predicting Price of Spot Instances Using Random Forest Regressor

Region	MeanSquaredError	MeanAbsoluteError
APNorthEast1	4.969769	0.179073
APNorthEast2	114.166456	1.277778
APSouth1	37.084668	0.494033
APSouthEast1	0.064842	0.025299
APSouthEast2	17.988840	0.337296
CACentral1	1.943412	0.118887
EUCentral1	9.493404	0.180981
EUWest1	4.295894	0.215713
SAEast1	3.439516	0.134340
USEast1	1.137628	0.082519
USWest1	1.479317	0.076733

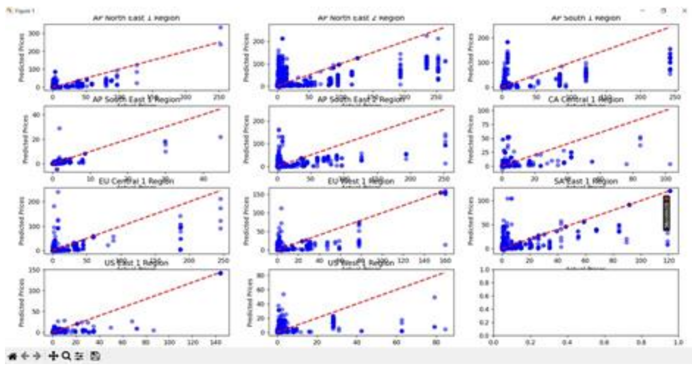

Fig. 2. Region-wise spot instance price prediction using XGBoost Regression

Table 2. Region-wise Error in Predicting Price of Spot Instances Using XGBoost Regression

Region	MeanSquaredError	MeanAbsoluteError
APNorthEast1	4.623801	0.214086
APNorthEast2	110.988923	1.366776
APSouth1	35.868513	0.537906

APSouthEast1	0.062595	0.029430
APSouthEast2	16.753965	0.367996
CACentral1	1.981949	0.127548
EUCentral1	9.185701	0.187455
EUWest1	4.385070	0.233918
SAEast1	3.538617	0.145718
USEast1	0.978867	0.084078
USWest1	1.334588	0.081077

MSE Comparison. The XGBoost program generally produces lower MSE values compared to the Random Forest Regressor program for most regions. For some regions, like "AP North East 1," "AP South East 1," "EU West 1," and "US East 1," XGBoost has notably lower MSE values, indicating better predictive performance in these cases. **MAEComparison.** Similar to the MSE results, the XGBoost program tends to achieve lower MAE values compared to the Random Forest Regressor program formost regions. Regions like "AP North East 1," "AP South 1," "AP South East 2," "CACentral1," "EUCentral1," and "EUWest1" show lower MAE values with XGBoost.

4.2 Discussion

The results that we have obtained by conducting these experiments have three major benefits. Firstly, it leads us to predict the prices of spot instances accurately for various regions so that we can make informed decisions on the pricing strategies, budgeting, and resource allocation. Secondly, we can get region-specific insights into the unique factors influencing prices in each region. Furthermore, accurate price predictions can lead to optimized allocation of resources. Businesses can better allocate inventory, advertising budgets, and other resources based on predicted prices, leading to cost savings and improved efficiency.

4.3 Comparison With Existing Approaches

In comparing our proposed approach with existing methodologies, it is discerned that the exploration of machine learning methods for spot price predictions is relatively limited. While [1] focuses solely on selecting spot instances with minimal interruptions, it falls short of providing insights into spot prices. Conversely, [2] analyzes spot instance price history logs, concentrating on dynamic price changes, and [3] examines pricing data to understand the impact of location on deployment costs, emphasizing price volatility. Meanwhile, [4] scrutinizes spot instance prices and assesses the effect of change on spot prices, yet it does not involve price prediction. Previously, spot price predictions have been tackled using ARIMA and LSTM recurrent neural network models, yielding RMSE values of 0.557 and 0.423, along with MAE values of 0.362 and 0.320, respectively [15].However, with proposed approach the corresponding lowest RMSE value is 0.2546 (MSE 0.064842 for AP South East 1 region) whereas the corresponding value of MAE is 0.025299, which are considerably lower. Hence the proposed approach accurately predicts the price of AWS spot instances.

4.4 Using Trained Model for Prediction

After training the model using proposed approaches, we can use it to predict the spot instance prices for specific instance types. Table 3 shows the predicted prices for 10 randomly chosen instance types in AP North East 1 region.

Table 3. Price Prediction of Spot Instances Using Trained Model for Randomly Generated Instance Type Samples

Instance Type	Predicted Price
c4.8xlarge	1.82
m3.xlarge	0.17
m2.2xlarge	0.06
c3.8xlarge	0.52
c3.2xlarge	0.20
c4.2xlarge	0.25
c3.2xlarge	0.20
m3.xlarge	0.17
r3.8xlarge	0.46
c4.xlarge	0.10

4.5 Dynamic Resource Provisioning – Practical Use Case

The proposed method finds useful application in the deployment of an Amazon Web Services (AWS) dynamic resource provisioning system. First, the trained model is loaded, and then the AWS EC2 Boto3 client is initialized. Basically, Boto3 functions as the Python equivalent of the AWS SDK. The trained model is then used to determine a price estimate for specified region.

When an instance launch request is received, our dynamic resource provisioning system checks to see if the predicted price is less than the specified threshold. In the event that this requirement is satisfied, the system launches a spot instance.

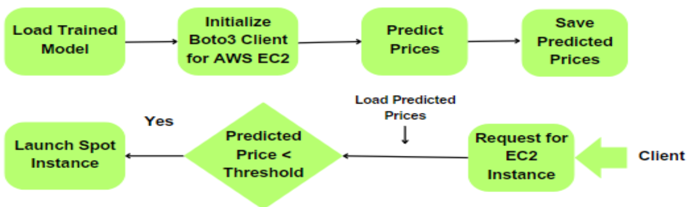

Fig. 3. Dynamic Resource Provisioning With AWS

5 Conclusion

Based on the output obtained, the XGBoost algorithm outperform the Random Forest Regressor in terms of both Mean Squared Error (MSE) and Mean Absolute Error (MAE) for most regions in spot instance price prediction. XGBoost demonstrates better predictive capabilities in capturing the relationships between features and prices in the dataset. This highlights the potential benefits of using XGBoost as a regression model for this specific prediction task. The benefits of the experiments that we have conducted include enhanced price prediction accuracy, insights into regional price trends, improved resource allocation, and the ability to make data-driven decisions that lead to increased profitability and strategic advantage in the market.

References

1. D. Katayama, K. Kasai and T. Koita: "Migration Destination Selection Algorithm for Spot-Instances using SPS,". In: IEEE International Conference on Big Data (Big Data), Osaka, Japan,2022,pp.6690-6692,doi:10.1109/BigData55660.2022.10020520.
2. K. Lee and M. Son: "DeepSpotCloud: Leveraging Cross-Region GPU Spot Instances forDeepLearning. In: "IEEE 10th International Conference on Cloud Computing (CLOUD), Honololu, HI,USA, 2017,pp.98-105,doi:10.1109/CLOUD.2017.21.
3. N. Ekwe-Ekwe and A. Barker: "Location, Location, Location: Exploring Amazon EC2Spot Instance Pricing Across Geographical Regions". In: 18th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing, 2018 (CCGRID), Washington, DC, USA,2018,pp. 370-373,doi:10.1109/CCGRID.2018.00059.
4. G. George, R. Wolski, C. Krintz and J. Brevik: "Analyzing AWS Spot Instance Pricing". In: IEEE International Conference on Cloud Engineering (IC2E), Prague, Czech Repub-lic, 2019,pp.222-228,doi:10.1109/IC2E.2019.00036.
5. A. J. Sanad and M. Hammad: "Reducing Cloud provisioning Cost Using Spot Instances hopping". In: International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), Sakhier, Bahrain, 2019, pp.1-6, doi:10.1109/3ICT.2019.8910325.
6. S. Caton, M. Baughman, C. Haas, R. Chard, I. Foster and K. Chard: "Assessing the Cur-rent State of AWS Spot Market Forecastability". In: IEEE/ACM International Workshop on Interoperability of Supercomputing and Cloud Technologies (SuperCompCloud), Dal-las, TX, USA, 2022,pp.8-15,doi:10.1109/SuperCompCloud56703.2022.00007.
7. V. K. Singh and K. Dutta: "Dynamic Price Prediction for Amazon Spot Instances". In: 48th Hawaii International Conference on System Sciences, Kauai, HI, USA, 2015, pp.1513-1520,doi:10.1109/HICSS.2015.184.
8. X. Song, R. Lin and H. Zou: "Amazon EC2 Spot Price Prediction Using Temporal Convolution Network". In: ICETIS2022; 7th International Conference on Electronic Technology and Information Science, Harbin, China, 2022, pp.1-6.
9. Baldominos, Alejandro & Sáez, Yago & Quintana, David & Isasi, Pedro: AWS PredSpot: Machine Learning for Predicting the Price of Spot Instances in AWS Cloud. International Journal of Interactive Multimedia and Artificial Intelligence. 7. 65–74. 10.9781/ijimai. 2022.02.003 (2022).

10. Nezamdoust, S.S., Pourmina, M.A. and Razzazi, F.: Optimal prediction of cloud spot instance price utilizing deep learning. *The Journal of Supercomputing*, *79*(7), pp.7626-7647 (2023).
11. Malik, M. and Bagmar, N.: Forecasting price of amazon spot instances using machine learning. *International Journal of Artificial Intelligence and Machine Learning (IJAIML)*, *11*(2), pp.71-82 (2021).
12. Mishra, Ashish & Yadav, Dharmendra: Analysis and prediction of amazon EC2 spot instance prices. International Journal of Applied Engineering Research. 12. 11205-11212 (2017).
13. Lancon, Jeffrey & Kunwar, Yejur & Stroud, David & Slater, Robert & McGee, Monnie: AWS EC2 Instance Spot Price Forecasting Using LSTM Networks. 2. 8 (2019).
14. Cai, Zhicheng & Li, Xiaoping & Ruiz, Rubén & Li, Qianmu: Price forecasting for spot instances in Cloud computing. Future Generation Computer Systems. 79. 10.1016/j.future.2017.09.038 (2017).
15. Baughman, M., Haas, C., Wolski, R., Foster, I. and Chard, K.: Predicting amazon spot prices with LSTM networks. In: *Proceedings of the 9th workshop on scientific cloud computing* (pp. 1-7), 2018.